

Second Language Learning and Teaching

Rahma Al-Mahrooqi
Christine Coombe
Faisal Al-Maamari
Vijay Thakur *Editors*

Revisiting EFL Assessment

Critical Perspectives

 Springer

Second Language Learning and Teaching

Series editor

Mirosław Pawlak, Kalisz, Poland

About the Series

The series brings together volumes dealing with different aspects of learning and teaching second and foreign languages. The titles included are both monographs and edited collections focusing on a variety of topics ranging from the processes underlying second language acquisition, through various aspects of language learning in instructed and non-instructed settings, to different facets of the teaching process, including syllabus choice, materials design, classroom practices and evaluation. The publications reflect state-of-the-art developments in those areas, they adopt a wide range of theoretical perspectives and follow diverse research paradigms. The intended audience are all those who are interested in naturalistic and classroom second language acquisition, including researchers, methodologists, curriculum and materials designers, teachers and undergraduate and graduate students undertaking empirical investigations of how second languages are learnt and taught.

More information about this series at <http://www.springer.com/series/10129>

Rahma Al-Mahrooqi · Christine Coombe
Faisal Al-Maamari · Vijay Thakur
Editors

Revisiting EFL Assessment

Critical Perspectives

 Springer

Editors

Rahma Al-Mahrooqi
Sultan Qaboos University
Al-Khuodh
Oman

Faisal Al-Maamari
Language Center
Sultan Qaboos University
Al-Khoudh
Oman

Christine Coombe
Dubai Men's College
Higher Colleges of Technology
Dubai
United Arab Emirates

Vijay Thakur
Department of Languages and Translation
Dhofar University
Salalah
Oman

ISSN 2193-7648 ISSN 2193-7656 (electronic)
Second Language Learning and Teaching
ISBN 978-3-319-32599-6 ISBN 978-3-319-32601-6 (eBook)
DOI 10.1007/978-3-319-32601-6

Library of Congress Control Number: 2016954556

© Springer International Publishing Switzerland 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Introduction: EFL Assessment: Back in Focus	1
Rahma Al-Mahrooqi	
Part I Assessment Literacy	
Assessment Literacy: Beyond Teacher Practice	9
Zineb Djoub	
Kurdish Tertiary EFL Teachers' Assessment Literacy in Alternative Assessments and Its Influence on the Ethicality of Their Assessments	29
Dler Abdullah Ismael	
Part II Theoretical Perspectives on Assessment	
A Critical Review of Washback Studies: Hypothesis and Evidence	49
Wei Wei	
Developing and Sustaining Outcomes Assessment in English as a Foreign Language Programs	69
Donald F. Staub	
Outcomes-Based Assessment in an English Language Program	85
Khadernawaz Khan and Umamaheswara Rao Bontha	
Assessment of EFL Through the Process of Problem-Based Learning	105
Melissa Caspary and Diane Boothe	
The Perception of Assessment as a Multilayer Dimension in the Armenian EFL Classroom	119
Marine H. Arakelyan	

Part III Skill-Area Assessment

EFL Assessment: Assessment of Speaking and Listening	133
Seetha Jayaraman	

Assessing Entry-Level Academic Literacy with IELTS in the U.A.E.	151
Christopher Morrow	

The Development, Validation and Use of a Test of Word Recognition for English Learners	171
David Coulson and Paul Meara	

Part IV Alternative Assessment

Alternative Assessment: Growth, Development and Future Directions	191
Vino Sarah Reardon	

Alternative Assessment: Student Designed Test Evidence in an Iranian EFL Context	209
Jafar Dorri Kafrani and Mohammad Reza Afshari	

From a Culture of Testing to a Culture of Assessment: Implementing Writing Portfolios in a Micro Context	221
Elizabeth Noel	

An Essential Tool for Continuous Assessment: The Learning Portfolio	237
Esra Gun Alayafi and Pınar Gunduz	

Believing in the Power of the Child: Reggio Recognizing the Affective	261
Nayyer Iqbal Ali Chandella	

Provoking Potentials: Student Self-Evaluated and Socially-Mediated Testing	287
Tim Murphey	

Part V Future Perspectives on Assessment

Electronic Intervention Strategies in Dynamic Assessment in an Omani EFL Classroom	321
Priya Mathew, Rahma Al-Mahrooqi and Christopher Denman	

The Future of E-assessments in the UAE: Students' Perspectives	339
Racquel Warner	

Introduction: EFL Assessment: Back in Focus

Rahma Al-Mahrooqi

Assessment is broadly defined as the activities teachers assign to learners to diagnose their learning proficiency or achievement for the purpose of directing and influencing the teaching-learning experience (Cheng, Rogers, & Wang, 2007). It is, hence, an integral part of any formal instructional endeavor. In fact, assessment is an important part of any well-structured educational process. While it is essential for ascertaining the achievement of educational goals and objectives, it is also a vital ingredient for continuing improvement and reform. However, assessing student performance is far from a straightforward process since it cannot be removed from the socio-historical context in which it exists (McNamara, 2000). It is for this reason that, traditionally, assessment has been concerned with the product and with finding weaknesses in student learning, usually at the end of a course (summative assessment). However, recently the emphasis has shifted from the product to the process and to the use of more creative, authentic, and dynamic assessment methods. Continuous or formative assessment is now favored despite the criticism it draws concerning its objectivity. Alternative testing methods, such as portfolios, peer assessment and self-assessment, have also come to the fore with researchers and practitioners now hailing their value for student learning.

The aim of this book, whose context is teaching English as a Foreign Language (EFL), is to focus on these practices to improve teacher literacy and to demonstrate to teachers and practitioners how they can make the best out of their assessment practices and the knowledge that has accumulated in the field. In addition to focusing on different assessment types such as alternative assessment, dynamic assessment, self-directed assessment, continuous assessment, and outcomes-based assessment, the book explores different assessment methods in skills such as reading, writing, listening and speaking. Furthermore, it attempts to forecast the future of assessment and where such concepts as alternative assessment and

R. Al-Mahrooqi (✉)
Humanities Research Center, Sultan Qaboos University, Muscat, Oman
e-mail: ralmahrooqi@gmail.com

dynamic assessment are heading. In doing so, it also shows how relatively new teaching methods such as communicative methodology and problem-based learning are reflected in assessment.

The book is in five parts. Part 1, *Assessment Literacy*, focuses on the education of the language teacher and the role that professional development programs plays to sensitize EFL teachers to core issues such as assessment for learning, teacher professionalism, ethical assessment, and the democratization of assessment. In Part 2, *Theoretical Perspectives on Assessment*, key issues such as washback, alternative assessment, and problem-based/outcomes based assessment are discussed. Part 3, *Skill-area Assessment*, presents a variety of assessment systems from the renowned international IELTS to the less known local in-house tests targeting specific skills reaching to the less traveled paths in the form of the validation of new measures. Part 4, *Alternative Assessment*, is what it says it is. Situated in the transition from a testing to an assessment culture, not only are examples of alternative assessment (e.g., student designed tests, portfolios) given, but also challenges and opportunities relating to students' involvement in AA's design and evaluation are described. In the final part, *Future Perspectives on Assessment*, the marriage between assessment and technological advances is embedded into a discussion of the grand edifices of cognitive (e.g., validity and reliability) and sociocultural theories (e.g., zone of proximal development). In short, this book brings together assessment, be it summative, formative, continuous, traditional, electronic, alternative, student-designed, teacher-based, for learning, of learning, or as learning, and presents a vivid description of both their connection to theory and their implications to practice.

The second chapter, "Assessment literacy: Beyond teacher practice" by Zineb Djoub examines how ESL/EFL teachers' assessment literacy affects their assessment views and practices. The author uses a survey administered to forty-five EFL/ESL teachers from different schools and institutions worldwide. The results revealed that, during their teaching experience, most participants had never been trained in language assessment. Instead, they received courses as part of their MA program. Hence, they possessed theoretical knowledge about assessment but they received inadequate training in terms of practice. Teachers perceived testing as an event that should neither promote anxiety among students nor establish and exert control over the teaching and learning experience. However, none of the teachers indicated willingness to involve their learners in the assessment process that would help them to reflect on their learning or monitor their progress to decide on how to improve it.

Chapter three, "Kurdish tertiary EFL teachers' assessment literacy in alternative assessments and its influence on the ethicality of their assessments" by Dler Abdullah Ismael, focuses on the link between assessment literacy in alternative assessment and the ethicality of teacher assessment practices using Kurdish English departments as a context for the study. The author examines the level of Kurdish EFL teachers' assessment and testing literacy and their perceptions and understanding of alternative assessment and how that influences their assessment-related practices.

In chapter four, “A critical review of washback studies: Hypothesis and evidence” by Wei Wei, the author examines the current understanding of washback effects in the context of language testing and education from three perspectives: (1) a historical perspective outlining the development of the concept of washback in both general education and language education; (2) a focus on results from empirical investigations to highlight the gaps between hypotheses and evidence; and (3) identification of the areas and directions for future washback studies.

Donald F. Staub’s chapter, “Developing and sustaining outcomes assessment in English as a Foreign Language programs” examines the principles and practices that are believed to be must-haves for successful outcomes assessment in EFL programs. The chapter also includes a discussion of common drawbacks that can result in failure of such principles and practices. Finally, the chapter proposes the distributed leadership model for EFL program leaders who are embarking upon an outcomes assessment process in order to increase the probability of success and the sustainability of their outcomes assessment initiative.

In chapter six, Khadernawaz Khan and Umamaheswara Rao Bontha reflect, based on action research, on outcomes-based assessment in an English language program in Oman. They explore the adherence of assessment to the teaching/learning outcomes of a course, and the implementation of alternative assessments (e.g. writing portfolios, continuous assessments and self-assessments) which are used to overcome the drawbacks produced by traditional paper based tests and exams. Teachers and learners’ perceptions regarding the achievement of the learning outcomes, conventional assessments and alternative assessments are also examined and analyzed.

Chapter seven “Assessment of EFL through the process of problem-based learning” by Melissa Caspary and Diane Boothe applies problem-based learning to EFL assessment in an attempt to meet learner needs and assess their progress. The PBL approach emphasizes assessment tools and models that address unique teaching styles and key competencies which, at the same time, stimulate critical thinking and effective teamwork. The chapter offers practitioners dynamic and creative strategies as well as assessment tools that cater to individual differences and create opportunities for collaboration and group synergy in order to promote and foster EFL academic achievement.

In chapter eight “The perception of assessment as a multilayer dimension in the Armenian EFL classroom”, Marine Arakelyan used questionnaires with open-ended questions administered to students and teachers to gauge participants’ perceptions of the multivoice notion of assessment. This notion views assessment not only as a diagnostic tool, but also as a trigger for new content creation. The author also offers her perceptions of the role ascribed to the practice of assessment in the language learning instructional processes and environment. Prior exposure to assessment in the context of language instruction was a variable the author looked at when examining teachers’ perceptions of assessment. Findings indicated that the views of teachers with previous exposure to assessment were in-line with the notion of ‘assessment for learning’, whereas the views of teachers with no prior exposure to

assessment and students' views indicated assessment and motivation for learners are analogous.

As its title indicates, the ninth chapter, "EFL assessment: Assessment of speaking and listening" by Seetha Jayaraman focuses on the assessment of both speaking and listening skills as the two have often been linked in teaching and assessment. It examines the methods and criteria involved in testing the English speaking and listening proficiency levels of a group of Arabic speaking undergraduate Omani students from the Dhofar region. This is done with the purpose of exploring those methods that suit these students best and help them attain the prescribed outcomes without experiencing high levels of anxiety.

Christopher Morrow's "Assessing entry-level academic literacy with IELTS in the UAE" examines the suitability of IELTS as a linguistic measure of Emirati students' language proficiency. In the chapter, Morrow offers the position that such an exam, with its advanced linguistic demands, is not well-suited for making valid and reliable decisions about the readiness of Arab students to begin college-level studies in English. Hence, he adds his voice to those of other professionals who call for a need to develop a more appropriate means of assessing students including by offering alternative assessment options.

Chapter eleven, "The development, validation and use of a test of word recognition for English learners" by David Coulson and Paul Meara reports on the process the authors used to develop a test of word recognition they called Q_Lex for EFL learners. In the test, words are hidden in nonsense letter strings which they predicted would slow recognition speed to a level that personal computers can easily measure. The test assesses learners on the basis of native speakers' reaction time norms. The authors describe the development and validation of this assessment tool and the measurement principles underlying it and emphasize the measures they took to improve its reliability. Finally, the chapter describes an experiment with Q_Lex which the authors conducted to investigate learners' word recognition abilities at different levels of proficiency.

The twelfth chapter "Alternative assessment: Growth, development and future directions" by Vino Reardon traces the growth and development of alternative assessment from a pedagogical perspective. She also examines the direction which alternative assessment has taken in recent years within special education programs, ESL and/or EFL classrooms, and elucidates the belief that this type of assessment is the most-suited to ESL/EFL situations for students with or without speech impairment.

In Jafar Dorri Kafrani and Mohammad Reza Afshari's chapter, "Alternative assessment: student designed test evidence in an Iranian EFL context", 120 junior high school male students were asked to design tests based on two chapters from an English course book they studied. The resulting questions were analyzed in terms of format, skills and sub-skills, and students and their teachers were interviewed. The results showed that students benefited from their tests, and the opportunity helped them to review the two chapters' content in detail. However, results also revealed that students paid inadequate attention to certain key sections of each chapter, and that their test-item formats lacked variety.

In the fourteenth chapter, “From a culture of testing to a culture of assessment: Implementing writing portfolios in a micro context”, Elizabeth Noel clarifies that, in recent years, there has been a focus on the use of portfolios as one form of alternative assessment in assessing college level writing. Despite the challenges associated with the use of portfolios, the author outlines the transformation that happened in the English for Academic Purposes department at the University of Technology and Business (UTB) from a culture of testing to a culture of assessment. The author also describes some of the changes occurring in the micro context of UTB, where portfolio assessment is being implemented to help solve curricular and other challenges.

Chapter fifteen, “An essential tool for continuous assessment: The learning portfolio” by Esra Gun Alayafi and Pinar Gunduz also focuses on the use of portfolios as a mode of continuous assessment. The authors describe the reasons behind using learning portfolios and then explain how their implementation took place at Sabancı University School of Languages (SL) in Turkey. The chapter also evaluates the current practices and procedures associated with the implementation of learning portfolios. The authors conclude by proposing future goals in light of the collected feedback.

The sixteenth chapter, “Believing in the power of the child: Reggio Recognizing the affective” by Nayyer Chandella, asserts that assessment for learning encompasses all factors influencing a student’s learning. To the author, assessment goes beyond being a quantification of test results. Good assessment practices, in her opinion, have to make use of student data to change teaching and learning in a positive way that will enable children to become successful learners and confident individuals. In this chapter, the author reports on a small scale research project based on the Reggio philosophy. She uses her own narrative to describe the Reggio approach to early childhood education and explores aspects of this approach adapted to the Pakistani context, with particular focus on portfolios and documentation as means of assessment.

Tim Murphey’s chapter, “Provoking potentials: Student self-evaluated and socially-mediated testing”, reports on an exploratory study conducted through action research where students are directed to evaluate themselves at two points in time. The author believes that the procedures he explored enlighten students about different aspects of learning and evaluation, and assist teachers in examining different aspects of classroom dynamics and learning potentials. According to the author, student self-evaluation and socially-mediated testing blend learning and assessment and theory with practice.

Priya Mathew, Rahma Al-Mahrooqi and Christopher Denman’s chapter, “Electronic intervention strategies in dynamic assessment in an Omani EFL classroom”, focuses on dynamic assessment (DA) as a method of formal testing and explores its suitability when applied in both electronic and face-to-face encounters. The sample included 12 Omani EFL learners studying in a foundation program who emailed their assignments throughout a semester to their instructor who then returned feedback using a word processor’s review function. Students were then assessed on their ability to incorporate the instructor’s explicit and implicit

feedback. A focus group interview with the students was held and a series of observations were carried out to explore emergent trends associated with DA. The results suggest that mediated electronic forms of DA attuned to participants' ZPD are more useful and conducive to learning than pre-scripted prompts resulting from assessors' guesses about the kinds of intervention learners may require during assessment. The chapter concludes by suggesting that electronic forms of DA can enhance Omani students' learning and, therefore, their application is recommended in the tertiary classroom.

The final chapter, "The Future of E-assessments in the UAE: Students' perspectives" by Racquel Warner features a qualitative study exploring teacher and student perceptions of e-assessment within a private higher education institution (HEI) in the United Arab Emirates. The author used questionnaires and interviews to examine participants' perceptions of virtual learning environments, e-assessment methods on virtual platforms, and the process of giving feedback on performance on e-assessments. The results indicated that most participants benefited in one way or another from e-assessments. However, they were reluctant to express full support for a transition to e-assessments as a sole method of summative evaluation. The study suggests ways of promoting the idea of e-assessment and assures both teachers and students that these new methods can actually increase reliability and validity of testing and hence can be regarded as an improvement over old methods.

With its nineteen chapters, this book is a forum where contributors present their research and innovative ideas and practices on the topic of EFL assessment and, in doing so, encourage renewed debate around the issue. Due to its breadth and variety, this book serves as an excellent reference for EFL teachers, practitioners, researchers and testing and assessment specialists. Each of its nineteen chapters is unique, examining important issues pertinent to assessment and its connection with teaching and learning in EFL contexts.

References

- Cheng, L., Rogers, W. T., & Wang, X. (2007). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment and Evaluation in Higher Education*, 33(1), 9–32.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.

Part I
Assessment Literacy

Assessment Literacy: Beyond Teacher Practice

Zineb Djoub

Abstract Language teachers nowadays should cope with the changing and challenging demands of society which requires more flexibility in assessment in order to support learning. Indeed, assessment is no longer used for merely measuring learning outcomes but also for creating more learning opportunities. To support the achievement of this goal, an increasing interest in developing teachers' assessment literacy via training and professional development courses has been gaining ground. Yet, reaching the intended objectives and effectively defining the contents and approaches of those courses depend on understanding the nature of assessment literacy. To this end, this paper aims to find out about this process through examining how ESL/EFL teachers' assessment literacy affects their assessment views and practices.

Keywords Assessment literacy • Enhancing learning • Teacher training

1 Introduction

Assessment of language learning has been the primary concern of several researchers, teachers, test developers, syllabus designers, etc. It is a vital component of the educational process which serves a variety of purposes such as diagnostic, achievement, progress, among others. Hence, the challenge which remains either unaddressed or not addressed properly is how teachers can make the most out of their assessment practice in a given educational context. In this respect, it has been widely recognized that language assessment literacy is an important aspect of teachers' professional knowledge (Coombe, Al-Hamly, & Troudi, 2009). Thus, the question which may be raised is: How can teachers' beliefs and knowledge of language assessment affect their assessment practices and attitudes within this process? To provide empirical evidence of what constitutes teachers' beliefs and

Z. Djoub (✉)

Abdelhamid Ibn Badis University of Mostaganem, Mostaganem, Algeria
e-mail: Algeriazdjoub@yahoo.fr

knowledge of language assessment and how these may affect their assessment approaches in English language teaching, a questionnaire was administered to English language teachers worldwide using a web based survey site called SurveyMonkey. This chapter will first introduce the concept of assessment literacy, its definitions and importance in language teaching and learning. Then, it will analyze and interpret the data obtained from the survey. Finally, a set of recommendations will be provided on how to train teachers to become more literate in assessment in ESL/EFL contexts.

2 Literature Review

2.1 Principles of Language Assessment

Language assessment refers to “the act of collecting information and making judgments on a language learner’s understanding of a language and his ability to use it” (Chapelle & Brindley, 2002, p. 267). It is, thus, an interpretation of the test taker’s ability to use some aspects of this language (Bachman & Palmer, 2010). It is worth noting that being able to use a language entails interacting with others, in a given setting, to create or interpret intended meanings within a particular discourse (Bachman & Palmer, 2010). Though assessment is undertaken for a variety of purposes (e.g. formative or summative assessment), the primary purpose remains to support learning which occurs when learners are, according to Cameron, Tate, Macnaughton and Politano (1998):

Thinking, problem-solving, constructing, transforming, investigating, creating, analyzing, making choices, organizing, deciding, explaining, talking and communicating, sharing, representing, predicting, interpreting, assessing, reflecting, taking responsibility, exploring, asking, answering, recording, gaining new knowledge, and applying that knowledge to new situations (p. 6).

Improving assessment practice remains the concern of several researchers who are attempting to find out how to make this process support learning. To do so, there has been a need to account for what constitutes good or “sound” language assessment, whose characteristics, as put forward by Stiggins (2007, cited in Coombe et al., 2009, p. 16), are:

- They arise from and serve clear purposes.
- They arise from and reflect clear and appropriate achievement targets.
- They rely on a proper assessment method (given the purpose and the target).
- They sample student achievement appropriately.
- They control for all relevant sources of bias and distortion.

So, in summary, effective or sound assessment is purposive and targets clear and relevant objectives which can contribute to both evaluating and developing learners’ language ability. Moreover, it uses appropriate assessment methods according

to the set objectives and communicates assessment results to all stakeholders who are involved in the process. Sound assessment also needs to develop valid and reliable grading by maintaining control over the variables that may distort its results. In addition to reliability and validity, Bachman and Palmer (1996, cited in Daalen, 1999) add other terms which refer to test usefulness such as authenticity, interactiveness, practicality and impact.

Additionally, sound assessment also entails involving learners in the assessment process and helping them move towards greater autonomy through introducing a wider variety of assessment methods, or what has been called alternative assessment. The latter is often connected to formative ‘assessment for learning’ (AFL) where assessment needs to “serve the purpose of promoting students’ learning...an assessment activity (is formative) if it provides information to be used as feedback by teachers and their students” (Black, Harrison, Lee, Marshall, & William, 2003, p. 10).

Following this description of the various concepts involved with assessment, a question which can be addressed is how teachers can effectively use assessment procedures in a given educational context. In fact, to avoid “the potential misuse or abuse of tests”, teachers need to be equipped with the knowledge and training required to practice effective assessment procedures (Taylor, 2009, p. 25). This has been referred to as assessment literacy.

3 Assessment Literacy: Definitions and Importance

It has been widely recognized that language assessment literacy is an important aspect of teachers’ professional knowledge (Coombe et al., 2009). Being literate in assessment means “having the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the results” (Inbar-Lourie, 2008, cited in Watanabe, 2011, p. 29). Accordingly, assessment literacy provides teachers with the knowledge and necessary tools to help them understand what they are assessing, how they need to assess it according to specific purposes, and what decisions they need to make in order to assess their learners effectively and maximize learning. According to Coombe et al. (2009), assessment literacy can be achieved through:

1. Understanding what a good assessment means while recognizing the different views about the nature of education which may lead to dissimilar approaches to assessment.
2. Providing professional development through both online training of teachers and through assessment workshops at all levels.
3. Being committed to significant change in educational practices.
4. Making assessment resources (especially online) available to language teachers to achieve successful professional development.

However, it needs to be maintained that assessment literacy does not only concern teachers, but also those involved in test development such as: policy makers, test developers, test administration, etc. Test takers or learners are also concerned “because they are the most important stakeholders and the greatest recipients of the benefits derived from the process and the product of language assessment” (Watanabe, 2011, p. 29). Therefore, learners need to understand the assessment process being implemented, its objectives and the criteria on which it is based. Indeed, Watanabe (2011) argues that this literacy is of crucial importance because, first, it helps relieve learners of their fear or anxiety towards the test they take, thereby avoiding negative washback. Second, because it allows them to get actively involved in the process of assessment and gain motivation from it.

4 Research Method

To investigate the effect of teachers’ assessment literacy on their assessment practices, a questionnaire was administered to English language teachers worldwide during April and May 2014. The results of this survey were collected online using a web based survey site called SurveyMonkey. The online survey was distributed through the author’s Twitter Network and English language teaching Networks such as Academia.Edu, Learner Autonomy Research Network, ESL International via LinkedIn, and the TESOL Arabia e-list.

This survey consisted of ten questions (four open questions, three semi-structured, and three structured questions). These questions were grouped into three sections. The first section attempted to find out about the participants’ teaching experience of English and their training related to assessment practices. The objective behind addressing these questions is to determine whether these teachers’ assessment practices are based on certain knowledge and skills developed from training programs, or are they mostly shaped by their teaching experience? The second section included five questions which aim to reveal their awareness of what constitutes sound assessment (what according to them is good language assessment? Are they aware of the importance of including alternative assessment or not?), as well as attempting to investigate how such knowledge is put into practice through examining their stated assessment objectives, approaches and procedures. In doing so, this section was looking to discover whether these teachers were encouraging their learners to learn and develop their language ability through assessment practices or just use assessment as an end in itself, i.e. limiting its scope to assigning grades, indicating learners’ success or failure at the end of a given term. This section’s aim was also to get an idea about test-takers’ assessment literacy through asking the participants whether they provide their learners with some knowledge about language assessment or not. In the last section of this survey, a space was devoted to the participants’ views regarding their institution’s supporting role and teachers’ needs, as far as assessment is concerned, in order to help enhance this process.

5 Results

The participants were forty-five EFL/ESL teachers from different schools and institutions worldwide. As their answers to the first question indicate, most of them (77.78 %) have more than 10 years of teaching experience, while only 17.78 % have experience of 5–10 years, and just 4.44 % have less than 5 years within this profession. So, the majority of the participants have considerable experience with teaching and thus with assessing their learners. When they were asked whether they had received training into how to assess language learners (Fig. 1), the majority (71.11 %) responded that they had while only 28.89 % said no.

However, the kind of training these teachers received in assessment processes may not guarantee their assessment literacy since not simply any kind of training may serve this aim. Therefore, those who received such training were asked about the kind of training they had received. Their answers demonstrate that most of them had taken courses on how to assess and test language learners as part of their Master’s programs. Still, a description of such courses and targeted objectives was not explicitly stated as the following examples show:

- In my master study we took a course named testing and its main objective how to test different English skills
- Assessment in ELT-MA Module
- During my Masters in TESL/TEFL studies, one course (3 credit hours) was in language testing but I don’t remember the name. It was about what and how to test

In addition to those MA courses, there were other teachers who gave just names of training programs without identifying their objective, duration nor the center or organization. For instance: *Evaluation and Assessment in General and Designing Good Tests, Continuous Assessment, Testing and Learning, Designing Language*

Fig. 1 Have participants had training in assessment?

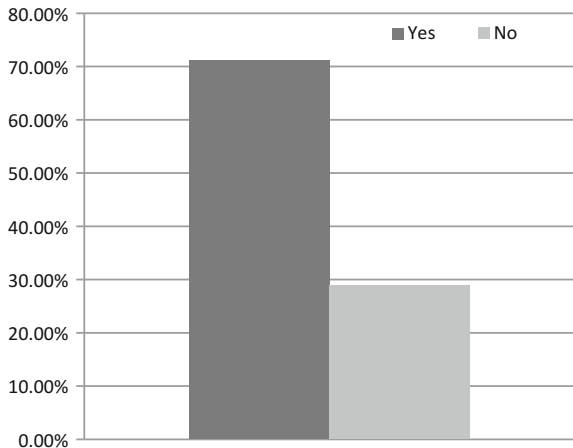
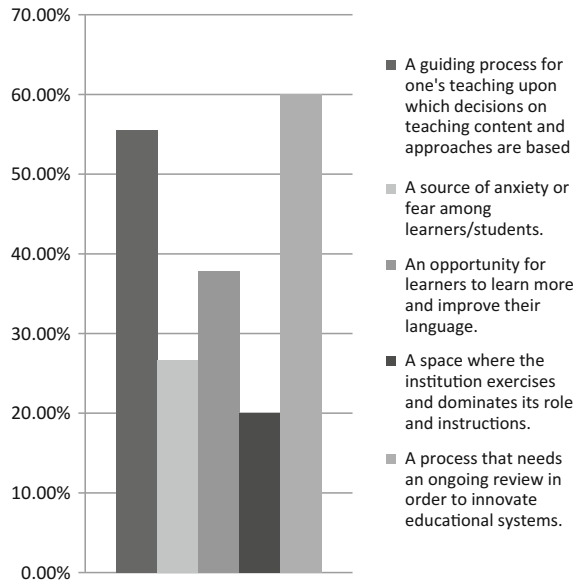


Fig. 2 The sample’s views concerning language assessment



Tests, etc. It is worth noting here, that there were two teachers who did not mention the name of the training program because they could not remember, as it was such a long time ago.

Furthermore, it is thought that teachers’ beliefs and understanding of teaching and learning may affect their actual assessment approach. Indeed, Richards and Rodgers (2001) affirmed that teachers possess assumptions about language and language learning, and that these provide the basis for a particular approach to language instruction. To find out about such an effect, the participants were asked about their views concerning assessment, i.e., how they view the assessment process in relation to their teaching practices. A set of options were provided here for selection, as well as a space for their additional or alternative answers. Figure 2 shows the data obtained for this question.

From this question we can see that assessment was regarded by 60 % of the teachers as a process which needs ongoing review in order to innovate educational systems. The second highest response (55.56 %) was from those considering it a guiding process for their teaching, including its content and approaches. 37.78 % of the questioned teachers believed that assessment is an opportunity for learners to learn more and improve their language, while only 26.67 % viewed it as a source of anxiety or fear among learners. Moreover, most respondents seemed to disagree with the idea that within this process there is much more space for institutional control and less or no room for the teacher or tester’s own voice, as only 20 % opted for this answer.

Moreover, there were other additional comments provided by 6.67 % of the teachers. Assessment was regarded as a way to determine students’ progress and

achievement as one teacher mentioned. Yet, another one pointed to the existing gap between what is taught and what is assessed stating: “When teaching at my university, an emphasis is placed on teaching students language skills, but many of the tests focus on grammar structures or filling in blanks with vocabulary from the book”. Likewise, one teacher referred to the dichotomy between what theoretical principles imply and actual practices indicate writing: “Assessment has the potential to be a great experience, but often it isn’t”. The need to train teachers in formative assessment forms was also raised by one teacher.

In the attempt to find out about their assessment literacy, the participants were also asked about their definition of sound assessment. This was done through an open-ended question, so that they could express themselves freely. The collected data reveals that, out of the questioned 45 teachers, 29 answered this question. Although, these answers differ from one teacher to another, they were grouped into the following categories, as shown in the following chart. There were criteria for sound assessment which pertain to the assessment tasks themselves, i.e., content; and those related to the process of assessment, i.e., the kind of assessment tools used, the way the assessment task was administered, feedback provided, etc. (Table 1).

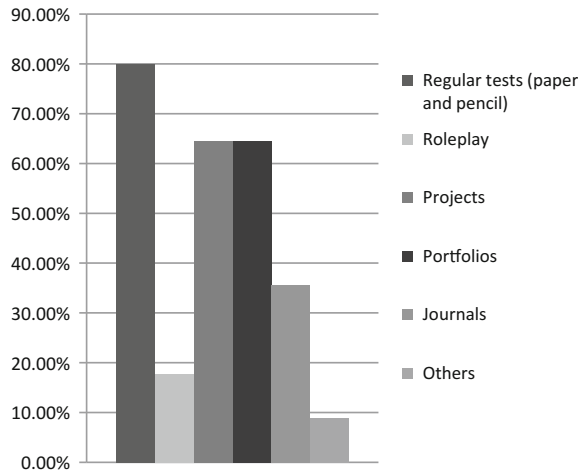
It is worth noting, however, that there were few answers provided for each category since the teachers’ definition of sound assessment covered just one or two criteria. Indeed, there were some teachers who put emphasis on what is known as the assessment principles: reliability, validity, authenticity, practicality and wash-back. Only two teachers stated these principles together, while others mentioned one or two of them (three referred just to authenticity, two others to reliability, another two teachers to validity and reliability, and one teacher wrote authenticity and reliability). For other teachers, relating assessment to teaching and covering the different language skills was considered a common feature of assessment content. Indeed, there were four teachers who maintained that sound assessment should assess directly what was taught in class, except when it is the case of language competency tests such as IELTS. On the other hand, four others pointed to the need to provide learners with opportunities to use the different language skills by assessing them. One of these teachers added that a good test should not only cover the language structure (language usage), but it should also assess learners’ productive and receptive skills (language use).

Moreover, for some teachers’ good assessment practices are attributed entirely to the process of its implementation. Clarifying the assessment criteria, objectives, and explaining the how and why were highlighted by four teachers. Another teacher referred to the importance of asking well-phrased questions. Three others saw that feedback needs to be clear for both teachers and learners. Other test criteria were also mentioned by teachers in relation to the assessment process, among them was the use of different assessment methods to gather ample data about learners’ performance, as noted by four teachers. It is worth noting, however, that assessment should not be conducted merely for the sake of collecting such data, it also needs to provide learners with the opportunity to learn and improve—a sentiment echoed by just two other teachers. Similarly, in spite of the increasing tendency towards

Table 1 Criteria of sound assessment according to the sample

	Criteria of sound assessment	Some examples
Content of assessment	Valid, reliable, and fair	Fair, reliable and tests student abilities
		Valid and reliable assessments should all be extensions of the curriculum
	Authentic, i.e., interactive, testing learner’s communicative competence	Testing the ability to apply the learned concepts into real life setting
		A good language assessment in a language classroom measures how well the student has internalized the language system and how competent the student is in using the language in social contexts
	Directly related to what was taught	Good language assessment asks students to produce language based on what they have been practicing in class
It should be directly related to what the students are learning at the time		
Assessment of different skills	Assessment that includes all four skills with clear criteria identifying each level for consistency	
	Weekly reviews or short quizzes that examine different skills	
Process of assessment	Clear assessment process including the objective and criteria	Students know what, why and how to do well
		Assessment where both teacher and learners understand and can use the goals, processes and outcomes
	Clear feedback provider for both teachers and learners	The one that gives the teachers clear feedback on learners’ performance
		Criterion-based fairness in marking feedback to students on how to improve their performance
	Using a variety of methods of information collection	A mixture of standard tests and ongoing monitoring
One that provides ample opportunities for the students to exhibit their proficiency		
A source of learning and progressing	This process needs to track students’ learning as well as contribute to their learning	
	Where the learner learns not just being tested. He should have the opportunity for feedback about his performance	
Prompting reflection	Summative assessment: reflection paper is good for the student to be aware of his own progress, checklists are good to the teacher to collect qualitative data of the student	
	Sound assessments are based on sound and achievable objectives that have ongoing formative assessments based on reflective practices that have been incorporated into the lesson plan, and utilize the input of both students and teachers	

Fig. 3 The sample’s use of alternative assessment tools



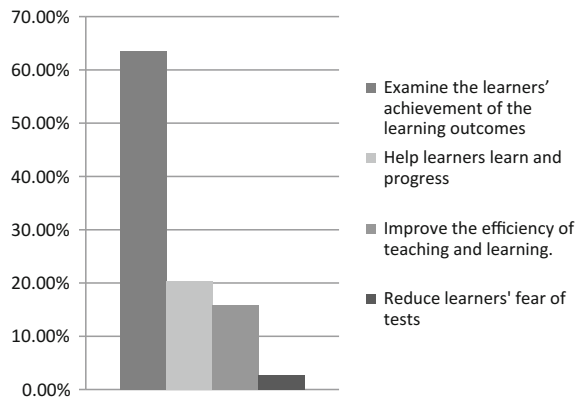
integrating alternative assessment forms in language teaching and involving learners in assessing themselves and their peers, only two teachers conceived this as important in their answers.

Question six was addressed to learn whether these teachers use any kind of alternative assessment approaches apart from final exams. Regular tests (paper and pencil), role play, projects, portfolios and journals were the choices provided within this question, besides asking the teachers to indicate other tools that they used to assess their learners which were not among the choices. As Fig. 3 shows, the majority of the sample (80 %) used regular pen and paper tests. Projects and portfolios were used by the same number of teachers (64.44 %); around half this number (35.56 %) assessed their learners through journal writing and only 18 % of the participants selected role play. Other assessment modes were also provided such as Quizzes (by four teachers), checklists and reflection (three teachers), presentations (two teachers), tests on iPad (one teacher), observations (one teacher), and oral interviews (one teacher).

In addition to identifying their assessment approaches, there is also a need to uncover their intention behind implementing them in order to find out about their assessment objectives and whether these match their assessment practices. To achieve this aim, an open question (N^o7) was addressed to the participants. From the collected responses, three main objectives of assessment were identified by the sample as follows: to examine the learners’ achievement of the learning outcomes, to help them learn from the teacher’s feedback and progress, to improve the efficiency of teaching and learning. The highest rate was recorded for the first objective as Fig. 4 demonstrates.

Concerning the kind of support these teachers provide their learners during their assessment process, it was found that providing constructive feedback on learners’ performance was selected by 88.89 % of the sample. Similarly, the majority (84.44 %) agreed with the need to familiarize their learners with the test format and

Fig. 4 The teachers' intentions of integrating formative assessment



clarify from the outset the assessment criteria and objective. Reviewing lessons covered by the test was regarded essential by 60 %. As Fig. 5 shows, other forms of support were provided for learners by some teachers, such as:

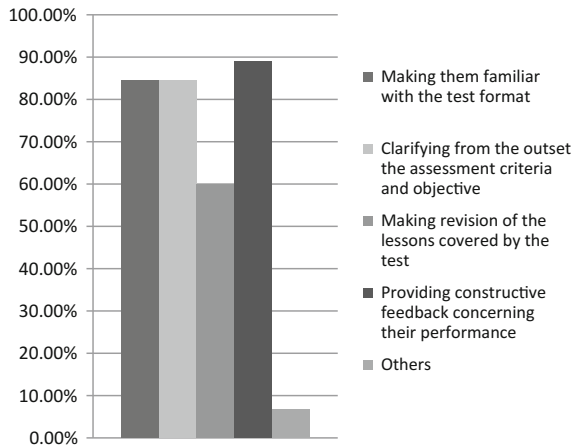
- “Handouts of best examples of students’ answers”.
- “Discussions”.
- “Praise the effort made. Advise with sensitivity and encouragement”.
- “Streamed tables focusing on different skills, so that learners can focus on the one skill they really need to improve.
- “Self-reflection on test performance and reasons why they achieved/didn’t achieve their goals—Goal-setting—Making study plans—Allocating some class time to study skills”.

As far as teacher’s autonomy in assessment is concerned, the participants were requested to mention to what extent their assessment practices are dictated by the institutions/universities/schools they belong to. It was found that 81.39 % of them stated that these play a dominating role within this process and thus there is no space for teacher freedom. Whereas, only 11.62 % replied that there is some freedom left for them mainly within formative assessment as the following examples show:

- “Other than final assessments we have a lot of flexibility with assessments, bearing in mind they are within the framework of the institution’s dictated learning outcomes”.
- “In my institution, the middle and end of semester practices are dictated by the institution but the assessment practices throughout the semester are in the hands of the individual teacher”.
- “Final assessments are quite regulated, but within the course it’s fairly liberal”.

Additionally, 6.97 % of these participants mentioned that this depends on the institution/school’s objective behind their assessment practices, as one teacher said: “It depends on the institution’s curriculum. Some are strict on the type of

Fig. 5 The kind of support provided for their learners



assessment to be done and others are interested only in scores that can be assigned to students”. Finally, the participants indicated the kind of support they need to make their assessment more effective. These ranged from providing teacher training courses and programs on how to assess language learning, to having at their disposal the necessary materials, time, and certain autonomy to achieve this objective. Figure 6 illustrates these findings.

Overwhelmingly, the kind of support which teachers looked for was the provision of teacher’s training and professional development courses in relation to language assessment, as stated by 60.46 %. These would cover the know-how of the assessment process, including understanding its purpose, criteria of assessment, what should be assessed, its approaches (why, how and when to implement them), in addition to how feedback should be communicated to learners. For other teachers (18.60 %) mentoring and collaborating among teachers and experts in the assessment process is crucial to help them gain more feedback about their practices and get involved in reviewing and improving them.

On the other hand, only 16.27 % of the participants mentioned that teachers should be given a certain degree of autonomy to make decisions regarding the assessment process, including selecting the type of assessment that matches their learners’ learning needs and interests, as one teacher stated:

creating the type of assessment that go along with what students have been doing in class. Classroom tests are not necessarily indicative of what students can do with the language. Assessment should allow a variety of usage to ensure that language is used in its appropriate sociocultural context.

Moreover, just 11.62 % of participants mentioned the materials needed within assessment (see Fig. 6) and the same rate was also noted for those who referred to the importance of allocating enough time for this process, as these teachers’ statements indicate:

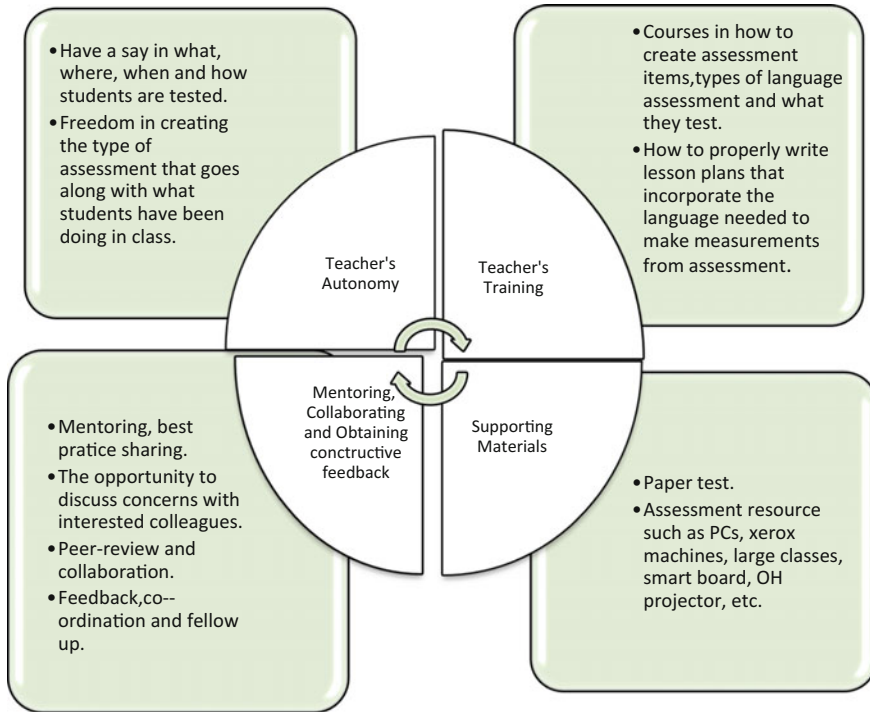


Fig. 6 The sample's needs within the assessment process

- “Time, time, time to create assessments, time to analyze the results, and time to make the necessary adjustments to the assessments”.
- “...enough time to focus on the main skills”.

6 Discussion

The survey results show that most of the teachers (45 teachers) have experience in teaching English which goes beyond 10 years. This is likely to contribute to their conceptions and attitude construction regarding the assessment process. During that teaching experience most of them had never been trained in language assessment. Instead, they received courses as part of their MA program. However, a clear distinction needs to be made here between receiving courses as part of an education where emphasis is put upon the development of knowledge and moral values required in all walks of life, and training which emphasizes knowledge, skills and behavior patterns required to perform a particular job (Rao, 2004). Thus, teacher

training is mostly concerned with the practice and skill of methodologies, not with the knowledge of background theories.

As a matter of fact, providing teachers with knowledge about assessment would not suffice to make it effective without equipping them with the necessary skills and strategies that can help them make decisions over what assessment tools to integrate into their teaching. They also need the ability to evaluate learners' needs and institutions' intended outcomes, as well as put them into practice along with feedback provision to enhance learner performance. In addition, since there are some teachers who do not remember even the type, name, objective or time of training they received, there is a need for continuous professional development (CPD) that can help teachers innovate and commit themselves to change their teaching and assessment practices.

It can be inferred that, overall, these participants must possess some knowledge and assumptions about language assessment as a result of attending such trainings and MA courses and their teaching experience. But, they may not have developed the necessary practical skills to assess their language learners, as their knowledge of assessment is not updated regularly since, as they stated, they have never gone through professional development courses. Some teachers who stated that they never received any training or courses into assessment displayed awareness of the importance and need for such training, and they also acknowledged that it is unacceptable for language instructors not to be trained in such a process whose results can be critical for their learners' future and decision makers' intentions and plans. This may raise issues related to the effectiveness of their assessment practices and trustworthiness of the assessment process. In fact, their dissatisfaction of its outcomes and effects was revealed when they highlighted the need for continuous review and reform to the process, in addition to pointing to the existing gap between theoretical principles on language assessment and what actual practices reflect.

Nevertheless, their views that assessment is not a source of anxiety or fear for learners, nor a space for institutions to exercise complete control, may imply that these teachers are aware that this process should help learners achieve their learning potential through helping them "appreciate challenge and shake off the fear of failure" (Clegg & Bryan, 2006, p. 218). They also seem to realize that they need flexibility to adjust their assessment according to their learners' needs. Still, the findings revealed that they were unaware of the means to achieve such a purpose. Using alternative assessment approaches was not regarded as a major feature of sound assessment by most of them, who considered reliability and validity as the most crucial criteria for any assessment. Focusing entirely on these criteria, however, may not encourage learners' creativity and language use, as Gipps (2006) maintains, these concepts "are now seen to have limited usefulness because of the way that they assume that all assessments are unidimensional, and that they are steps towards producing a single 'true score' to summarise the educational achievement level of a student" (cited in Murphy, 2006, p. 43).

In fact, these teachers were unaware of the potential of such assessment approaches in helping learners to learn and progress, and improve the teaching/learning process on the basis of assessment feedback. With participant's

stated objectives for integrating them was limited entirely to the examination of learners' achievement of the learning outcomes, whereas no teachers indicated any intention of involving their learners in the assessment process through providing them with the opportunity to reflect on their learning process or monitor their progress and make the necessary decisions to improve it. Integrating such forms is not a guarantee of subsequent AFL opportunities, as they may simply be used for grading purposes (Murphy, 2006). So, making the right selection of alternative assessment approaches does not suffice without considering the purpose behind implementing them, which remains crucial to determining their usefulness.

Moreover, since the washback effect can engender either learners' frustration or motivation to learn and improve, the affective aspect of assessment also needs to be catered for as part of the teacher's support within this process. Though most of the participants did not consider assessment as a source of anxiety and fear for learners, no kind of psychological support was provided by them to help their learners overcome their fear of exam taking. This could imply that they are not aware that "being assessed is undoubtedly an emotional business" which is likely to be remembered by learners (Clegg & Bryan, 2006, p. 218).

Finally, most of the teachers' answers refer to the dominating role of their institutions and schools over the entire assessment process, a role deemed not necessarily helpful by most teachers. Rather, teacher training into language assessment was conceived as a prerequisite for the effectiveness of this process. By the same token, the participants were calling for their institutions/schools' support to provide them with the necessary professional development courses which can help them achieve their "continual, intellectual, experiential, and attitudinal growth" in regard to their language assessment process (Richards, 1989, p. 4).

7 Conclusions, Implications and Recommendations

From this study, it can be concluded that the participants have not received the necessary training into educational assessment in order to maximize its effectiveness. Nor have they gone through professional development courses during their career. Their beliefs and views concerning what assessment means for them in general and what constitutes sound assessment in particular reflect their lack of assessment literacy. And, hence, they seem mostly unaware of the need "to recognize that assessment procedures can and should contribute to student learning as well as measure it", besides assessing learning "in a wide variety of ways, and indeed be reported in ways that recognize diversity rather than mask it" (Murphy, 2006, p. 44).

Though they mentioned their implementation of some alternative assessment forms, their use remains for the sake of grading rather than for learning. Thus, traditional assessment practices are still prevailing where the focus is entirely on the learning outcome instead of the process. This might be due to their lack of assessment literacy and thus their narrowed vision of what learner assessment

should focus on. Another possible reason, however, would be that using alternative assessment for grading is part of the institution or school's policy because as stated previously most of the participants have no control over their assessment practices.

As with any study, there are some limitations that affect the generalizability of these results. This study needs to be carried out on larger numbers of participants to improve the validity of the findings. Also, other data collection tools can be used such as interviews and observations of the English language teachers' assessment practices. Another limitation is that the study did not examine the teachers' assessment literacy in relation to a given language skill, i.e., writing, speaking, etc. Future studies would benefit from gaining more empirical evidence to investigate teachers' assessment literacy regarding a particular language area.

To help learners overcome their fears and anxiety of exam taking, it is crucial for teachers to understand learner psychology in relation to language assessment. To this end, researchers need to examine this relationship and provide evidence regarding psychological support for learners (Watanabe, 2011). There are also other questions that still need answers. For instance, how can we train teachers in assessment literacy and keep them up to date with the latest innovative approaches to assessment so that they can cope with the changing and challenging demands of society?

In fact, this study revealed that the participants' teaching experience alone has not allowed them to learn about how language assessment needs to be conducted more effectively, and they have therefore failed to develop their assessment literacy and share it with their learners. Their lack of assessment literacy is reflected through their views and practices in assessing their learners. Indeed, their views demonstrate their lack of awareness of what constitutes sound assessment, and their assessment practices act as instruments of justification, measurements and limitation rather than tools to enhance and enable self-regulated learning and judgments (Bryan & Clegg, 2006).

Therefore, teacher training into assessment literacy is advocated here as part of initial teacher education and should be supported beyond this stage through continuous professional development courses. This training needs to support teachers in developing "multi-dimensional awareness" and "the ability to apply this awareness to their actual contexts of teaching" (Tomlinson, 2003, p. 2). To do so, it first needs to sensitize them to the importance of being literate in assessment and the benefits derived from such a process over both learning and teaching, besides introducing them to the major assessment principles in the EFL/ESL context. Second, it needs to equip them with the necessary strategies and techniques of language assessment which are related to particular contexts. Finally, this training needs to encourage teachers' self-evaluation over their assessment practices and provide guidance into the process. This can be done by making them aware of procedures such as observations, checklists, questionnaires, etc., and showing them how to use them continuously and effectively to gain more insights into these practices, thus linking theoretical concepts with experience.

Appendix 1: Questionnaire

1) Would you please indicate your teaching experience:

- Less than 5 years
- 5 to 10 years
- More than 10 years

2) Have you ever been trained into how to assess you language

learners? YES NO

3) If your answer is YES, would you please indicate the kind of training you received.

.....

.....

.....

.....

.....

4) How do you consider your assessment practices?

- A guiding process for your teaching upon which decisions on
teaching contents and approaches are based.
- A source of anxiety or fear among learners/students.
- An opportunity for learners to learn more and improve.
- A space where the institution exercises and dominates its
role
and instruction.
- A process that needs an ongoing review in order to innovate
educational systems.
- Others?

5) What is your definition of sound assessment?

.....
.....
.....
.....
.....

6) What kind of alternative assessment do you integrate in your assessment of your learners?

- Regular tests (paper and pencil)
- Role play
- Projects
- Portfolios
- Journals
- Peer-assessment
- Others?

7) What is your main objective(s) from using such alternative assessment approach(es)

.....
.....
.....

8) To support your learners along the assessment process, you:

- Make them familiar with the test/exam format.
- Clarify from the outset the assessment objective and criteria.
- Make revision of the lessons covered by the test/exam.
- Provide constructive feedback concerning their performance.
- Others?

- 9) To what extent your assessment practices are dictated by the institutions/ universities /schools you belong to?
.....

- 10) What kind of support do you need to make from your assessment more effective?
.....
.....
.....
.....
.....

References

Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.

Cameron, C., Tate, B., Macnaughton, D., & Politano, C. (1998). *Recognition without rewards*. Winnipeg, MB: Peguis Publishers.

Chapelle, C., & Brindley, G. (2002). Assessment. In N. Schmitt (Ed.), *An introduction to applied linguistics* (pp. 268–288). London: Arnold.

Clegg, K., & Bryan, C. (2006). Reflections, rationales and realities. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 216–227). New York: Routledge.

Coombe, C., Al-Hamly, M., & Troudi, S. (2009). Foreign and second language teacher assessment literacy: Issues, challenges and recommendations. *Research Notes*, 38, 14–18.

Daalen, M. V. (1999). Test usefulness in alternative assessment. *Dialog on language Instruction*, 13(1 & 2), 1–2.

Gipps, G. (2006). How assessment frames student learning. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23–36). New York: Routledge.

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25, 385–402.

Murphy, R. (2006). Evaluating new priorities for assessment in higher education. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 37–47). New York: Routledge.

Rao, R. (2004). *Methods of teacher training*. New Delhi: Discovery Publishing House.

Richards, J. C. (1989). *Beyond training: Approaches to teacher education in language teaching*. Retrieved from <http://markandrews.edublogs.org/files/2010/05/training-v-development-11.pdf>.

Richards, J. C., & Rodgers, T. (2001). *Approaches and methods in language teaching*. Cambridge: Cambridge University Press.

- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.
- Tomlinson, B. (2003). *Developing materials to develop yourself: Humanising language teaching*. Retrieved from <http://www.hltmag.co.uk/jul03/mart1.htm>.
- Watanabe, Y. (2011). Teaching a course in assessment literacy to test takers: Its rationale, procedure, content and effectiveness. *Research Notes*, 46, 29–34.

Kurdish Tertiary EFL Teachers' Assessment Literacy in Alternative Assessments and Its Influence on the Ethicality of Their Assessments

Dler Abdullah Ismael

Abstract Assessment has witnessed a great transformation in the past decade or so because it is considered an integral part of the instructional process. This transformation has recognized the impact of such factors as ethicality and professionalism on assessment and test performance. This chapter aims to critically analyse the effect of language assessment literacy in alternative assessment on the ethicality of English language assessment practices among Kurdish tertiary EFL teachers. After discussing concepts related to alternative assessment, the chapter provides a description of the current assessment situation in some Kurdistan Region English departments and then it discusses the level of Kurdish EFL teachers' assessment literacy in alternative assessment and how that influences their assessment practices.

Keywords Assessment literacy • Alternative assessment • EFL

1 Introduction

Generally speaking, nowadays, teaching in many parts of the world is in the midst of a great transformation because of teachers' increased expectations for their students to achieve high standards of performance and for their learning to continually improve (Hargreaves, 2000). Probably one of the most important aspects of the teaching process that has witnessed this transformation is assessment. At least 250 studies have reported that the use of assessment to promote learning in the classroom has improved student achievement (Earl & Katz, 2006). Assessment is, according to Coombe, Troudi, and Al-Hamly (2012), "an integral part of the teaching-learning process" (p. 20). Therefore, the evaluation of students' progress is considered a major part of a teacher's job (Brumfit & Johnson, 1979). However, until recently, assessment processes were based only on traditional standardised 'pen and paper' tests (Ataç, 2012), which I and possibly many TESOL practitioners

D.A. Ismael (✉)

University of Sulaimani, Sulaymaniyah Region of Kurdistan, Iraq
e-mail: daii201@exeter.ac.uk; dler276@gmail.com

believe are inadequate. Fortunately, the field of language assessment has now widened to recognize various factors other than language ability that impact test performance, and it now takes into account ethics and professionalism issues. Presently, language testers investigate the use of language assessment, test developers' and users' ethical responsibilities, language assessment fairness and the impact and consequences of assessment use on instructional practice and societal values (Bachman, 2007). Thus, as engines of reform, tests can influence students, curricula, and the whole educational systems (Winke, 2011).

Despite the significance of assessment and testing, in this study I will demonstrate the low level of Kurdish tertiary EFL teachers' assessment literacy in alternative assessments in my professional context as an EFL teacher in some English departments of the universities of the Kurdistan Region, Iraq. Then, I will show the influence of assessment literacy on assessment ethicality in that context. Here, as an introduction, it would be reasonable to present a brief description of the present situation of English language assessment in those English departments. In these departments, most English language assessment procedures are still based on the traditional standardised 'pen and paper' tests alone. Thus, these English departments have testing systems rather than assessment systems as there is a series of monthly exams (out of nearly 30 marks) and a final year exam (out of 60 marks) while daily alternative assessments are allocated approximately 10 marks. Generally, this reveals that those departments have not embraced, first, the current widened scope of language assessment that pays attention to various factors that impact test performance, ethics and professionalism (Bachman, 2007), secondly, how assessment results have effective implications for improving overall teaching and learning qualities (Wolf, Herman, Bachman, Bailey, & Griffin, 2008), and third, the understanding of the essentiality of alternative assessments.

To critically analyse and understand the effect of language assessment literacy in alternative assessments on the ethicality of English language assessment, we should first discuss the conceptualization of such recently developed alternative assessments. This is because assessment literacy in alternative assessments is possibly much more influential in terms of assessment ethicality and validity.

2 Conceptualization of Alternative Assessments

Assessment and testing is a universal facet of social life. Throughout history, people were tested to prove their abilities, and they continue to be tested for different reasons today. However, currently tests have proliferated rapidly (McNamara, 2000, p. 3) resulting in the appearance of a vast array of different test types and procedures.

Among those recently-appeared assessment types are alternative assessments. I believe that alternative assessment practices are the most effective in the students' learning process. Alternative assessments are multiple measures that show students' learning, achievement, motivation and attitudes in instructional classroom activities

while performing real-life tasks. Many terms emerge for this sort of assessment such as alternative assessment, authentic assessment, performance assessment or portfolio assessment (Ataç, 2012). In such assessments, instruction and assessment are intrinsically integrated (Ripley, 2012). For example, portfolios are collections of students' work accumulated over time-throughout the learning process and used to assess their competencies in a given standard or objective (Ripley, 2012). Another impressive kind of alternative assessment is self-assessment, which is regarded as an integral component of language learning. Basically, no self-assessment is thought to mean no self-awareness, or knowledge of one's level, strengths, weaknesses and preferred way of learning (Cummins & Davesne, 2009). Additionally, self-assessment encourages learner awareness and hence confidence, an understanding of evaluation, and the ability to see errors as helpful (Alderson & Banerjee, 2002). There are many alternative assessment methods such as conferences, debates, demonstrations, diaries/journals, dramatizations, exhibitions, games, observations, peer-assessment, projects, story retelling and think-alouds, alongside almost as many ways of recording them, for example, anecdotal records, checklists, learner profiles, progress cards, questionnaires and rating scales (Tsagari, 2004).

There are two basic purposes of language assessment, namely, 'assessment for learning' and 'assessment as learning' that, I believe, are related to alternative assessments. Assessment for learning is to gather information to modify learning activities, to target instruction and resources, and to give feedback to students. Assessment as learning is used to develop and support metacognition for students by focusing on their roles as critical connectors between assessment and learning. This can be done through self reflection and critical analysis of students own learning. These two purposes are not to be confused with assessment of learning which is a summative measure used to confirm students' knowledge and abilities (Earl & Katz, 2006).

Following this brief definition and conceptualization of alternative methods of assessment, let us elucidate the theoretical background of these assessments.

3 Theoretical Background of Alternative Assessments

First of all, to understand the theoretical basis of alternative assessments, it is helpful to explain the two contrasting assessment approaches that constitute two major tenets of language assessment: traditional and alternative. I will base this contrast only on positivism and interpretivism without presenting other relevant theories such as cognitivism, socio-cultural approach, constructivism, etc. Basically, traditional and alternative assessments are based on two different underpinning paradigmatic assumptions. It is said that traditional language assessment follows positivism, which sees language ability as parallel to objects in the physical world. Whereas, alternative assessment is informed by interpretivism, which considers language proficiency as part of the social world. Acknowledging the latter, language knowledge can be pursued in ways other than the scientific

method. Another distinguishing characteristic is related to the assessor's relationship with the assessed. Positivist testers prize objectivity, with the assessor remaining neutral and disinterested in the object of inquiry. But interpretivist testers find it impossible to separate facts from the subjectivity of values in relation to people and the social world. Thus, the abilities that an assessor tries to assess are seen as socially constructed not as external and independent of the assessor (Lynch & Shaw, 2005).

With regard to abilities, the strength of alternative assessment is its potential to measure language proficiency. In this respect, Shohamy (1998, as cited in Piggini, 2012) urges that the definition of language proficiency should be critically examined because decisions based on single tests can change test-takers' lives. Therefore, assessing both academic and social English language skills may go some way towards appeasing Shohamy while providing a clearer picture of students' English proficiency (Stephenson, Johnson, Jorgensen, & Young, 2003). I believe that the most suitable assessment type required for more comprehensive language proficiency is alternative assessment, which can occur at various points in time and in various ways both inside and outside classrooms. This belief is also congruent with the recommendation of the TESOL International Association (2010) to use various performance-based assessment tools and techniques in assessment.

As an EFL teacher, I believe that when Coombe et al. (2012, p. 20) made the statement that assessment should be "an integral part of the teaching-learning process" they are advocating the use of a variety of alternative assessment techniques in the classroom. These alternative assessment practices are integral components of teaching and learning because they provide more informative measures, have more of a focus on strengths and progress and treat each learner as a unique person. This, in turn, encourages on-going assessment in a culture-inclusive environment where several perspectives are possible. The result is the improvement and guidance of learning and collaborative learning (Ataç, 2012). Alternative assessments are typically less formal, gathered over a period of time, formative in function, often low-stakes in terms of consequences, and provide positive washback effects. Further advantages include the fact that they provide easily understood data, are more integrative than traditional tests and are easily integrated into the classroom activities (Alderson & Banerjee, 2002). Thus, since a major consequence of the use of a test is its impact on instruction or washback (Bachman, 2005), alternative assessment results could be used to improve instruction (Tsagari, 2004). This is facilitated by using tests as techniques for gathering systematic evidence to base instructional decisions on. This, in turn, can effectively enhance educational processes (Brumfit & Johnson, 1979). All this will hopefully promote learning and enhance educational access and equity (McNamara, 1998, as cited in Alderson & Banerjee, 2002). That is why I do not think that classroom achievement tests that follow traditional approaches to testing (Brumfit & Johnson, 1979), and that are possibly informed by positivism, can fulfil the above-mentioned recommendation regarding the integration of instruction and assessment.

Having outlined the two contrasting aspects of language assessment and the theoretical background of alternative assessments, it will be elucidating to next examine the interrelation between alternative assessments and critical language testing (CLT), one of its underpinning theories.

4 Alternative Assessments and Critical Language Testing

Generally, it is realized that testing has an impact on test-takers, stakeholders and society (Hamp-Lyons, 2001). However, testing as a criterion for graduation affects students differently from testing designed for informing instructional decisions (Pitoniak et al., 2009). Focusing on these effects, CLT begins with the assumption that language testing is not neutral, but rather it is a product of cultural, social, political and educational agendas that affect teachers' and learners' lives. Hence, test-takers can be thought of as political subjects in political contexts. CLT also explores whose agendas are behind tests, what visions of society tests presuppose, whose knowledge tests are based on and whether it is negotiable. CLT then examines test meanings and scores and their openness to interpretations (Dai Quang, 2007).

To follow CLT principles, testers should always consider the basic principles of testing that are of great significance in classroom assessment; these include reliability, validity, reference-points, record-keeping (Earl & Katz, 2006), as well as fairness and washback (The EALTA Executive Committee, 2006). This is in order to ensure teachers' inferences about students' learning are credible, fair and free from bias (Earl & Katz, 2006). Regarding bias, I think it is more important than the other principles of language testing because it directly influences the ethicality of language assessment. Bias is the systematic unequal impact on specific subgroups in assessments and includes—whether they are fully included, how they are treated, how they perform, how their performance is scored and how their test scores are used (Bachman, 2005). Taking all these assessment issues into consideration is to achieve ethical language assessment.

Concerning ethicality and traditional testing, my understanding is that ethical assessment cannot be adequately achieved by traditional testing alone. Traditional testing determines the quality of testing by adopting accepted models and procedures for testing accuracy; whereas, little attention is paid to the test uses and their importance in test-takers' lives and their status in society. By listening to the opinions of test-takers it is possible to get an idea of the power of tests (Shohamy, 2001). In this regard, traditional testing does not focus much on the testing experience and the meanings and feelings that tests create in the test-takers' minds (Shohamy, 2001). Therefore, I believe that ethicality of traditional language testing is possibly violated by ignoring the impact tests bring about to students' lives. Their major purpose and focus is not on lives but on creating quality tests that can accurately measure the knowledge; as a professional field of endeavour, it is limited

in scope due to its strict rules and procedures for appropriate practices (Shohamy, 2001). Furthermore, traditional testing techniques such as multiple-choice, fill-in-the blanks, matching, etc., are often incompatible with the current ESL/EFL classroom practices (Tsagari, 2004). I believe that these issues represent a major critical concern for CLT.

Personally, I advocate CLT in attempting to challenge the psychometrics of traditional language testing and to support interpretive approaches to language assessment (Dai Quang, 2007). This signifies a paradigmatic shift in which many new criteria for understanding validity are considered such as consequential, systemic and interpretive validity (Dai Quang, 2007); most importantly, I believe, is the consideration of ethical issues that are associated with the effects (consequential validity) of tests.

CLT supports alternative assessments like portfolios for developing more democratic testing methods in which test-takers and local bodies are more active (Dai Quang, 2007). Regarding the democratization of testing and assessment, Moss (as cited in Lynch & Shaw, 2005) argues for an interpretivist hermeneutic approach that encompasses many alternative assessment essential qualities as it acknowledges the effectiveness of the context of assessment, and formulates validity as a consensus through dialogues between stakeholders: teachers, students and parents, not between disinterested external experts. Moss also challenges the generalizability principle of traditional testing in which educators should generalize from one performance to all similar performances and contexts (Lynch & Shaw, 2005). This disapproval of traditional testing demonstrates that alternative assessments, informed by interpretivism, provide what Shohamy calls democratic testing, in which I believe there exists a high degree of assessment ethicality. Shohamy's democratic model of testing gives some guidelines for making testing more democratic, with the aim of limiting and controlling the power inherent in the use of some tests. By adopting critical pedagogy, institutions may achieve an ethical discourse, which helps students to question and challenge dominating beliefs and practices (Riasati & Mollaei, 2012), including those beliefs about language testing and assessment. To foster such a discourse of ethicality, Shohamy's principles of CLT regarding questioning the informing values, agendas, goals, needs, purposes and actual uses of tests should be taken into account in language assessment (Bachman, 2005). This may be possible by adopting an anti-authoritarian, dialogical and interactive approach, such as Freire advocates, to examine relational power issues for students (Chandella & Troudi, 2013). Power relations usually have elements of domination, exploitation and subjection resulting from the immobilization and prevention of any reversibility of movement (Lynch & Shaw, 2005). However, concerning the democratization of assessment, Foucault adds that substituting alternative assessments for traditional testing does not necessarily change the power relations, for example, students do not necessarily control their portfolio specifications. Nevertheless, Foucault considers ethics as the practice of freedom; this can be found in portfolio assessment which focuses on validity and ethics affected by power relations. In a portfolio, the freedom lies in the students' abilities to shape the portfolio process and form (Lynch & Shaw, 2005).

Personally, I believe in the democratic model of language assessment, which is epitomised by alternative assessments such as portfolios (Foucault, as cited in Lynch & Shaw, 2005). Also, I think that democratic assessment must be more widely championed, because as Howe (1994, as cited in Lynch, 1997) mentions the democratic model is the most moral approach, providing a viable alternative that includes voices which have historically been banned from negotiating educational issues (Lynch, 1997). This is in keeping with attempts through critical theory, critical pedagogy and critical research to enlighten, empower and emancipate people from oppression (Brown & Jones, 2001, pp. 101–102). This oppression sometimes constrains social and educational practice and produces results contrary to those desired by participants (Reason & Bradbury, 2001, p. 95). To reduce negative power relations and unfairness in language assessment, the knowledge of and training in alternative assessments and CLT principles is fundamental. Hence, we should pay special attention to teacher assessment literacy.

5 Alternative Assessments and Language Assessment Literacy

From the beginning, ESL/EFL teachers should know why language assessment literacy is important. They should know that, first, assessment is a widespread characteristic of educational systems, second, it enables teachers to share their classroom results with other teachers in order to develop a community of teachers that fosters learning, third, and very importantly, it is suggested that assessment literacy is an influential aspect of teachers' professional development (Newfields, 2006). Moreover, one of the eleven TESOL standards of the TESOL International Association deals with ESL/EFL assessment, which is regarded as an undertaking requiring solid pedagogical knowledge (Thibeault, Kuhlman, & Day, 2010). However, unfortunately, almost all teachers carry out assessments without having learned the principles of sound assessment (Coombe et al., 2012, p. 20).

Assessment literacy for language teachers includes having:

1. The ability to employ a variety of assessment measures with minimal bias,
2. The ability to construct, administer and score tests,
3. The ability to evaluate the reliability, item difficulty, item facility and content validity of tests,
4. The ability to statistically determine the cut-off point of examination,
5. The ability to appropriately intervene when students engage in unethical behaviours during tests and
6. Competence in communicating assessment results to parents, peers and students (Newfields, 2006).

These assessment literacy requirements are based on the 1990 Standards for Teacher Competence in the Educational Assessment of Students that are published

jointly by the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (Newfields, 2006). ESL/EFL teachers should become aware of these assessment literacy points so as to be capable of assessing ESL/EFL standards related to the goals of using English in socially and culturally appropriate ways to communicate in social settings and to achieve academically in all content areas (Short, 2000).

Because of its significance for ESL/EFL teachers, there are some recommendations that might enhance assessment literacy as follows:

1. Beta-testing the test and revising it after examining the testee responses,
2. Paying special attention when grading a test to the cut-off points and questions that over 90 % of test-takers answered correctly or incorrectly,
3. Explaining the descriptive statistics clearly when mentioning test scores to students,
4. Making assessments educationally valid and sufficiently clear to stakeholders when deciding how to grade a course,
5. Improving micro-assessment and daily feedback skills and
6. Consulting with peers and working to rectify the problem of unethical assessment practices (Newfields, 2006).

However, we should not forget that in some contexts, teachers are not involved in assessment decision-making processes due to top-down managerial approaches (Troudi, Coombe, & Al-Hamly, 2009). Which is why, even if teachers have acceptable assessment literacy in terms of the above points, they do not necessarily follow acceptable practices. I think that ethicality involves all the above points in some way or another because basically these are followed to maximize the ethicality and validity of language assessment.

6 Ethicality of Language Assessment and Alternative Assessments

Semantically speaking, ethics, morality and fairness are members of the same semantic set (Hamp-Lyons, 2001). Ethicality includes the issues of harm, consent, fairness, deception, privacy and confidentiality (Lynch, 1997). It also covers validity, absence of bias, access, administration and social consequences (Kunnan, 2003, as cited in Bachman, 2005). Furthermore, ethics for language testers also involve whose voices are heard, whose needs are met and how society determines what the best course of action is when fairness is in conflict (Hamp-Lyons, 2001). Generally, the agreed upon core principle of ethical language testing is that no test-taker will be harmed by tests (Lynch, 1997). This is related to test validity as well since ethical issues and practices to protect participants' rights not to be harmed, coerced or manipulated socially, psychologically, emotionally and physically is certainly an essential part of test validity (Lynch & Shaw, 2005).

For further consideration of the ethicality of language testing, we should revisit CLT, which addresses the questions of societal values, consequences, test developers' and users' ethical responsibilities, and how and why language assessments are used (Bachman, 2007). This is because test-takers' rights not to be harmed remain constant whether dealing with traditional or alternative assessments despite the fact that educators' definitions of coercion and ethical responsibility might differ (Lynch & Shaw, 2005). However, I am in agreement with Lynch (1997) in that if denying some test-takers entrance to university or preventing some from accessing specific social and economic resources is considered harmful, no high-stakes test would be considered moral at all. An interesting point to contemplate though is that tests are regarded as a means for finding differences in abilities but they do not create those differences (Lynch, 1997).

I and possibly many language testers believe that achieving fairness is not an easy task because we should base our understanding and analysis on contradictory strategies for considering fairness from different stakeholders' perspectives including taxpayers, education department officials, businessmen, political parties and governments (Alderson & Banerjee, 2002). That is why making a so called *fair* test is quite difficult, fairness is a concept for which there is no single perspective determining the outcome, but rather a complex combination of multiple viewpoints (Alderson & Banerjee, 2002).

Some examples of possible misuses of language testing might be the use of IELTS with applicants for immigration to New Zealand, and also using TOEFL and other proficiency tests to measure test-takers' achievements and growth in instructional programmes (Alderson & Banerjee, 2002). Shohamy (1997, as cited in Alderson & Banerjee, 2002) states that language tests containing unfair content or methods are unethical. She also argues that tests used to control and manipulate stakeholders rather than showing proficiency levels are unethical as well; hence, she advocates CLT. There is a heightened interest in the ethics and role of testing in society according to McNamara (1998 as cited in Alderson & Banerjee, 2002), who anticipates a renewed awareness of the socially constructed nature of test performance and test score interpretations, and an awareness of the issues of testing in the context of English as an international language. Thus, ethically speaking, the credibility of the evidence on which we make test decisions (Bachman, 2005) is, I believe, extremely important for language testers who are considered independent moral agents that can refuse participation in procedures that violate their personal moral beliefs. This is parallel to the basic intent of the Code of Ethics, which, simply put, calls for testers to adopt ethical practices (Alderson & Banerjee, 2002).

Furthermore, to maximize fairness and accessibility of tests, the influence of construct-irrelevant knowledge of test materials should be minimized and unnecessarily controversial, inflammatory, offensive or upsetting test material must be avoided (Pitoniak et al., 2009). However, giving equal chances to students might not sufficiently enhance ethicality because if a broader sense of test validity and ethicality is considered, testers cannot limit their attention to providing only fair chances to test-takers, since this does not consider the fairness or the consequences of language testing decisions. Therefore, there should always be justifications for

any decisions resulting from differential test performances in accordance with the requirements of equality (Lynch, 1997) and thus, ethicality will be maximized.

Regarding the legalization of ethicality, the impetus for appropriateness and responsiveness of assessing young ELLs is supported by some legal requirements and ethical guidelines in the form of case law, public law, and ethical codes from professional organizations that support using sound assessment tools, practices and interpretations (Garcia, Lawton, & Diniz de Figueiredo, 2010). Also, beginning last century, because it became obvious that schooling was a key to social mobility and that achievements in school were used for entry into the higher education or the workplace, many jurisdictions instituted standardized testing to try and ensure fair, accurate, and consistent opportunities for all students (Earl & Katz, 2006). The introduction of legal support and standardised tests over the years can be viewed as an indication of the fundamentality and significance of ethicality and fairness in English language assessment. Moreover, almost all codes (which represent the widely accepted beliefs about validity, reliability, washback and fairness of language tests) consist of a set of expectations for the purpose of potentially judging language testers and improving the quality of language assessment fairness (Jia, 2009).

Fairness can be also achieved through teachers' assessment literacy. In this respect, teachers should be aware of using multiple methods in assessment for ensuring fairness for students (Troudi et al., 2009). A range of assessment approaches and multiple measures allows students to show their knowledge in many ways, and can reveal a composite picture of student learning (Earl & Katz, 2006). Most of these multiple measures can take the form of alternative assessment methods, which the author argues might be the best way to achieve language assessment ethicality for ESL/EFL students.

After discussing assessment literacy and ethicality of language testing and assessment, let us investigate the possibility of the influence of language assessment literacy on language assessment ethicality.

7 The Influence of Language Assessment Literacy on Language Assessment Ethicality Through Alternative Assessments

First of all, it is deemed urgent for ESL/EFL teachers to be benchmarked so that they can instil better self-knowledge in students, so as to make them less reliant on teachers' sometimes inaccurate assessment of their progress. Also, teachers possibly need repeated benchmarking courses, especially when they enter a new teaching situation or encounter new kinds of learners (Hamp-Lyons, 2001). Secondly, Mahboob and Tilakaratna (2012) recommend that first time TESOL teachers should be equipped with the knowledge of assessment, i.e., raise their language assessment literacy to empower students and measure their development properly. This proper

measurement with empowered students guarantees, to a large extent, the ethicality and fairness of ESL/EFL assessment in a democratic testing context. I think that all stakeholders and practitioners need to be aware of the power of testing and comprehend its ethical issues (Shohamy, 2001). Thus, when they have an idea of the extent to which language testing influences students and society in general, language testing policymakers, specialists, and test users are forced to minimize the negative consequences and to maximize the positive consequences of using high-stakes tests of L2 ability (Stoyhoff, 2008). Of course, this will foster the ethicality and fairness of testing and assessment.

I believe that assessment literacy has a major role in assessment ethicality and that teachers need to have knowledge about the following issues: Awareness regarding who designs assessment procedures, appropriateness of assessment procedures to the learners, how information on students' learning is collected, how information is assessed and stored, accuracy and fairness of assessment results, and students' views of those assessment procedures (The EALTA Executive Committee, 2006). ESL/EFL teachers should also understand the issues and concepts of assessment such as accountability bias, language proficiency, and testing accommodations. Additionally, they need to have an awareness of different purposes of assessment for measuring language knowledge or ability, and that tests need to be equitable, accurate, consistent, and practical to administer, that is, to be fair, valid, reliable and easy, respectively. In this aspect, performance-based assessments are the best measure of the criteria that cover formative and summative assessments (TESOL International Association, 2010). Furthermore, ESL/EFL teachers must know that assessments for English native-speakers and ELLs differ. For example, assessments for ELLs might contain cultural bias such as unfamiliar images or references, or contain linguistic bias, for instance, some language items may be more difficult for ELLs due to their complex language (TESOL International Association, 2010). For the next phase, ESL/EFL teachers should know that these assessments can be used to show language growth over time and to find areas that need more focus. To this end, they should know about portfolio assessment, which is a collection of students' work that reflects progress over time and its samples are based on class activities or home assignments (TESOL International Association, 2010). They must also understand that self-assessment and peer-assessment methods must be used regularly to push students to monitor and control their learning. In addition, ESL/EFL teachers have to be sure about their students' prior experience with the test questions and answer formats (TESOL International Association, 2010). If the above-mentioned types of literacies are possessed by teachers, this can certainly be a contributing factor to ensure and maximize fairness and ethicality of language assessment in general and English language assessment in particular.

So far in this paper, literacy in alternative assessments and ethicality of language assessment and the impact of assessment literacy on assessment ethicality has been elucidated. Next, these issues will be considered with reference to Kurdish tertiary EFL teachers' academic context.

8 The Importance of Assessment Literacy for the Ethicality of Language Assessment in the Context of Kurdish Tertiary EFL Teachers

The recommendation of Mahboob and Tilakaratna (2012) to urgently equip teachers with the knowledge of assessment in order to empower students and measure their developments properly is not followed in the English departments of the Kurdistan Region universities. That is why the ethicality and fairness of ESL/EFL assessment, as outlined in Shohamy's (2001) democratic testing context, does not exist in those departments. This is because the majority of the Kurdish tertiary EFL teachers have not had any training courses to equip them with assessment knowledge. There are simply no such training courses in the Kurdistan Region that focus on English language assessment and testing or cover these subjects adequately. Thus, almost no teachers have had any opportunity to attend relevant courses, except those very few teachers that have studied abroad. As a result, I believe that nearly all Kurdish EFL teachers have a low level of assessment literacy that probably affects the fairness of their assessments negatively. For instance, I was not aware of language assessment standards, professional morality (Alderson & Banerjee, 2002), the significance of language assessment literacy, assessment literacy benefits for teachers, and most importantly the use of multiple measures in language assessments. All of which are necessary to ensure fairness for students (Troudi et al., 2009) and to protect students from the misuse and abuse of tests (Alderson & Banerjee, 2002). I strongly believe that I and many Kurdish EFL teachers are not aware of these assessment issues. This understanding is based on my experience observing Kurdish tertiary EFL teachers regarding their assessment literacy and their practice of alternative assessments. My observation is valid because I was part of the situation as an insider, working as a lecturer and researcher in several English departments for nearly 5 years from 2008 to 2012. During that time, I was the deputy head and then the head of an English department, the coordinator of quality assurance in another department and visiting lecturer in another one, in addition to attending many meetings and workshops in Hawler in which the representatives from all English departments in the Kurdistan Region were present. Even in those meetings and workshops there were no mention of assessment and testing or alternative assessments. My observation has been confirmed by the data I collected for my EdD study in 2015, in which I devoted two items in the questionnaires. In the first item, almost all teachers (nearly 92 %), mentioned that they have never had a course on assessment and testing or alternative assessments. Though they claim that they know something about it, their answers to the open-ended question on what they know about alternative assessments revealed merely their views on alternative assessments and did not indicate any knowledge of the concepts involved. This speaks to their limited knowledge of assessment in general and alternative assessments in particular. Also, all of the 12 interviewed teachers mentioned that they have never received any standards or guidelines for doing assessments, testing and alternative assessments from their

departments to follow. So, when teachers have not attended any training course on assessment or alternative assessments and have not received any guidelines or standards to follow, they logically do not have adequate assessment literacy. Therefore, it can be reasonably assumed that Kurdish tertiary EFL teachers are severely lacking assessment literacy and that, therefore, our assessments might, to a certain extent, be unethical and invalid.

Additional evidence of our possible unethical assessment is that, logically and educationally, having language assessment literacy influences the ethicality of language testing and assessment as it leads teachers to assess ESL/EFL students' proficiency more comprehensively and fairly by using various techniques more frequently throughout the academic year. Assessment literacy also makes teachers aware of ethical issues such as harm, consent, fairness, deception, privacy, confidentiality (Lynch, 1997) validity, absence of bias, access, administration, and social consequences (Kunnan, 2003, as cited in Bachman, 2005). Therefore, any teacher who has the knowledge of these issues will assess language proficiency so differently (almost surely more ethically) than others. So, as Kurdish tertiary EFL teachers have a low level of knowledge of these assessment issues, they perhaps cannot assess Kurdish EFL learners ethically.

9 The Effectiveness of Alternative Assessments for a More Comprehensive and Fairer English Language Assessment in the Context of Kurdish Tertiary EFL Teachers

Because decisions based on English tests can change test-takers' lives (Shohamy, 1998, as cited in Piggini, 2012) both academic and social English language skills must be assessed accurately to provide a clearer picture of students' English proficiency (Stephenson et al., 2003). Therefore, I think that alternative assessments which occur at various points in time and in various ways both inside and outside the classroom (Tsayari, 2004), are the best procedures to assess academic and social English, and this is in line with the TESOL International Association's (2010) recommendation of using various performance-based assessment tools and techniques. However, regarding the present assessment practices of Kurdish tertiary EFL teachers, I have realized that they do not have adequate knowledge of alternative assessment methods and they do not implement them sufficiently. That is why, in my estimation, the methods of assessing English language proficiency in the English departments of the Kurdistan Region are not adequate and possibly not fair, since it is chiefly done by traditional testing, whether monthly or as formal final-year exams by which 90 % of marks are allocated for 'pen and paper' tests. These tests are not multi-dimensional and thus cannot possibly give a comprehensive idea of Kurdish tertiary EFL students' proficiency of English. However, there are some classroom performance-assessments, but these are very restricted and do not present equal

opportunities for all students. Even in those restricted practices, some Kurdish EFL teachers find it very difficult to remain unbiased because sometimes students have different levels of engagement depending on the teacher. This is probably due to the large number of students (approximately 40–50 students per class), or it could be related to gender discrimination, i.e. females are more engaged by single male teachers or vice versa—leading to bias towards a certain gender during assessments. In these few cases, assessment (even assessment involving grades) is dependent on teacher-student personal relationships, which is surely unethical.

To avoid such unethical assessments and develop more democratic tests, I suggest adopting alternative assessments for language assessment that are supported by CLT, and in which test-takers and local bodies are more active (Dai Quang, 2007). For example, piloting tests and asking for students' opinions (The EALTA Executive Committee, 2006) is a procedure for developing such a democratic assessment. To this same end, Shohamy's guidelines for making testing more democratic, Freire's anti-authoritarian, dialogical and interactive approach (Chandella & Troudi, 2013) and Foucault's consideration of ethics as the practice of freedom can be achieved in portfolio assessment, an assessment for which students' have the freedom to shape the portfolio process and form (Lynch & Shaw, 2005).

I believe that this democratic model using portfolio assessment is quite suitable for my professional context. Yet the majority of Kurdish tertiary EFL teachers most likely have little knowledge and experience of that model of assessment and the related assessment practices, principles, and issues. Instead they mostly use multiple-choices, fill-in-the-gaps, matching tasks, etc. that are often incompatible with current ESL/EFL classroom practices, the tasks cannot integrate instruction and assessment properly. Hence, there is an urgent need for having training courses for these teachers to equip them with the knowledge of alternative assessments and enable them to adopt democratic assessment.

Regarding this democratic approach, I believe that many Kurdish tertiary EFL education majors might not have positive views on their learning process when considering alternative assessments because they just want to graduate and be employed. Student resistance is only one side of the equation, as not only are students indifferent before graduation, but also some school managers' are indifferent about their graduates' level of proficiency, since they are not expected to have many skills or much knowledge of the English language when they become EFL teachers at schools. I think that this indifference affects assessment ethicality negatively as there is no interest and focus on using assessments to increase the English proficiency of students.

10 Conclusions

In conclusion, broadly speaking, I have realized that language assessment literacy of ESL/EFL teachers, which Mahboob and Tilakaratna (2012) state urgently needs raising, plays a major role in language assessment ethicality and fairness. Raising

assessment literacy means being aware of many assessment issues such as assessment procedures, designs, appropriateness, accuracy, fairness, the difference between assessments for English native-speakers and ELLs, students' rights to know about test questions, answer formats, the need for self-assessment and peer-assessment, taking students' viewpoints, language testing standards and professional morality. I strongly believe that these literacies enhance and may even maximize assessment ethicality effectively. This will hopefully go some way towards protecting all from the misuses of tests (Alderson & Banerjee, 2002). To this end, I also conclude that the most appropriate methods that ensure the ethicality of ESL/EFL assessments in a democratic assessment context is alternative assessments of language proficiency. Also, it can be concluded that awareness and knowledge of alternative language assessment practices can certainly be contributing factors to ensure fairness and ethicality of English language assessment.

Regarding the context of Kurdistan Region's English departments, firstly, I have realized that the current assessment practices of Kurdish tertiary EFL teachers are still based mostly on the traditional 'pen and paper' tests, with most departments allocating nearly 90 % of scores for them. This situation is certainly flawed, and these traditional tests, at the very least, need to be combined with alternative assessment practices. This indicates that these teachers have not yet been exposed to the recent advances in language assessment, which covers various aspects such as ethicality, professionalism, CLT concerns, TESOL International Association standards on assessment, and how assessment improves overall teaching and learning. The lack of incorporated alternative assessments reveals the limited assessment literacy possessed by Kurdish EFL teachers. More importantly, they may not have enough knowledge and experience of alternative assessment methods, which are necessary practices for achieving language assessment ethicality.

Finally, I can suggest that the democratic model of assessment is quite suitable for most Kurdish tertiary EFL teachers and their students. However, this model is not followed now and is not well-known in the English departments of Kurdistan Region universities. Instead, traditional testing techniques are mostly adopted. Nevertheless, I believe that even if such a democratic model of assessment is adopted, some Kurdish EFL students might not be willing to contribute much to getting the most out of alternative assessments as they just want to graduate and be employed. Overall, I can say that English language assessment in my context is, to some extent, inadequate and unethical, and lacks many influential developments that have been achieved through alternative assessments in a democratic assessment context. Therefore, I recommend some training courses on the adoption of TESOL assessment standards, alternative assessment practices, and assessment ethicality and validity to help Kurdish tertiary EFL teachers to increase their English language assessment literacy. Following this they need to then practice the innovative assessments with their Kurdish EFL students for their assessments to be more ethical and valid.

References

- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (Part 1). *Language Teaching*, 35(2), 79–113.
- Atac, B. A. (2012). Foreign language teachers' attitude toward authentic assessment in language teaching. *Journal of Language & Linguistics Studies*, 8(2), 7–19.
- Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly: An International Journal*, 2(1), 1–34.
- Bachman, L. F. (2007). *Language assessment: Opportunities and challenges*. Paper Presented at Meeting of the American Association of Applied Linguistics (AAAL), Costa Mesa, CA.
- Brown, T., & Jones, L. (2001). *Action research and postmodernism, congruence and critique*. Buckingham: Open University Press.
- Brumfit, C. J., & Johnson, K. (Eds.). (1979). *The communicative approach to language teaching*. Oxford: Oxford University Press.
- Chandella, N., & Troudi, S. (2013). Critical pedagogy in language education: Challenges and potentials. In R. Akbari & C. Coombe (Eds.), *Middle East handbook of applied linguistics* (pp. 42–61). Dubai: TESOL Arabia Publications.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and commendations. In C. Coombe, S. Stoyloff, B. O'Sullivan, & P. Davidson (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge: Cambridge University Press.
- Cummins, P. W., & Davesne, C. (2009). Using electronic portfolios for second language assessment. *The Modern Language Journal*, 93(1), 848–867.
- Dai Quang, V. (2007). *Critical applied linguistics: Concerns and domains*. Hanoi: Scientific Research Management Office, College of Foreign Languages—VNU.
- EALTA Executive Committee. (2006). *EALTA guidelines for good practice in language testing and assessment*. Retrieved from <http://www.ealta.eu.org/guidelines.htm>.
- Earl, L. M., & Katz, M. S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Manitoba: Manitoba Education, Citizenship & Youth.
- Hamp-Lyons, L. (2001). Ethics, fairness(es) and developments in language testing. In Elder, C., Brown, A., Grove, E., Hill, K. Iwashita, N., Lumley, T., & O'Loughlin, K. (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 222–227). New York: Cambridge University Press.
- Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers and Teaching: Theory and Practice*, 6(2), 151–182.
- Garcia, E., Lawton, K., & Diniz de Figueiredo, E. H. (2010). *Assessment of young English language learners in Arizona: Questioning the validity of the state measure of English proficiency*. Los Angeles: Civil Rights Project.
- Jia, Y. (2009). Ethical Standards for language testing professionals: An introduction to five major codes. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 13(2), 2–8.
- Lynch, B. K. (1997). In search of the ethical test. *Language Testing*, 14(3), 315–327.
- Lynch, B. K., & Shaw, P. (2005). Portfolios, power, and ethics. *TESOL Quarterly*, 39(2), 263–297.
- Mahboob, A., & Tilakaratna, N. (2012). *A principles-based approach for English language teaching policies and practices*. Alexandria, VA: TESOL International Association.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Newfields, T. (2006). Teacher development and assessment literacy. *Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48–73). JALT: Tokyo.
- Piggin, G. (2012). What are our tools really made out of? A critical assessment of recent models of language proficiency. *Polyglossia*, 22, 79–87.

- Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.
- Reason, P., & Bradbury, H. (Eds.). (2001). *Handbook of action research*. London: SAGE Publications.
- Riasati, M. J., & Mollaei, F. (2012). Critical pedagogy and language learning. *International Journal of Humanities and Social Science*, 2(21), 223–229.
- Ripley, D. (2012). Implementing portfolio-based language assessment in LINC programs: Benefits and challenges. *TESL Canada Journal*, 30(1), 69–86.
- Shohamy, E. G. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Pearson Education.
- Short, D. J. (2000). *The ESL standards: Bridging the academic gap for English language learners*. ERIC Clearinghouse on Languages and Linguistics (ED447728).
- Stephenson, A., Johnson, P. D., Jorgensen, M. A., & Young, M. J. (2003). *Assessing English language proficiency: Using valid results to optimize instruction*. San Antonio, TX: Harcourt Assessment.
- Stoyhoff, S. J. (2008). *Developments in English language assessment APEC strategic plan for english and other languages*. Mankato: APEC Human Resources Development Working Group, Minnesota State University.
- TESOL International Association. (2010). *TESOL/NCATE standards for the recognition of initial TESOL programs in P-12 ESL teacher education*. Retrieved from [http://www.tesol.org/docs/books/the-revised-tesol-ncate-standards-for-the-recognition-of-initial-tesol-programs-in-p-12-esl-teacher-education-\(2010-pdf\).pdf?sfvrsn=0](http://www.tesol.org/docs/books/the-revised-tesol-ncate-standards-for-the-recognition-of-initial-tesol-programs-in-p-12-esl-teacher-education-(2010-pdf).pdf?sfvrsn=0).
- Thibeault, C. H., Kuhlman, N., & Day, C. (2010). ESL teacher-education programs: Measuring up to the TESOL/NCATE yardstick. *CATESOL Journal*, 22(1), 48–59.
- Troudi, S., Coombe, C., & Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546–555.
- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. Center for Research in Language Education, CRILE Working Papers, 58. Retrieved from <http://www.ling.lancs.ac.uk/groups/crile/docs/crile58tsagari.pdf>.
- Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628–660.
- Wolf, M. K., Herman, J. L., Bachman, L. F., Bailey, A. L. & Griffin, N. (2008). *Recommendations for assessing English language learners: English language proficiency measures and accommodation uses (Part 3 of 3)*. CRESST Report 732. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Part II
Theoretical Perspectives on Assessment

A Critical Review of Washback Studies: Hypothesis and Evidence

Wei Wei

Abstract This chapter aims to demonstrate the current understanding of washback effects in the context of language testing and language education from the following three perspectives: (1) demonstrating the historical development of the concept of washback in both general education and language education, (2) elaborating the results of the empirical investigations and highlighting the gaps between hypotheses and evidence and (3) identifying the directions and areas for washback studies in future.

Keywords Assessment · Washback · Impact

The negative impact of standardized tests on the teaching, learning and implementation of national new curriculum in general education has been widely discussed and empirically tested in the United States (Smith, 1991a, 1991b; Madaus, 1988; Power & Alderman, 1983), Canada (Wideen, O'Shea, & Ivany, 1997), Europe (Kellaghan, Madaus, & Airasian, 1982) and Asia (Gorsuch, 2000; Zeng, 1999). The concepts of constructive alignment between external high-stakes assessments and classroom teaching have been proposed as one of the solutions (Herman, Webb, & Zuniga, 2007). Moreover, the idea of formative assessment or assessment for learning has started to receive increasing attention from educational policy makers, such as the Assessment Reform Group (ARG) in the UK and the international Organization for Economic Cooperation and Development (OECD). More recently, in the field of applied linguistics and English education, the attempts of using high-stakes tests (mostly summative assessments) to reform classroom language teaching and learning have been made by researchers and education policy makers in a wide range of contexts such as Hong Kong, Sri Lanka, Greece and Eastern Europe. The international exam developers such as Cambridge ESOL and ETS have also started to include the evidence of positive or intended washback effects in their research plans and validation reports.

W. Wei (✉)

University International College, Macau University of Science
and Technology, Macau, China
e-mail: soroswei@gmail.com

This chapter begins by reviewing the key concepts in general education and educational assessment theory (Messick, 1989, 1996), such as alignment between curriculum and assessment, consequential validity, and assessments for learning and formative assessments. Following on from this, the empirical studies of washback effects in the past 20 years in the field of language education will be reviewed with special reference to the following four areas, (1) hypotheses which have or have not been empirically tested, (2) research design, the selection of research instruments and evidence, (3) the nature or purposes of the test and contexts being investigated, and (4) recent attempts (post 2000) to identify the mediating factors.

1 Key Concepts in Washback and Impact Studies

In the field of testing, washback sometimes refers to consequential validity (Messick, 1989; Gronlund & Waugh, 2009), that is, the validity of a test is concerned with the consequences of using the assessments (Gronlund & Waugh, 2009, p. 47) and Messick's (1989) comment that judging validity in terms of whether a test does the job it is employed to do ... requires evaluation of the intended or unintended social consequences of test interpretation and use (1989, p. 84). Regarding the washback effects, according to Madaus (1988), it is testing, not the official stated curriculum, that is increasingly determining what is taught, how it is taught, what is learned, and how it is learned (1988, p. 83). Alderson and Wall (1993) simply define washback as the influence of testing on teaching and learning. It is the practice of where the teacher and students do extra teaching and learning with the express purpose of passing the test. They proposed fifteen hypotheses in washback effects, which will be discussed in detail later. Later, Bailey (1996) summarized three basic perceptions of washback: (1) washback is generally defined as the influence of testing on teaching and learning; (2) it is widely present and is important and (3) relatively little empirical research has been done to document its nature or the mechanisms by which it works.

In relation to the impact of a test, Bachman and Palmer (1996) stress that this covers a broader area than washback does. They define the impact of a test as having at least two levels: a micro level, in terms of individuals who are affected by the particular test use, and a macro level, in terms of the educational system or society (1996, p. 30). From the definitions of washback and impact, it is clear that (1) washback is a part of impact; (2) washback is one of the micro-level impacts (in terms of individual teachers and students); and (3) the influence of tests on other individuals (parents, for instance) and the educational system (i.e., school, teacher training program, curriculum and textbook) are examples of impact.

2 Consequential Validity Studies in General Education

This section considers impact studies in general education. Through searching the American-based database J-STOR, 13 empirical studies that focused on test impact in general education were reviewed. Nine of these studies examined the close relationship between the test and the teaching, and two of these took educational culture into account and showed how culture influences the test and the teaching. Another two studies explored the issues of teachers' professional development under the high-stakes test system. General information (the author, the investigated examination, research methods and main findings) regarding these thirteen empirical studies is summarized in Table 1.

To summarize the main findings of these studies in general education, the following themes have been used: washback of the test on teacher and teaching, washback of the test on learners and learning, impact of the test on school, and impact of the test on parents, although none of the studies investigate the relationships among all these factors.

Eight studies report the impact of testing on teachers and teaching as negative in that it decreases the variety of teaching approaches, narrows the curriculum, reduces the teaching time, practises the commercial test preparation book and creates great pressure on teachers. For example, in Smith's (1991a, 1991b) study of the impact of the Iowa Test of Basic Skills in America on teachers work, she finds that teachers experience negative emotions as a result of the publication of test scores and subsequently decide to change their methods to avoid low scores. The teachers interviewed believed that scores were used against them, despite the perceived invalidity of the tests themselves. In addition, the test substantially reduced the instruction time, and narrowed curricular offerings and modes of instruction. Therefore, the test potentially reduce(s) the capacities of teachers to teach content and to use the methods and materials that are incompatible with standardized testing forms (1991a, 1991b, p. 8). In Herman and Golan's (1993) survey study, the results show that the respondents felt pressure to improve students' performance in tests. As a result, testing affects (their) instructional planning and delivery (1993, p. 21), and a substantial amount of time is spent on preparing students for testing.

Two studies consider the learner: one (Huhta, Paula, & Pitkanen-Huhta, 2006) explores the cognition process of the learner during their test preparation; the other (Halpin & Halpin, 1982) explores the impact of testing on learners' attitudes and its relation to learning outcomes. Huhta et al. (2006) identify four basic repertoires (feelings and experiences) in students preparing for the examination in terms of how they define their studies in preparing for the exam and what factors they attribute to their success or failure in the exam:

Table 1 Summary of research findings on washback

Authors	Investigated test	Research methods	Main findings
Kellaghan et al. (1982)	Classroom-based standardized test in Ireland (low-stakes)	Interview Survey	Displays the effect of the standardized test on schools, parents, teachers and pupils. This study concluded that there is a little impact of the test on them
Halpin and Halpin (1982)	Classroom-based summative test (low-stakes)	Compare students' performance who knew or not knew there was a test at the end of study	The students who studied to take a test not only achieved more, but also retained their learning longer than those students who studied without the pressure of test
Powers and Alderman (1983)	SAT in US, high-stakes test	Survey	The students hold a favourable attitude to the commercial test preparation materials and the material increased candidates confidence before the SAT, but there is little effect on the performance of those candidates who had experienced the SAT before
Mehrens and Kaminski (1989)	California achievement test in US	Document analysis Survey	The pressure from the test may force schools and teachers to do illegitimate test preparation work. Commercial materials for test preparation may be unethical, which make students score having weak inference about their knowledge, depending on how closely they match the tests
Smith (1991a, 1991b)	State-mandate Iowa Test of Basic Skills (ITBS), high-stakes	Interview Classroom observation	Testing influences teachers' attitudes towards the test/education. Testing reduces the time for instruction, narrows the possible curriculum, and reduces teachers' ability to adapt and create material/activity; the format of the test leads to teaching filled with multiple-choice questions
Smith (1991a, 1991b)	State-mandate Iowa Test of Basic Skills (ITBS)	Classroom observation Interviews Survey	Eight types of test preparation activities: (1) no special preparation for test, (2) training pupils test wiseness or test-taking skills, (3) reminding students to get a good night's sleep and breakfast, (4) preparing pupils for tests by reviewing the content of ordinary instruction, (5) teach students the test preparation commercial textbook, (6) releasing stress, (7) coaching the questions from former tests, and (8) cheating in the examination, such as providing extra time on timed portions of tests, providing correct answers and altering marks on answer sheets
Smith and Rottenberg (1991)	State-mandate Iowa Test of Basic Skills (ITBS)	Classroom observation Interviews Survey	Testing affects what elementary schools teach: in a high-stakes environment, schools neglect material that external tests exclude; and external testing affects school organization by placing general boundaries on placement and instructional opportunities
Hall, Kleine, and Paul (1992)	High-stakes Non-reference Test (NRT) in general	Survey	In teaching, the curriculum has been changed to match the skills/questions contained in NRT; most of the educators (principals, teachers, testing coordinator) admitted the existence of inappropriate testing practices (such as

(continued)

Table 1 (continued)

Authors	Investigated test	Research methods	Main findings
Herman and Golan (1993)	High stakes test in general	Survey	increasing time limits, giving hints); nearly one third of teachers noted they teach to test The effects of testing on teaching and on teachers attitudes; teacher feels pressure to improve students test scores; school administrations give substantial time to test preparation; testing affects instructional planning and delivery; non-tested subjects also get some attention; teachers generally take responsibility for their students' performance; and teachers have doubts about the efficacy of testing
Shepard and Bliem (1995)	Standardized test and performance test in elementary school (classroom-based)	Interview Survey	Parents did not see the score in standardized tests as the main indicator for their children's progress and the schools performance. The main information the standardized test rather than the performance test provided was the comparative information with other children, but for the elementary school stage, the parents can get such information from talking with teachers
Wideen et al. (1997)	School leaving examination in Science in Canada for Grade 12 students	Classroom observation Interview	The science test affects curriculum and instruction at classroom level. The data showed a decrease in the variety of instructional approaches from Grade 8 to Grade 12 (when the examination occurs)
Chudowsky and Behuniak (1998)	Academic performance test: (CAPT) was designed to promote high quality instruction	Group interview with teachers	Impact on curriculum and instruction Impact on teacher expectations (teacher and school have high expectations of student) Impact on students behaviour and attitudes Impact on parents behaviour and expectations (test motivated students to learn) Impact on professional development
Gorsuch (2000)	University entrance examination in Japan, high-stakes	Interview Survey	Obvious exam effects on the formal local school and informal classroom cultures are found. The educational culture, which focuses on grammar-translation-based university entrance examination, constrains curriculum innovation
Huhtia et al. (2006)	Matriculation Examination in Turkey	Recording and analyzing an oral diary Group discussion	This research discovered four basic repertoires (feelings and experiences) in students preparing for the examination The details will be presented below
Sternberg (2007)	Test in general	Observing Interview	Students do better on assessments when the materials on which they are assessed are familiar and meaningful to them; individuals in different cultures may think about concepts and problems in different ways; when children are taught in culturally appropriate ways, their achievement increases

- (1) The first type of student depicts themselves as either diligent or lazy. Their predictions and explanations for success or failure in the test are based on their efforts in studying.
- (2) The second type of student defines themselves as skilled enough or not. They relate their success in the test to their understanding and their method of acquiring knowledge.
- (3) The third type of student is confident they will pass the exam without becoming stressed. They relate their success or failure to personal characteristics like the ability to concentrate or a tendency to be nervous.
- (4) The last type of student considers themselves to be either lucky or unlucky test candidates. Therefore, they believe that failure in a test is not their fault; their expectations and explanations for success or failure in the test are based on the randomness of the test result.

In the second study, Halpin and Halpin (1982) introduce a class-based summative test to their class. Comparing two groups of students, the result indicates that the (first group of) students who studied and took a test at the end of the program not only achieved more but also retained their learning longer than those students (in the second group) who studied in order to learn rather than for a test (1982, p. 32). In other words, their experiment infers that testing is a valuable promoter in the learning process. Even if the students feel negative about it, testing seems to cause them to work harder, learn more, and remember longer was learned (1982, p. 37). It is therefore worthy of investigating the extent to which a test can influence students learning outcomes and whether this effect can be related to students' characteristics (e.g. socioeconomic background, age and gender) or subjects they have learned.

There are five studies that indicate a test's impact on a school's routine work. Most of them (Mehrens & Kaminski, 1989; Smith, 1991a, 1991b; Smith & Rottenberg, 1991; Herman & Golan, 1993) describe the school as a mediator in the mechanical process of transferring the impact of the test onto their teachers. In other words, the pressure from a high-stakes test on a school is likely to be transferred to teachers, which results in test-like teaching. In Iowa, America, Smith and Rottenberg (1991) report that in high-stakes environments, the school neglects material that external tests exclude (1991, p. 9). In addition, external testing affects school organization by placing general boundaries on placements and instructional opportunities (1991, p. 10). That is to say, schools would use the pupils' performance in an external high-stakes test to group them into different levels and grades to study. Furthermore, Herman and Golan (1993) explain this phenomenon more explicitly: standardized testing has substantial effects on schools and their teaching and learning processes: schools send out the message to their teachers about the

importance of test-curriculum alignment, and teachers design their instructional program with such alignment in mind (1993, p. 23). Therefore, in the class, substantial time and attention are devoted to ensuring that students are taught tested subjects, are given practice in expected test content, and spend time in special test-preparation activities (1993, p. 24). These results show that some schools are influenced by the test and tend to pass the pressure from the test on to their teachers and students, although further investigation needs to be conducted to investigate what kinds of schools are more vulnerable to the political pressures.

Two studies focus on the relationship between parents and tests: one examines parents' different attitudes towards the performance test and standardized test; the other depicts how a test triggers parents' participation in their children's learning process. Shepard and Bliem (1995) report that the parents in elementary schools do not see test scores as the main indicator of their children's performances in school. They prefer to communicate with teachers and receive feedback. In addition, they are more concerned about their children's relative standing of their academic performances compared with other students in the same or other areas; they even believe that the performance test can lead to more creative learning. The other study, conducted by Chudowsky and Behuniak (1998), concludes that a high-stakes test can re-arouse parents' interests in assisting their children in learning, and the more high-stakes a test is, the greater the concern will be. In the case of the newly introduced Connecticut Academic Performance Test (CAPT) in America, Chudowsky and Behuniak (1998) note that (the new test encourages) some parents to be more interested in their children's homework and preparation for the test; however, since the new test in this research is not a very high-stakes test, parents were more concerned about SATs, which really counted (1998, p. 35).

3 The Tensions Between Summative and Formative Assessments

After realizing the destructive role of standardized text on learning and teaching, formative assessment and the idea of assessment for learning, has been proposed as one of the solutions. Previous discussions and empirical studies of summative and formative assessments have been focused on the following two areas: (1) what the main differences or tensions between summative and formative assessments are and its pedagogical implications to teachers' classroom practices, and (2) what factors make the results or feedback from summative and formative assessments more formative and therefore benefit and promote the learning process more effectively.

In relation to the first thread, the theoretical discussions (Harlen & James, 1997; William & Black, 1996; William, 2000) have summarized the following differences: the collected information, nature of results or feedback, the purposes of assessments. In contrast to summative assessment, William (2000) presents a powerful framework to define the characteristics of formative assessments: questioning, feedback, sharing quality criteria and student self-assessment. Moreover, the concepts of feedback and self-assessment stand out from the previous discussions as two of the most distinguished differences (Harlen, 2005; William, 2000). In relation to feedback, the feedback from formative assessments are considered as intrinsically motivating (Ames & Archer, 1988), informational rather than judgmental (Brookhart, 2001) and directional which highlights the gap between actual and expected performances and the strategies to learn (Ramaprasad, 1983). In relation to self-assessment, Ross and Bruce (2007) divide it into three components based on the work on student self-assessment: self-observations, self-judgements and self-reactions. As a result, to facilitate the process of self-assessment, the previous discussions suggest that (1) students and teachers share the same understanding of the learning outcomes or the constructs being assessed in tests (Sadler, 1989), (2) students need to be aware of the gap between their actual performances and expected performances and the strategies to narrow down that gap (Harlen & James, 1997). Last, Harlen (2005) makes a wider range of suggestions on classroom instruction: explaining to students the purpose of tests, using assessments to convey a sense of progress, providing feedback that inform learning and developing students self-assessment skills (2005, p. 211).

For the second thread, to summarize the results from empirical studies, the identified factors which may undermine the formative function of feedback or results from the assessments include (1) practitioners lack of understanding of the nature and purposes of summative assessments (Taras, 2008), (2) cultural influence (Kennedy, Chan, Fok, & Yu, 2008) and (3) students' interpretations and usage of the results (Brookhart, 2001). Taras (2008) conducted a small scale study with lecturers at the School of Education in a British university. The results suggested that the lecturers did not really understand the features and the differences between summative and formative assessments, even though the classroom formative assessments have been used extensively in practice. Researchers continued attempts to make the results from formative and summative assessments more formative seem to have another barrier, which is the cultural issue. Kennedy and his colleagues (2008) argue that because of the exam-oriented culture within the education system in Southeast Asia, formative assessments should not be proposed as a tool to replace summative assessments, which are widely believed to have negative backwash effects to classroom teaching and learning. Rather, when designing assessment tasks for formative and summative assessments, the designers and

policy makers are advised to be aware of those conditions which may ‘provide the conditions for students to learn’ (2008, p. 204) and the process they have to follow to learn more effectively. Last, Brookhart’s study (2001) provides further evidence that what matters are not the forms of assessments but students’ interpretations of the results, self-assessment and the integration of the test results into their learning process. This study finds that the successful students did integrate their learning with the assessment results by feeding the information from their self-interpretation and self-assessments into their learning, and neglecting the distinctions between summative and formative assessments.

4 The Hypotheses in Washback Empirical Investigations

Although formative assessments have been widely accepted as a way to promote better learning and classroom teaching in general education, washback studies in the field of language testing have still primarily focused on the high-stakes summative assessments, for example Senior High School Leaving English Tests, National Matriculation English tests, IELTS and TOEFL. Moreover, the studies in general education appear to express a pessimistic attitude towards the consequences and impacts of high-stakes test, while the researchers in applied linguistics seem to hold a mixed feeling. The following empirical investigations (see Table 2) examine what Alderson suggested in 1993, namely, the Fifteen Washback Hypotheses (see Table 3). Most of the empirical studies are relevant to the investigation of the first hypothesis—a test will influence teaching, but only four of them take the effect of washback on the learner or learning into account.

Table 2 Empirical investigations that test the hypotheses

Investigators	Year	Context and test
Li	1990	National Matriculation English test in China
Alderson	1993	O-level English test in Sri Lanka
Watanabe	1996	University entrance examinations in Japan
Alderson	1996	TOEFL in America
Hayes	1996	IELTS in New Zealand
Cheng	1997 and after	Hong Kong Certificate of Education Examination (HKCEE)
Shohamy	2001	Test of Arabic as a second language and EFL oral examination in Israel
Andrew	2002	Hong Kong’s Advanced Supplementary Use of English oral examination

Table 3 The Alderson and Walls hypotheses (1993)

About	Hypotheses
Teaching	A test will influence teaching
Teaching	A test will influence WHAT teachers teach
Teaching	A test will influence HOW teachers teach
Teaching	A test will influence the RATE and SEQUENCE of teaching
Teaching	A test will influence the DEGREE and DEPTH of teaching
Learning	A test will influence learning
Learning	A test will influence WHAT learners learn
Learning	A test will influence HOW learners learn
Learning	A test will influence the RATE and SEQUENCE of learning
Learning	A test will influence the DEGREE and DEPTH of learning
T & L	A test will influence attitudes to the content and methods of teaching and learning
Others	Tests that have important consequences will have washback
Others	Tests that do not have important consequences will have no washback
Others	Tests will have washback on all learners and teachers
Others	Tests will have washback effects for some learners and some teachers, but not for others

5 Hypotheses Relevant to Teaching

In Alderson's fifteen washback hypotheses, five hypotheses are relevant to teaching. Regarding the hypothesis of 'a test will influence what teachers teach', the previous empirical studies agree that there are washback effects on teaching content. For instance, in Alderson and Wall's (1993) washback study of the newly-introduced and high-stakes O-level examination in Sri Lanka, they conclude there is washback on teaching content, saying that it is clear that there is a narrowing of the curriculum as teachers finish or abandon their textbooks and begin intensive work with past papers and commercial publications to prepare their students for the exam. At this point, there is obvious exam impact on the content of the teaching (1993, p. 5). More recently, Cheng (1997) argues that the most dramatic change due to the introduction of the 1996 Hong Kong Certificate of Education Examination (HKCEE) lay in the content of teaching—an area of high washback intensity (1997, p. 50).

With respect to the hypothesis of 'a test will influence how teachers teach', the previous investigations depict surprisingly different pictures. Some studies suggest that a test is powerful enough to promote expected teaching methodology reform (e.g. Li, 1990), while some studies suggest that there is little evidence to draw any conclusion of washback effects on classroom teaching methods (Alderson & Wall, 1993); other studies challenge this idea and report that such washback effects on teaching are usually negative (Qi, 2005; Cheng, 1997, 1998, 1999) and varied by teachers educational background and working experiences (Watanabe, 1996). Last but not least, some studies (Shohamy, Donitsa-Schmidt, & Ferman, 1996,

Shohamy, 2001) provide evidence that washback effects on classroom teaching can be either enhanced or reduced, due to other factors, such as the subjects' status in education and the tests stakes for teachers and students.

Li (1990) investigated the newly introduced Matriculation English Test by interviewing secondary English teachers in China. This test was originally designed to facilitate the change in classroom teaching methodology. The investigation showed that as a result of the test, teachers' teaching focuses have started to shift from linguistic knowledge to communicative ability, assisted by imported authentic textbooks and reading materials. Fourteen years later, Qi's research (2004) on the same test suggests that little of the intended washback effect on teaching methods remains in English classrooms: the focus on teaching is on linguistic knowledge and NMET-oriented use of English. The language tasks used in English classes are very test like. That is to say, the intended washback effects on English teaching (communicative tasks) are not evident in class, but NMET-tasks are observed. Qi's research notes that the high-stakes test fails to initiate a change in teaching methodology due to the contradiction between the test's social function (selecting candidates for higher education) and promoting function (changing the teaching methodology at classroom level). Cheng's (1997, 1998, and 1999) investigations on the new HKCEE highlight some possible explanations for this problem. By introducing the new test, test developers wanted to encourage teachers to spend more time on teaching speaking with more types of communicative activities (pair work, group work and presentation). Cheng compares the indicators of teaching methodology (teacher talking time, types and length of tasks in class) before and after the introduction of the HKCEE. It shows that although the teachers demonstrate a positive attitude towards the test, these indicators do not change substantially. Then, she suggests that the change brought by a new test is very superficial, and that teachers' core beliefs in teaching English have not changed significantly. Wall and Horak's (2000) statement about superficial change in the process of innovation helps to support Cheng's findings: an innovation may require change on three different levels: content, methodology and attitudes. It is easier for teachers to change the content of their teaching than to change their behaviour and easier for them to change their behaviour than to change their attitudes or values (2000, p. 503). In contrast, in Wall and Alderson's study, although the classroom observation shows connections between the newly-designed O-level English test and English teaching in terms of teaching materials, they still find that there are no relationships between the methodology a teacher uses and the new test (1993, p. 62). They believe that the reason for this was teachers' poor understanding of communicative teaching techniques and their not knowing how to deal with the new textbooks. Watanabe's (1996) investigation in Japan points out that a teacher's academic background and existing beliefs about English teaching and learning limit the washback effect on teaching. Finally, Shohamy's (1996) research in Israel compares the washback effects of the ELT test and the Arabic test on teaching. She notes that as the status of the language (Arabic) decreases for political and economic reasons, the washback effects on the (Arabic language) teaching method decrease accordingly.

The hypothesis, ‘a test will influence attitudes to the content and methods of teaching’, has also been examined in different contexts. Li (1990) observes that, in the face of a new test, English teachers in China generally had a positive attitude and believed the new test would definitely be helpful in assisting their methodology reform at the classroom level. In contrast, Cheng’s studies in 1997 and 1998 show that although most of the English teachers in Hong Kong held a positive and welcoming attitude to the new test, teachers’ beliefs in language teaching changed little. For example, their English classes were still teacher-centred, lecturing and explaining grammar still took up most of the instructional time.

6 Hypotheses Relevant to Learning

In Alderson’s fifteen washback hypotheses, apart from the second hypothesis, a test will influence learning, there are another five hypotheses relevant to learning. As mentioned previously, it appears that fewer studies have aimed to investigate the effect of washback on learning. Three out of nine studies discuss the washback effects on what learners learn, and four studies examine if there is a relationship between the test and the learners attitudes to testing/learning/learning strategies.

Andrew, Fullilove and Wong’s investigation (2002) on the Hong Kong Advanced Supplementary Use of English Oral Examination provides some evidence of the washback effects on what learners’ learn. The research, comparing the transcripts of candidates’ performance in the oral test with the exercises written in the textbook and commercial test-guidance books, concludes that the UE Oral test is exerting some influence on students’ performance in spoken English. However, the learning outcomes vary significantly. Andrews adds that the test may have led to improved performance, but in others only a superficial learning outcome, such as the ability to conform to the requirements of exam format, or to produce memorized phrases (2002, p. 207). Regarding learners’ attitudes, Li (1990) mentions that the high-stakes test MET seemed to bring the learner back to learning English, and there seems to be a new awareness of time and resources and a new enthusiasm for after-class learning of English (1990, p. 401). This change is not that big and perhaps still at the superficial level, but she insists that as time goes on, there will be a more marked tendency and a more penetrating change (1990, p. 402). In conclusion, previous empirical studies show that washback effects from a high-stakes test on learning may have a superficial outcome, but it is far from substantial improvement.

7 Other Hypotheses

The remaining hypotheses, which are relevant to the nature of the test, or to the uses to which scores will be put (Alderson & Wall, 1993, p. 120), have rarely been examined. Just one study, that is, Shohamy’s research in Israel, compared two

examinations, the Arabic test and the EFL test, and then indicated that the higher-stakes test (English test) had more significant washback effects than had the lower-stakes test.

According to Shohamy's investigations (2001) in Israel, as a part of the Israel Matriculation test, the EFL oral test is very high-stakes, which exerts a continuing influence on teachers, learners and even the other stake-holders of the test, such as the commercial test guidance book publisher, and the language learning software company. In contrast, the Arabic language test was placed in a subordinate position as a low-stakes test. The washback effects from the Arabic language test on both teachers and learners are clearly observed at the very beginning of its introduction. However, as time passes, a decreasing number of stake-holders (teachers, learners, schools, parents, and commercial textbook publishers) maintain their interest in this subject. Therefore, Shohamy concludes that the stakes of the test and status of the language will influence the washback effects on teaching and learning, in terms of the instruction content, methodology and their attitudes.

8 Research Design of Washback Studies

The research design of a washback study has been an issue for years (Bailey, 1996) as the researchers have been confronted with the challenge of determining how much of what takes place in a language class can be evidentially linked to the introduction or use of a high-stakes test (Wall & Horák, 2006, p. 13; Messick, 1996, p. 24; Alderson & Wall, 1993, p. 17). The book *Washback in Language Testing: Research Contexts and Methods* summarises that there are two main areas of washback or backwash studies: the first type refers to those relating to traditional, multiple-choice, large scale tests, which are perceived to have had a mainly negative influence on the quality of teaching and learning (2004, p. 3); and the second type refers to those studies where a specific test or examination has been modified and improved upon in order to exert a positive influence on teaching and learning (2004, p. 6).

In relation to the first type, the cause-effect link between the test and teaching is suggested by the observed differences of one English teacher teaching test-preparation courses and ordinary English courses, or a group of teachers teaching the same test-preparation course. The approaches such as classroom-observation and an after-class interview or questionnaire are employed to identify the relationship between the test and certain teaching behaviours. The researcher always observes different types of English classes (test preparation courses and regular academic courses) taught by the same teacher. If one teacher teaches the test preparation and regular academic courses in a similar way, then, it is not possible to suggest the existence of washback effects. However, if there is a washback effect on teaching, the teacher may employ a different method in different courses. For instance, Alderson and Hamp-Lyons (1996) investigate the washback effects from previous versions of the TOEFL on teachers in an American language centre. In order to remove the

influence from teachers own pedagogical beliefs, educational background and other factors, this research compares two teachers who teach TOEFL preparation courses and non-TOEFL courses. The results show that one teacher teaches TOEFL and regular academic courses in a similar way, whereas the other one employs two different methods. Then, the researchers conclude that their study shows clearly that the TOEFL affects both what and how teachers teach, but the effect is not the same in degree or in kind from teacher to teacher (1996, p. 295). Therefore, the result infers that the TOEFL alone does not cause washback. Hayes and Read (2004) examine the washback effects on the teaching of IELTS in two language courses: one clearly test-focused and the other with a stronger EAP orientation (2004, p. 99). Through interviewing and observing teachers, the study concludes that there is clear evidence of washback effects in the sense that teacher and students were narrowly focused on practice of the test tasks rather than the development of academic language proficiency in a broader sense (2004, p. 111). However, the teaching from a more experienced teacher who used to be an IELTS examiner seems to be more test-oriented than the less experienced teacher who held an MA in linguistics. Hayes and Read (2004) suggest that the level of teaching experience and educational background are the reason for different washback effects on teachers.

The second type of study usually includes a baseline study before the official introduction and implementation of the test. By comparing the differences, such as classroom teaching methods, teachers' explanations of their attitudes and behaviour, classroom assessment, and teaching materials, researchers suggest the existence of washback effects. Newly introduced tests in washback studies refer to those tests that have been introduced as a lever to initiate top-down reform, such as the Hong Kong Certificate of Education Examination (HKCEE) in Cheng's (1997, 1998, 1999) studies, Hong Kong's Advanced Supplementary (AS) Use of English (UE) oral examination in Andrew's (2004) study and impact study of New TOEFL in Central and Eastern Europe in Wall and Horák's research (2006, 2008).

Although previous impact and washback studies in general education present a complex relationship between a high-stakes test and teaching practice at the classroom level, many empirical studies in ELT try to build one-to-one connections between some aspects of teaching (e.g., methods, teaching materials, teachers beliefs, motivation) and a high-stakes test. Without offering any empirical evidence, Prodromou (1995) states that there is no one-to-one relationship between tests ... and their effect on the classroom, ..., before a test has an impact on classroom practice, it is mediated by factors such as the place of examinations in particular societies, the teacher's competence, and the resources available within the school system (1995, p. 13). This limitation in washback studies has still not been fully addressed as Saville and Hawkey (2004) again point out: Test washback, limited in scope to effects on teaching and learning, cannot really be substantiated without full consideration of the social consequences of test use (2004, p. 75). The remainder of this chapter reviews relevant studies that identify mediating factors of washback effects and impact of a test.

9 Recent Studies and Mediating Factors

The test, especially a high-stakes test, is frequently used to engineer innovation, to steer and guide the curriculum (Alderson, 2004, p. xi). Employing better kinds of assessment practices is widely recognized as a tool to promote better learning and teaching (Cheng & Curtis, 2004; Wei, 2014). Moreover, it is believed that the introduction of any test would necessarily lead to a washback effect, a single and uniform response (Burrows, 2004, p. 125) to every teacher and student. However, the idea that a test leads to good change or a test guarantees a desirable change, is under increasing criticism, as it is rarely the case that the consequences of testing are those that the designers intended (Fulcher, 2010, p. 6). Responding to the idea of using tests to make positive influence on classroom teaching and learning, Alderson (2004) notes that there are limits of what a test developer can achieve, and much more attention needs to be paid to the reasons why teachers teach the way they do (2004, p. xi). More explicitly, Watanabe (2004) summarises that one of the key findings of the research in the field to date is that washback is a highly complex rather than a monolithic phenomenon (2004, p. 19). In some recent washback studies, researchers (e.g., Saville & Hawkey, 2004; Burrows, 2004; Tsagari, 2009; Wei, 2014) provide increasing evidence to support this idea: such uniformity is unlikely to happen.

Burrows (2004), investigating a relatively low-stakes test in EFL immigration classes in Australia and its washback effects on classroom assessment, suggests that the teachers might not all respond to the new test in the same way and the test designers should take certain teacher variables (e.g. education background, training and working experiences) into account when designing implementation strategies for new tests. More recently, Tsagari (2009), analyzing 29 First Certificate in English-Cambridge ESOL (FCE) students' diaries, suggests that the analysis of the diaries revealed that the exam's influence was not direct but mediated through a variety of factors (2009, p. 7). Tsagari (2009) identifies at least three mediating factors in detail:

- (1) Mediated through textbook: In the absence of any official syllabus, the syllabus of the exam textbook became the course syllabus and determined the content of teaching and classroom assessment leading to textbook washback
- (2) Mediated through English teachers: The teacher reshaped the textbook in terms of methods, for example, by adding extra techniques and structuring the lessons of the day in her own way leading to teacher washback
- (3) Mediated through other stakeholders: learners, school atmosphere and parents.

Tsagari (2009) proposes a list of potential factors that she believes should be taken into account in the model of washback: textbook writers and publishers, teachers, students, schools, parents, the local education system and local society (2009, p. 8).

Aiming to investigate to what extent a high-stakes test can change the classroom teaching, Wall (2005; Wall & Horak, 2006, 2008) borrows Henrichsen's (1989)

framework of educational innovation and proposes four categories of implementation factors, which she believes can facilitate or hinder the success of the innovation (2005, p. 159). Apart from the textbooks and examination, there are two other categories: the first category is called the user system, by which she means the context that the innovation is being introduced into (2005, p. 203): classroom and school factors, education administration, and political, economical, cultural and geographic context. The second category is called the user, by which she means the basic abilities, beliefs and values (2005, p. 231) of teachers and students who are directly affected by the innovation. The factors she reports and discusses under the category of the users include the characteristics of the teachers and students, which propose a very comprehensive framework to analyse the teacher as a mediating factor (Burrows, 2004; Tsagari, 2009). The factors that are relevant to teachers are presented as follows:

1. The teachers' attitudes towards education
2. The teachers' attitudes towards the change in classroom teaching
3. The teachers' attitudes towards language teaching
4. The teachers' attitudes towards examinations
5. The teachers' attitudes towards English
6. The teachers' attitudes towards new ideas (openness)
7. The teachers' level of education
8. The teachers' abilities to teach the new curriculum
9. The teachers' personal lives revealed that to cope with the new textbooks and examination
10. The teachers' economic situation
11. The teachers' level of interest in the new curriculum
12. The teachers' goals at work

In relation to the new development on the washback effects on language learning, Wei (2014) investigated the links between test takers' awareness of the constructs of a newly introduced integrated skills tasks (PTE Academic) and learning strategies they employed to prepare for the test, concluding that test takers' choices of learning strategies can only be partly explained by their various levels of understanding and awareness of the new test. To be more specific, there appeared to be an association between language learners' uses of learning strategies and their awareness of what skills were being assessed and the nature of the audio input in the new test with integrated skills tasks.

10 The Future of Washback Studies

This chapter has observed that the trend of washback effects studies has gradually moved from asking whether washback exists to 'what does washback look like? What brings washback about? Why does washback exist?' (Alderson, 2004: ix).

The focus of the washback studies has moved from investigating simple cause-effect relationships between test and classroom teaching to a greater understanding of the whole much more complex picture. Many empirical studies have recently been conducted to identify the factors that mediate washback effects to classroom teaching and learning. Further studies may borrow a theoretical framework from other areas to empirically test the effectiveness of these potential mediating factors. For example, more studies may need to be conducted to assess the possible links between high-stakes tests and (1) teachers and learners motivation to teach and learn (Dornyei, 2001), (2) the development of language teachers' cognitions (Borg, 2006) and (3) learners' understanding, interpretations and responses to the results from high-stakes tests (William, 2000).

References

- Alderson, J., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*, 115–129.
- Alderson, J., & Hamp-Lyons, (1996). TOEFL preparation courses: A study of washback. *Language Testing*, *13*(3), 280–297.
- Alderson, J. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.), *Washback in language testing: Research contexts and methods* (pp. ix–xiii). London: Lawrence Erlbaum.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*, 260–267.
- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. *System*, *30*, 207–223.
- Bailey, K. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, *13*(3), 257–279.
- Bachman, L., & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London: Continuum.
- Brookhart, S. (2001). Successful students formative and summative uses of assessment information. *Assessment in Education*, *8*(2), 153–169.
- Burrows, S. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian Adult Migrant English Program. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 113–128). London: Lawrence Erlbaum.
- Cheng, L. Y. (1997). How does washback influence teaching implications for Hong Kong. *Language and Education*, *11*(1), 38–54.
- Cheng, L. Y. (1998). Impact of public English examination change on students' perceptions and attitudes towards their English learning. *Studies in Educational Evaluation*, *24*, 279–301.
- Cheng, L. Y. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, *15*, 253–271.
- Cheng, L., & Curtis, (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3–18). London: Lawrence Erlbaum.
- Chudowsky, N., & Behuniak, P. (1998). Using focus groups to examine the consequential aspect of validity. *Educational Measurement: Issues and Practice*, *17*(4), 28–38.
- Dornyei, Z. (2001). *Teaching and researching motivation*. Harlow, England: Pearson Education.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.

- Gronlund, N., & Waugh, C. (2009). *Assessment of student achievement* (9th ed.). New Jersey: Person Education Ltd.
- Gorsuch, G. (2000). EFL educational policies and educational cultures: Influences on teachers' approval of communicative activities. *TESOL Quarterly*, 34(4), 675–710.
- Hall, L., Kleine, & Paul, F. (1992). Educators' perceptions of NRT Misuse. *Educational Measurement: Issues and Practice*, 11, 18–22.
- Halpin, G., & Halpin, G. (1982). Experimental investigation of the effects of study and testing on student learning, retention, and ratings of instruction. *Journal of Educational Psychology*, 74(1), 32–38.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365–379.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *The Curriculum Journal*, 16(2), 203–207.
- Hayes, B., & Read, J. (2004). IELTS testing preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 97–112). London: Lawrence Erlbaum.
- Henrichsen, L. E. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan, 1956–1968*. New York: Greenwood Press.
- Herman, J., & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement: Issues and Practice*, 12(4), 20–25.
- Herman, J., Webb, N., & Zuniga, S. (2007). Measurement issues in the alignment of standards and assessments. *Applied Measurement in Education*, 20, 101–126.
- Huhta, A., Paula, K., & Pitkanen-Huhta, A. (2006). Discursive construction of a high-stakes test: The many faces of a test-taker. *Language Testing*, 23(3), 326–350.
- Kellaghan, T., Madaus, G., & Airasian, P. (1982). *The effects of standardized testing*. Boston and London: Nijhoff Publishing.
- Kennedy, K., Chan, J., Fok, P., & Yu, W. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice*, 7(3), 197–207.
- Li, X. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual and Multicultural Development*, 1(15), 393–404.
- Madaus, G. (1988). *The influence of testing on the curriculum. Critical Issues in Curriculum: Eighty-seven yearbook of the national society for the study of education*. Chicago: The University of Chicago Press.
- Messick, S. (1989) Validity. In LINN, R. L. (Ed.), *Educational measurement*. New York, Macmillan.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241–256.
- Mehrens, W., & Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 45(1), 14–22.
- Power, D., & Alderman, D. (1983). Effects of test familiarization on SAT performance. *Journal of Educational Measurement*, 20(1), 71–79.
- Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, 49(1), 13–25.
- Qi, L. X. (2004). Has a high-stakes test produced the intended changes? In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 171–190). London: Lawrence Erlbaum.
- Qi, L. X. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing*, 22(2), 142–173.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28, 4–13.
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23, 146–159.
- Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 145–165.

- Saville, N., & Hawkey, R. (2004). The IELTS impact study: Investigating Washback on teaching materials. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 73–96). London: Lawrence Erlbaum.
- Shepard, L., & Bliem, C. (1995). Parents' thinking about standardized tests and performance assessments. *Educational Researcher*, 24(8), 25–32.
- Shohamy, E. (2001). *The power of test—a critical perspective on the users of language tests*. Essex England: Pearson Education Limited.
- Shohamy, E., Donitsa-Schmidt, A., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298–317.
- Smith, M. L. (1991a). Put to the test: The effects of external testing on teachers. *Educational Researcher*, 20(5), 8–11.
- Smith, M. L. (1991b). Meanings of test preparation. *American Educational Research Journal*, 28(3), 521–542.
- Smith, M., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, Winter, 10(4), 7–11.
- Sternberg, R. (2007). Culture, instruction, and assessment. *Comparative Education*, 43(1), 5–22.
- Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education*, 9(2), 172–192.
- Tsagari, D. (2009). *Revisiting the concept of test washback: Investigating FCE in Greek language schools*. University of Cambridge ESOL Examinations Research Notes, March 35.
- Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.
- Wall, D., & Horak, T. (2006). *The TOEFL impact study: Phase 1. The baseline study*. TOEFL Monograph 34. Princeton, NJ: Educational Testing Service.
- Wall, D., & Horak, T. (2008). *The TOEFL impact study: Phase 2. Coping with change*. TOEFL iBT Research Series No.05. Princeton, NJ: Educational Testing Service.
- Wall, D., & Horak, T. (2000). Using baseline studies in the investigation of test impact. *Assessment in Education*, 14(1), 99–116.
- Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *Language Testing*, 13(3), 318–333.
- Watanabe, Y. (2004). Methodology in washback studies. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 19–36). London: Lawrence Erlbaum.
- Wei, W. (2014). Can integrated skills tasks change students' learning strategies and materials? *Language Learning Journal*. doi:10.1080/09571736.2014.905970.
- Wideen, M., O'Shea, T., & Ivory, G. (1997). High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4), 428–444.
- William, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative from summative functions of assessment. *British Educational Research Journal*, 22(5), 34–37.
- William, D. (2000). Integrating summative and formative functions of assessment, Keynote address to the European Association for Educational Assessment, Prague: Czech Republic, November 2000. Retrieved November 11, 2013 from http://eprints.ioe.ac.uk/1151/1/Wiliam2000IntegratingAEA-E_2000_keynoteaddress.pdf.
- Zeng, K. (1999). *Dragon Gate: Competitive examinations and their consequences*. London: Cassell.

Developing and Sustaining Outcomes Assessment in English as a Foreign Language Programs

Donald F. Staub

Abstract Educational organizations are charged with one critical task: effectively and efficiently ensuring student learning. Traditionally, the determining factor for whether educational institutions had imparted knowledge on their students was simply to count the number of graduates. English as a Foreign Language (EFL) programs have followed this tradition, equating quality with numbers of successful program completers. Over the past two decades, the so-called accountability movement has put increasing pressure on schools to demonstrate quality by evidencing student learning through the assessment of learning outcomes. EFL programs are increasingly being asked to develop and implement learning outcomes assessment programs. To do so, however, can be arduous, and, if not approached thoughtfully, can lead to failure. This chapter explores the principles and practices that are generally believed to be must-haves for successful outcomes assessment programs. This is followed by a discussion of common pitfalls that lead to failure of such initiatives. Finally, the chapter proposes that EFL program leaders who are embarking upon an outcomes assessment process consider the Distributed Leadership model as a means for increasing the probability of success and sustainability of their outcomes assessment initiative.

Keywords Assessment · EFL · Outcomes

1 Introduction

Educational organizations are charged with one critical task: effectively and efficiently ensuring student learning. Historically, the prevailing assumption was that if a student had completed the required coursework, then they had mastered all relevant content, and the institution was perceived as having successfully executed its role. In the case of many English as a Foreign Language (EFL) programs, verifi-

D.F. Staub (✉)
Isik University, Istanbul, Turkey
e-mail: staubdonald2@yahoo.com

cation of competence is generally through a single, end-of-year high-stakes exam. Thus, the quality of the EFL program has traditionally been measured solely by the number of students who successfully pass the exam. Over the course of the past two decades, this belief has been subject to much scrutiny, driven by calls for greater accountability by internal (e.g. governing boards, administrators, faculty members) and external (parents, students, politicians, and taxpayers) stakeholders, who are demanding solid evidence of learning—not simply evidence of teaching. A tangible outgrowth of this movement has been the emergence of outcomes assessment as a means for substantiating learning. This has inspired a substantial body of literature providing detailed discussions of relevant principles and practices of outcomes assessment (e.g. Baker, Jankowski, Provezis, & Kinzie, 2012; Banta, Jones, & Black, 2009; Maki, 2004; Suskie, 2010; Walvoord, 2010).

As is often stated, the goal in implementing any outcomes assessment initiative is the establishment of a process that consistently fosters improvement in student learning. While this may look good on paper, the stark reality is that examples of schools demonstrating a closed assessment loop—from design to implementation to analysis and action to consistent improvement in learning—are difficult to come by (Banta & Blaich, 2011; Hutchings, 2010; Miller, 2012). The absence of cases exemplifying success has been attributed to a number of barriers and missteps. Examples range from schools that focus too closely on assessment for accreditation rather than learning (Hersh & Keeling, 2013), to failing to turn data into action (Blaich & Wise, 2011; Bresciani, 2012), to insufficient faculty involvement (Bresciani, 2009; Hutchings, 2010), to educational organizations themselves not knowing how to learn very well (Tagg, 2007). Meanwhile, the literature on outcomes assessment can be broadly characterized as focusing on the macro level, such as a system or district (Lennon et al., 2014; Bresciani, 2009a) or an institution (Blaich & Wise, 2011; Maki, 2004).

Where program-level guidance does exist (e.g. Bresciani, 2009), it is generally not directed at any particular field; even more rare is a discussion of outcomes assessment for leaders of English as a Foreign Language (EFL) programs. Yet, there are two separate driving forces that should inspire leaders of EFL programs to more seriously consider the virtues of a rigorous outcomes assessment system. First, the aforementioned accountability movement is gathering momentum globally; accreditation and quality assurance, and, by extension, rankings, are now part of the day-to-day lexicon of higher education. That EFL programs will be held accountable for quality is not an issue of *if*, but *when*. Second, EFL programs provide a value-added service in a crowded higher education market place. As university market shares are shaved off by competitors, demonstration of quality will replace the mere existence of an EFL program; an associated issue related to competition and program quality is student retention. Certainly, EFL program leaders cannot simply flip a switch and expect an outcomes assessment program to power up. As mentioned, EFL programs, similar to most higher education programs, do not have a tradition of assessing learning outcomes. Therefore, the aforementioned principles for effectiveness and efficiency require careful consideration and strategizing at the developmental stage in order to ensure successful and

sustained implementation and a continuous cycle of improved student learning. In other words, while these principles and practices may be instructive to EFL program leaders, their application at the program level may appear intimidating.

It is for this reason that a Distributed Leadership model (e.g. Spillane, Diamond, & Halverson, 2001, 2004) may offer a framework worth exploring for EFL leaders faced with the task of implementing and sustaining an effective outcomes assessment system. The Distributed Leadership (DL) model offers a unique, but arguably compelling perspective on examining leadership. While most leadership models explore the personalities and actions of individuals, the DL model views the activity of leadership as the focal point. In doing so, DL posits that leadership is not the result of one individual's actions, but rather a complex web of social interaction between the leader, followers, and the situation (Spillane et al., 2001, 2004). Therefore, whether analyzing or planning an initiative, the focus becomes how leadership is, or is not, diffused throughout a unit. For an EFL program leader, the question shifts from "How am I going to make this happen?" to "How can I facilitate successful implementation and sustainability of this initiative?"

This chapter will review the most commonly accepted principles and practices in the outcomes assessment literature today. This will be followed by a discussion of the barriers that seem to be inhibiting successful implementation of outcomes assessment programs. The final section will explore the concept of Distributed Leadership, particularly given the backdrop of the barriers discussed in the previous section, and present it as a viable framework for EFL leaders to consider when implementing outcomes assessment programs.

2 Principles and Practices

Educational organizations are charged with one critical task: effectively and efficiently ensuring student learning. In the past, this meant that the institution enrolled the student at the beginning of his or her academic endeavor, provided a list of courses for that student to take while at the institution, and hoped that the student graduated at the other end. While this approach was sufficient in the past, over the last two decades this philosophy has changed dramatically. Internal and external stakeholders have exerted pressure on educational institutions to demonstrate that their students are not merely going to class, but that they are learning. Schools have been required to respond by devising systems to demonstrate that they are actually paying attention to what students are purportedly learning. And, if the students are not acquiring the knowledge, skills, and experience that they were promised in the first place, then the onus is on the institution to make appropriate changes to improve the student's opportunity for educational success.

Given this context, the outcomes assessment movement has garnered increasing attention over the last two decades (e.g. Angelo, 1999; Banta, 1993, 1996; Cross, 1998; Ewell, 1988; Palomba & Banta, 1999). The previous metrics, or *outputs* (e.g. students matriculated, students graduated, grade-point-averages) are no longer

sufficient to determine whether an institution has provided value-added to its students (e.g. Angelo, 1999; Kuh & Ikenberry, 2009). Today, higher education institutions must provide evidence that their students are demonstrating achievement of specified learning outcomes, as identified and monitored at the institution, program, and classroom levels. Learning outcomes measure changes in students' knowledge, skills, and behaviors over time—*vis a vis* the unit of analysis (i.e. institutional, program, or course level). In order to carry out an effective and efficient outcomes assessment program—that is, to consistently monitor and improve learning at all levels—educational institutions must carefully plan, implement, and work to sustain such initiatives. With this increase in attention toward outcomes assessment comes the need for guidance, which in turn has inspired a substantial body of literature providing useful direction concerning relevant principles and practices of outcomes assessment (e.g. Baker et al., 2012; Banta et al., 2009; Maki, 2004; Suskie, 2010; Walvoord, 2010).

Some sources highlight specific principles or practices, such as communication and sharing evidence (Blaich & Wise, 2011), planning (New Leadership Alliance, 2012), or meaningful, measurable, and mission-driven assessment (Baker et al., 2012). Others provide more comprehensive, detailed component descriptions and recommendations (e.g. Banta et al., 2009; Bresciani, 2009, 2012; Maki, 2004). Banta et al. (2009) outline the three Phases of Assessment: Planning, Implementing, and Improving and Sustaining. When planning the implementation of an outcomes assessment initiative, for an EFL program, particularly if such a system is non-existent, then it is certainly advisable to break the principles and practices into these three progressive stages.

3 Planning

Planning is what EFL program leadership must embark upon as early as possible. To begin, an EFL assessment committee should be constituted. It is particularly important to have a program-level committee as this is where responsibility for assessment resides (Banta & Blaich, 2011). These committees become the face and voice of assessment as the initiative is developed and begins to spread throughout the organization. Bresciani (2009) provides a useful list of guiding questions to be considered during the formation of the assessment committee, such as who will be on the committee, for how long will they be on the committee, and what support or rewards will they receive for membership. Certainly, this will vary depending on the size and structure of the program and workload distribution. This is where leadership commitment, in the form of providing time and resources to those who will enact the initiative, is crucial.

The assessment committees work with relevant stakeholders in order to identify expectations for student learning (Maki, 2004) which lead to the generation of assessment questions (Blaich & Wise, 2011) and ensure that assessment is meaningful, manageable, and mission-driven (Baker et al., 2012). Naturally, stakeholder

groups would include the students themselves, the faculty members who are teaching core subjects to the EFL program completers, and potential employers. The EFL assessment committee will also facilitate the process of establishing a common language for the initiative (e.g. what is a goal versus an objective) and, importantly, a “shared conceptualization” of why the program is undertaking the establishment of an outcomes-based assessment program (Bresciani, 2012). The plans devised by the committee include specification of how evidence and changes will be disseminated on a regular basis. Transparency of the process is often referred to as a critical factor in ensuring success of an outcomes initiative (Blaich & Wise, 2011; Bresciani, 2012; Maki, 2004; New Leadership Alliance, 2012; Jankowski & Provezis, 2011).

4 Implementation

Implementation is the subsequent phase. At this stage, the EFL program assessment committee moves from input gathered from stakeholders and extant data to identification of a specific set of learning outcomes that they wish to measure. They also devise an assessment plan for each outcome, keeping in mind that it is not necessary to assess all outcomes every year (Bresciani, 2009). Attempting to do so may prove burdensome from a workload perspective, as well as overwhelming and demoralizing to those responsible for collecting, analyzing, and reporting results. As individual outcomes are identified and defined, and assessment plans are devised, it is instructive to bear in mind the SMART acronym:

- **Specific**—the outcome should specify the group that should be achieving the outcome. The outcome should also only assess one specific skill, behavior, or ability. For example, the outcome “Students will read and summarize a text” is, in fact two separate outcomes. One outcome will assess the students’ ability to read and comprehend a text. The second outcome will assess their ability to summarize the text. The more specific the outcome, the greater the chance of identifying the root cause of any issues.
- **Measurable**—the outcome clearly identifies a numerical value that will change as a result of learning. For example: “75 % of Track 3 Writing students will receive 3 or higher on the 5-point rubric for the summarization exercise”. Banta and Blaich (2011) recommend multiple measures to increase reliability and validity. For example, in order to determine whether students have mastered the skill of summarization, there may be multiple exercises assessed over the course of a term, along with a summarization item on a final exam.
- **Achievable**—the target indicated should be realistic and attainable within the given learning period. If only 50 % of students successfully completed a lecture note-taking exercise last semester, it may be unrealistic to expect 75 % to do so this semester. The more often assessments are conducted, the more realistic the projected targets will be.

- Relevant—is the outcome aligned with the vision, mission, and goals of the program? Is the outcome based on stakeholder input, or on reliable data indicating an area worthy of focus?
- Time Frame—the time period for learning and assessment are defined. The wording of the outcome should specify the amount of time in which the skill, knowledge, or behavior should be acquired; e.g. One semester, or by the time the student has completed the program.

Beyond identification and definition of individual outcomes, the implementation phase includes the following steps: assessment of the outcomes, analysis of data, establishment of action plans, and reporting of results and plans. Completion of this cycle is commonly referred to as “closing the loop”, and it is generally facilitated by instructors and staff who have a firm understanding of the process. Bresciani (2009) notes that it is important to have a support structure in place to provide assistance to faculty members who are responsible for these steps. Assistance may manifest as access to professional development or participation in conferences, or due compensation for their efforts, such as overtime pay or release time. Likewise, the support may be in the form of technology that can assist with analysis, storage, and reporting of results.

5 Improving and Sustaining

Improving and Sustaining an outcomes assessment program are the hallmarks of a successful outcomes assessment system. As will be discussed below, this is quite often the phase that remains out of reach. In this stage, institutions and programs have established outcomes systems where, on a regular basis: outcomes are assessed, data is collected and analyzed, evidence-based changes in programs and practices are devised (New Leadership Alliance, 2012), and ultimately there is evidence of improved student learning and improved efficiency in processes. One of the keys to sustainability is consistent communication and improvement through the utilization of results (Banta & Blaich, 2011; Bresciani, 2012; New Leadership Alliance, 2012). Similarly, the EFL assessment committee must ensure “an entirely public process” where assessment evidence is “widely shared and discussed on campus” (Blaich & Wise, 2011, p. 12), which may come in the form of faculty-led forums and the posting of results of dialogues on a website (Maki, 2004). The National Institute for Learning Outcomes Assessment at the University of Illinois at Urbana-Champaign has established an initiative to push for greater transparency of outcomes assessment reporting. Other keys to sustainability are ongoing faculty development and the establishment of an environment of trust (Banta & Blaich, 2011; Banta et al., 2009; Bresciani, 2012).

As previously mentioned, one way to conceptualize the process of developing, implementing and sustaining a program of outcomes assessment is through the three-phase approach (Banta et al., 2009). I would like to propose that EFL program

leaders envision the process through a different framework—what I refer to as the Environment approach—whereby EFL program leaders can analyze their institution from the perspective of organizational culture and determine where barriers to development and implementation may arise. This approach analyzes the institution in terms of an Enabling environment, an Attractive environment, and a Sustainable environment. The Enabling environment essentially asks whether institutional leadership is receptive and supportive (Banta & Blaich, 2011; Banta et al., 2009), whether in the form of resources (e.g. human, technological) or in the creation of an atmosphere of trust that is non-threatening (Maki, 2004).

An Attractive environment provides structural features that encourage faculty and staff to engage in the outcomes assessment process. Outcomes work that is perceived as confusing, burdensome, or not aligned with the goals and needs of the EFL program often dissuades any engagement in the process. In contrast, release time, rewards, targeted professional development opportunities (Banta & Blaich, 2011; Bresciani, 2012), and encouraging research and scholarship that is aligned with the outcomes assessment initiative (Baker et al., 2012) may strengthen buy-in from instructors and staff.

The Sustainable environment emerges when the assessment initiative no longer relies on an individual or committee. It is apparent when assessment and data drive conversations for change and improvement, rather than the necessity for change coming from an external source. Another indicator is when anecdotes are rejected and data is accepted as the only valid evidence. The EFL program that collects, analyzes and then shares data regularly with its stakeholders, and invites their input in a continuous pursuit of improvement has fostered the Sustainable environment. As Banta et al. (2009, p. 3) concisely put it, effective assessment emerges over time. Whether educational leaders wishing to establish an outcomes assessment system view this task from the three-phase approach, or the Environmental framework—or a combination thereof—it is critical that they assume a long term view of the process. Rushing a process may result in short-term success, but there is a good chance of long term failure as people feel overwhelmed and under-motivated. In the end, it will be the students who pay the price. A consistent, persistent long-term approach to development and implementation of an outcomes assessment program will increase the probability of success.

6 Barriers

With so much good advice available, why are improvements in student learning as a result of assessment the exception rather than the rule? (Banta & Blaich, 2011).

An outcomes assessment program provides a valuable means for improving student learning at the institutional and program levels. In order to do so, an organization or a program must develop an outcomes assessment program that effectively and efficiently closes the loop on the assessment cycle—on a continuous

basis. That is, data is collected, analyzed, and acted upon to improve learning. There is, however, relatively little evidence that schools are experiencing success in closing the loop (Banta & Blaich, 2011; Hutchings, 2010; Miller, 2012). Despite the plentiful availability of sources that carefully define and describe best principles and practices for establishing outcomes assessment programs, a growing body of literature is substantiating the reasons why schools struggle with achieving success in implementation. The major challenges that have been cited range from focusing on outcomes for accountability rather than improvement, an inability of schools themselves to engage in deep learning for substantive change, failing to convert results from data into action, and, overall, institutional cultures that do not value collaboration and transparency for the sake of improved learning.

Hersh and Keeling (2013) allude to what is perhaps the most prevalent reason for the lack of success in implementing outcomes assessment programs: institutions responding to external demands for accountability. Often, such responses to external bodies are transactional, generating little systematic or systemic change (p. 4). Banta and Blaich (2011) concluded that the indifference to action is perpetuated by the belief among many faculty members that assessment is an “externally motivated and bureaucratic process” (p. 24), which minimizes time with students. In other words, these institutions and programs have failed to generate *transformational* or deep change that would lead to perhaps completely different approaches to delivery of instruction. Tagg (2007) refers to this as double-loop learning (citing Argyris & Schön, 1978). As opposed to examining the “governing values” (p. 38) behind the policies and processes that may actually be the root cause of ineffective change, institutions justify the status quo by relying on defensive routines and refusing to publicly report performance results. The condition that Tagg is alluding to is able to persist because many institutions lack a culture that collaboratively examines student learning—from its design to its assessment—what Hutchings (2010) points to as the absence of faculty involvement in the process, which could be explained in part by the “excruciatingly slow” work on common learning outcomes (Miller, 2012).

Blaich and Wise (2011), in attempting to uncover why so many institutions with ostensibly successful outcomes assessment programs have not transformed learning, determined that a major issue is the translation of data into action. The common procedure is for institutions to gather data and simply circulate results among a small group, and then “shelve them if nothing horrible jumps out—and sometimes even if it does!” (p. 12). The issue is that assessment data gathering is not followed by faculty presentations on the nature of the data, nor with faculty-driven discussions about how to respond. Blaich and Wise go on to point out that unless the data reveals something “truly devastating” there is little to no response. Bresciani (2012), in her case study, attributes an unsuccessful assessment program to ineffective communication in the planning stage. She realized that the major pitfall was that key stakeholders had not agreed on a “shared conceptualization” of what metrics or data would be collected, or with which audience they would share the data.

7 Distributed Leadership

Measuring and assessing learning outcomes is critical to ensuring that students have successfully mastered the skill, competency or knowledge. But where and how this is done is still an underdeveloped area (Lennon et al., 2014, p. 10).

English as a Foreign Language (EFL) programs, similar to most higher education programs, do not have a tradition of assessing learning outcomes. For EFL program leaders who wish to implement and sustain a successful outcomes assessment program, the task may seem intimidating when considering the lengthy list of principles and best practices prescribed by the literature. Yet the evidence indicates that few schools and programs establish successful outcomes assessment programs (e.g. Hutchings, 2010). If we boil down the barriers to success, we are left with: “little or no collaboration...”; “insufficient shared planning...”; “no transparency with data...”; “ineffective communication...” (Banta & Blauch, 2011; Hutchings, 2010; Miller, 2012). In other words, we are left with the notion that many organizations do not have people that can talk and work with each other—for the good of the students.

It is for this reason that Distributed Leadership (DL) may offer a framework worth exploring for EFL leaders faced with the task of implementing and sustaining an effective outcomes assessment system. Distributed Leadership (DL) has its roots in primary and secondary education in the United States, however it is gaining broader appeal across the educational spectrum and in different countries. (e.g. Bolden, Petrov, & Gosling, 2009; Jones, Lefoe, Harvey, & Ryland, 2012; Pont, Nusche, & Hopkins, 2008; Van Ameijde, Nelson, Billsberry, & Van Meurs, 2009). While most leadership models explore the personalities and actions of individual leaders, DL views the *activity* of leadership as the focal point. Spillane et al. (2001, 2004) posit that leadership is not the result of one individual’s actions, rather it is a complex web of social interaction between the leader, followers, and what they refer to as the situation. Therefore, when planning an outcomes assessment initiative, the focus becomes the way in which leadership is diffused throughout a unit. In the case of an EFL program leader, the question shifts from “How am I going to make this happen?” to “How can I facilitate successful implementation and sustainability of this project?” Two similar, yet distinct, queries.

The distributed perspective of leadership calls into question the generally accepted notion that leadership is the exclusive domain of those in leadership positions, such as a president or the head of a department. Rather, Distributed Leadership, as the name suggests, is derived from the idea that leadership emerges from the efforts of a variety of individuals within an organization—both positional and non-positional leaders. The central notion is that leadership is manifested when a leader’s cognition is stretched, or distributed situationally, over aspects and actors (Spillane & Sherer, 2004). Actors, according to the authors, may be both leaders and followers, for without followers a leader cannot lead. Therefore, the focus of leadership shifts from a single individual to the “interplay between the actions of multiple people” (p. 37) utilizing particular tools and artifacts within a particular

situation. Spillane et al. (2004, p. 25) explain that this “collective leading requires multiple leaders working together, each bringing somewhat different resources—skills, knowledge, perspectives—to bear”. In sum, the unit of analysis is not the leader but the activity and the “web” of leaders, followers, and the situations that constitute the leadership itself. In the specific case of the EFL program embarking on an outcomes assessment process, the Distributed Leadership framework may be compelling because of the large size and structure of most EFL units, in addition to the potential for expanding and deepening engagement across the program.

As mentioned throughout this chapter, there is a respectable body of literature dedicated to the principles and practices associated with successful outcomes assessment programs. (e.g. Angelo, 1999; Banta et al., 2009; Bresciani, 2009, 2012; Maki, 2004). In contrast, there is a body of literature indicating that evidence of successfully implemented initiatives is scant. As Hutchings (2010) has concluded, “Unfortunately, much of what has been done in the name of assessment has failed to engage large numbers of faculty in significant ways” (p. 3). This sentiment has been echoed by Hersh and Keeling (2013), who lament that too often assessment is “orphaned to the province of a small group of dedicated faculty and staff” (p. 9), which can easily lead to exhaustion and marginalization. Distributed Leadership works to adjust these imbalances by drawing a greater number of participants into such critical processes. This can be accomplished by examining the core concepts associated with Distributed Leadership (i.e. leaders, followers, cognition, and the situation), and exploring the ways in which Distributed Leadership may play out in the context of learning outcomes assessment in an EFL program.

8 Leaders and Followers

While it is critical that positional leaders (e.g. president, dean, department chair) are supportive of initiatives to assess outcomes (Banta & Blaich, 2011; Bresciani, 2009; Maki, 2004), it is also essential that other individuals within the organization or unit assume non-positional leadership roles in the assessment process. These may be EFL instructors leading working groups in the designing assessments or discussing results data. In one EFL program of 600 students in a university in Istanbul, Turkey, there is a testing office dedicated to the development and administration of placement and exit exams. However, all instructors are given the responsibility of developing a number of assessments that will determine 40 % of their students’ final grades. This allows instructors to have a greater understanding of the assessment process, while also giving them greater ownership in assessing their students. From this situation, a number of instructors have emerged as non-positional leaders in founding the program’s outcomes assessment process, while some others have assumed leadership positions in a recently expanded testing office. As the number of non-positional leaders grows, there is a corresponding increase in the number of followers who are brought along because of their

colleagues' influence. Likewise, there is an overall increase in engagement with the outcomes assessment process and student learning—a rising tide lifts all boats.

In an EFL program where top heavy leadership was once concentrated in three positional leaders, there is now much greater depth and breadth of leadership across the program. Indeed, the leadership of the three positional leaders has been distributed by spreading the influence of the non-positional leaders and their followers. This has resulted in much greater attention to student performance through collaboration among the teaching staff while attempting to make sense of and interpret assessment evidence (Banta & Blaich, 2011) and devise action plans.

In the lexicon of Distributed Leadership, the cognition of the positional and non-positional leaders has been stretched across the organization (i.e. the EFL program) as the collaboration expands. As Maki (2004) suggests, one can evidence a “collective commitment” via the *structures* (i.e. an expanded assessment office), the *processes* (i.e. shifting a percentage of student assessment responsibilities to classroom instructors), and the *practices* (e.g. no classes on Tuesday afternoons so that teachers have dedicated collaboration time) (emphasis mine).

In another instance from the Turkish university, an instructor took it upon herself to devise a survey to gauge professional development needs among the teaching staff. Based on the results of the survey she recruited instructors from within the teaching staff to provide professional development in areas where they had knowledge and experience. She then created a professional development calendar for the semester. In turn, those who provided the training sessions became the de facto go-to people in their respective areas of expertise. Thus, not only were the 1-h training sessions offered, but there was also the advent of a distributed resource center through the initiative of individual instructors.

Another instructor saw the need for an Academic Support Center for freshman students of English. Students needed assistance in writing papers and studying for TOEFL and IELTS. She began by offering her free time to her own students. After 1 year, there were five volunteer instructors who provide tutoring to any freshman student of English who requested assistance. The positional leader may or may not have perceived this need, but this instructor did, and she was not only able to found the center, but she has attracted a cadre of followers in the other volunteer instructors.

A final, and perhaps most relevant example for our purposes here, occurred when another instructor saw that outcomes assessment was not given much attention within the EFL program. The university had recently conducted some activities related to the Bologna process,¹ but there was no direction provided by

¹The Bologna Process is an effort by the European Union to create a European Higher Education Area. The primary goal is to develop a process of standardization across Europe that allows students and graduates have their degrees and transfer credits recognized throughout the Union. At the local level, each university desiring involvement in the process is required to assure quality of processes and learning. As these requirements make their way to the university and program levels, depending on the organizational culture, they may or may not come with explanation or training.

the institution, other than “complete this within 2 weeks.” In the case of the outcomes, at a staff meeting, the instructor provided an overview of the principles and the process of outcomes assessment, then asked if any of the instructors saw a need to assess any outcomes. Five instructors and two administrators expressed desire to begin the outcomes assessment process. Since then, all seven have been through one full cycle of the project and are currently developing action plans, based on their data collection and analysis. One of the projects, because of positive results, has moved from the pilot stage to program-wide implementation. Moreover, the EFL program recently began an accreditation process. Five of the seven who have been leading the outcomes assessment initiative are now on the leadership team for the accreditation process. Based on their experience with outcomes assessments, these five are able to effectively lead the accreditation process, as well as serve as ambassadors for the process and engage more followers. This may not have happened if the one instructor had not initiated the outcomes assessment process a year prior.

In sum, the positional leader of the program has enabled this environment, which allows instructors and staff to explore ideas and expand them into formal and informal entities that ultimately result in improved learning opportunities for the students of the EFL program. In addition, from a distributed perspective, Spillane and his colleagues describe this as a multiplicative rather than additive model (2004, p.16). That is, the interactions among two or more leaders in carrying out a particular task may amount to more than the sum of those leaders’ practice. Whether it is the positional or non-positional leaders, their cognition (vision and leadership) has been stretched (distributed) across the organization (EFL) via the interactions of these actors (leaders and followers).

9 The Situation

Spillane et al. (2004, p. 10) contend that Distributed Leadership has three essential constituting elements: the web of leaders, followers, and the situation. I have examined the actors, now it is important to explore the situation. The situation is comprised of the many facets of an organization that either enable or constrain a leader’s work. The situation may be the organizational culture or the structure—physical or organizational. It may be policies and procedures, or the symbols, tools, and other designed artifacts that are part and parcel of day-to-day leadership practice (p. 21). They further explain that a leader’s thinking and practice is mediated by these *artifacts*: they serve as constituting components of leadership practice, not simply as devices or means that allow individuals to do what they want to do (p. 23). Thus, when a leader (positional or non-positional) creates, for example, a memo, a report, a new policy, a new program, or a new office that is a means by which leadership is being distributed. Likewise, existing buildings, policies, and organizational structures provide conduits—and barriers—to leadership. As agents, leaders must choose to utilize the situation, or make efforts to

amend the situation in order to facilitate the distribution of their cognition (leadership). To that end, leaders in an EFL program working to establish a successful outcomes assessment initiative must keep in mind that the situation plays a critical role in determining how effectively they distribute their leadership. In other words, it is not simply interactions with others, but the means through which interactions take place. Likewise, the structures (physical and organizational), written documents, policies, emails, or celebrations mediate cognition—either stretching it or constraining it.

According to Spillane and his colleagues (Spillane et al., 2004), the situation is multi-dimensional. In the context of EFL outcomes assessment, the situation may include the organizational climate, including such indicators as degrees of inquiry, and accountability. The situation may also assume processes, such as the aggregation, disaggregation, and analysis of assessment data, the formulation of action plans based on results, as well as the eventual reporting of results, in either hard or soft copy. And, certainly, the situation may include structures such as a building or an office, an organizational hierarchy, and policies and positions. The most common situational aspects mentioned in the assessment literature include the need for continuous professional development (Blaich & Wise, 2011) and the need for structured time so that faculty can plan assessment in addition to analyze and reflect upon data (e.g. Hutchings, 2010) so that improvements in learning can be made.

When we revisit some of the EFL cases mentioned previously, we can see how the situation completed the Distributed Leadership triangle (leader + follower + situation) and contributed to the distribution of the leader's cognition. In the case of the instructor who started the professional development sessions, she was able to pursue her vision in the first place because of the collaborative culture (organizational culture) already in existence within the EFL program. She was also able to distribute her cognition by administering the needs-survey (an artifact) to not only gather data, but also to create awareness. She also took advantage of the fact that no classes are scheduled for Tuesday afternoons (structure), thus ensuring that there would be empty classrooms available (structure) and that instructors would be able to attend the training sessions. In the case of the instructor who developed the Academic Support Center for freshman students of English, she did not have the same physical structure as the instructor in our previous example, as they work in different buildings. Thus, she needed to create a center (a structure) by placing one small desk in the limited space in her own office, and asking each volunteer instructor to do the same (structure). Like the professional development leader, she took advantage of the collaborative culture (organizational culture) and the teaching schedule (structure)—instructors could choose to substitute 3 h per week in the center for one class. In this way, she was able to attract other volunteer instructors to expand the center. Finally, in the case of the instructor who developed the outcomes assessment project, he was aware that the Bologna Process and notions of quality assurance were in the minds of the instructors and staff. He also understood that some staff and instructors were open to change if it would improve student learning (organizational culture). He was also given permission (organizational culture) to utilize one staff meeting (structure) to introduce outcomes assessment.

Handouts and forms (artifacts) were created and distributed to ensure clarity of the project and the process. In all three of these instances, the leaders were able to navigate the organizational structure by avoiding barriers and discovering leverage points that “enable the movement and generation of knowledge” (Spillane et al., 2004, p. 27). These examples may appear common and uninteresting. However, if we return to the discussion of common barriers to successful implementation, we can see that the sharing, collaboration, and transparency that are missing from the unsuccessful examples are prevalent in these examples. Thus, one could argue that the unique feature is that the leaders in these three EFL cases, acting of their own accord, were able to take advantage of the *situation* and were able to distribute their leadership across their organization.

10 Conclusion

The English as a Foreign Language (EFL) program often constitutes the largest department or program in a higher education institution. The EFL program may also have the largest number of students and instructors in the institution. As such, monitoring quality across the unit is critical. Until recently, in many EFL programs, quality was measured by the percentage of students who were able to fulfil the stated requirements and advance to their academic programs. However, pressure is increasing from external forces (e.g. taxpayers, parents, quality assurance/accrediting bodies, rankings agencies, competitors). As this situation progresses, there is ever-more need to demonstrate that students are acquiring specified knowledge, skills, and behaviors. And, if they are not, then stakeholders want to know how the program is going to respond to the situation. Outcomes assessment is broadly viewed as an ideal way to monitor achievement as well as drive change toward more effective teaching and learning. This is especially true if educational leaders heed the advice carefully explicated in some of the more well-respected texts in the field (e.g. Banta et al., 2009; Baker et al., 2012; Maki, 2004; Suskie, 2010; Walvoord, 2010). Certainly, designing, developing, and implementing a sustained outcomes assessment effort is rarely a smooth ride, and this can be especially so for EFL programs embarking on this process as they can be large, unfamiliar with such processes, and lacking institutional resources. Implementing such an initiative involves culture change, and it is well-documented that organizational culture is one of the primary barriers preventing successful implementation (Blaich & Wise, 2011; Hersh & Keeling, 2013; Tagg, 2007). Thus, Distributed Leadership may provide a useful framework for EFL leaders who are embarking upon this intimidating task. Distributed Leadership shifts the focus of leadership from the individual, positional leaders to the actual activity of leadership, providing some explanation as to how leadership can be stretched across an organization through both individuals (i.e. leaders and followers), as well as the so-called situation. The pitfalls associated with implementation of outcomes assessment initiatives—lack of participation, lack of consensus, lack of deep learning—may be

mitigated through Distributed Leadership as it generates broader, more committed involvement in the outcomes assessment cycle. Greater engagement results in greater dedication to the vision of the program and commitment to its success. In the end, the unrealistic image of a single, positional leader influencing deep learning in the organization transforms into a more sensible notion of vision and commitment distributed across the instructors and staff, thus enhancing the possibility that the drive for improved learning becomes woven into the culture of the program.

References

- Angelo, T. A. (1999). Doing assessment as if learning matters most. *AAHE Bulletin*, 51(9), 3–6.
- Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.
- Baker, G. R., Jankowski, N., Provezis, S., & Kinzie, J. (2012). *Using assessment results: Promising practices of institutions that do it well*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Banta, T. W. (1993). *Making a difference: Outcomes of a decade of assessment in higher education. The Jossey-Bass higher and adult education Series*. San Francisco: Jossey-Bass.
- Banta, T. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine for Higher Learning*, 43(1), 22–27. Retrieved from http://college.wfu.edu/assessment/wp-content/uploads/Closing_The_Assessment_Loop-Banta_Blaich-2011.pdf.
- Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. San Francisco, CA: John Wiley & Sons.
- Blaich, C. & Wise, K. (2011). *From gathering to using assessment results: Lessons from the Wabash National Study National Institute for Learning Outcomes Assessment*. (NILOA Occasional Paper No.8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Bolden, R., Petrov, G., & Gosling, J. (2009). Distributed leadership in higher education rhetoric and reality. *Educational Management Administration & Leadership*, 37(2), 257–277.
- Bresciani, M. J. (2009). *How to build and sustain a culture of assessment in your college and department*. Retrieved from <https://manoa.hawaii.edu/assessment/workshops/pdf/Bresciani%20MAC%202009-09-09.pdf>.
- Bresciani, M. J. (2012). Recommendations for implementing an effective, efficient, and enduring outcomes-based assessment program. *Community College Journal of Research and Practice*, 36, 411–421.
- Cross, K. P. (1998). Why learning communities? Why now. *About Campus*, 3(3), 4–11.
- Ewell, P. T. (1988). Outcomes, assessment, and academic improvement: In search of usable knowledge. *Higher Education: Handbook of Theory and Research*, 4, 53–108.
- Hersh, R. H. & Keeling, R. P. (2013). *Changing institutional culture to promote assessment of higher learning*. (NILOA Occasional Paper No.17). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Hutchings, P. (2010). *Opening doors to faculty involvement in assessment*. (NILOA Occasional Paper No.4). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Jankowski, N., & Provezis, S. (2011). *Making student learning evidence transparent: The state of the art*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

- Jones, S., Lefoe, G., Harvey, M., & Ryland, K. (2012). Distributed leadership: A collaborative framework for academics, executives and professionals in higher education. *Journal of Higher Education Policy and Management*, 34(1), 67–78.
- Kuh, G., & Ikenberry, S. (2009). *More than you think, less than we need: Learning outcomes assessment in American Higher Education*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- Lennon, M. C., Frank, B., Humphreys, J., Lenton, R., Madsen, K., Omri, A., & Turner, R. (2014). *Tuning: Identifying and measuring sector-based learning outcomes in postsecondary education*. Toronto: Higher Education Quality Council of Ontario.
- Maki, P. L. (2004). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus.
- Miller, M. A. (2012). *From denial to acceptance: The stages of assessment*. (NILOA Occasional Paper No.13). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
- New Leadership Alliance for Student Learning and Accountability. (2012). *Committing to quality: Guidelines for assessment and accountability in higher education*. Washington, DC: author. Retrieved from <http://www.chea.org/pdf/Committing%20to%20Quality.pdf>.
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education. Higher and adult education series*. San Francisco: Jossey-Bass.
- Pont, B., Nusche, D., & Hopkins, D. (Eds.). (2008). *Improving school leadership, Volume 2: Case studies on system leadership*. Paris: Organisation for Economic Co-operation and Development.
- Spillane, J. P., & Sherer, J. Z. (2004). *A distributed perspective on school leadership: Leadership practice as stretched over people and place*. San Diego: AERA.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2001). Investigating school leadership practice: A distributed perspective. *Research News and Comment*, April 2001, 23–28.
- Spillane, J. P., Halverson, R., & Diamond, J. B. (2004). Distributed leadership: Toward a theory of school leadership practice. *Journal of Curriculum Studies*, 36(1), 3–34.
- Suskie, L. (2010). *Assessing student learning: A common sense guide* (2nd ed.). San Francisco, CA: John Wiley & Sons.
- Tagg, J. (2007). Double loop learning in higher education. *Change: The Magazine for Higher Learning*, 39, 36–41.
- Van Ameijde, J. D., Nelson, P. C., Billsberry, J., & Van Meurs, N. (2009). Improving leadership in Higher Education institutions: A distributed perspective. *Higher Education*, 58(6), 763–779.
- Walvoord, B. E. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco, CA: John Wiley & Sons.

Outcomes-Based Assessment in an English Language Program

Khadernawaz Khan and Umamaheswara Rao Bontha

Abstract Language learning and teaching programs are generally concerned with course design, preparation of syllabi and study plans, and assessment methods. Once prepared, these are meticulously implemented in the language classroom. The teaching and the learning process generally receives primary importance and assessment is often neglected. Assessments are done mechanically to grade students' performance. It is often found that some students who get good grades fail to achieve the intended learning outcomes of the course. Therefore, there is a need to design assessments based on the teaching/learning outcomes. The present chapter is based on action research. It explores both the need for assessment to adhere closely to the teaching/learning outcomes of a course, along with the implementation of alternative assessments such as writing portfolios, continuous assessments and self-assessments to bridge the gap that is left by traditional paper based tests and exams. Perceptions of teachers and learners regarding the achievement of the learning outcomes, conventional assessments and alternative assessments are recorded.

Keywords Alternative assessments • Learning outcomes • Outcomes-based assessment • Outcomes-based education • Perceptions

K. Khan (✉) · U.R. Bontha (✉)
Dhofar University, Salalah, Oman
e-mail: Khader_khan@hotmail.com

U.R. Bontha
e-mail: maheshlfdc@yahoo.co.in

1 Introduction

The English language has now pervaded across the continents as a language instrumental in connecting cultures and improving trade relations across the globe. It is now feverishly being taught and learned beyond its native borders, and in fact, voluminous research is being undertaken in ELT. The majority of this teaching and research is being done in what Kachru (1997) calls, the outer and expanding circles rather than in the center countries in which it was born.

In the present scenario, teaching and learning English has become one of the aims of the governments in the Gulf, as elsewhere, whose education policies reflect their firm belief in the economic development of their nations through the English language. Consequently, large sums of money are poured into training their human resources in acquiring and using the English language. The Sultanate of Oman has realized the fact that the English language can help in its economic growth and hence has placed more emphasis on arming its youth with the knowledge of this language. The government has helped private individuals to establish colleges and universities across the Sultanate. To maintain the academic standards on par with international academic institutions, the Ministry of Higher Education has established a watch dog called the Oman Academic Accreditation Authority (OAAA). Each Higher Educational Institution (HEI) is mandated to provide a General Foundation Program (GFP) to ensure that the students entering higher education have acquired the required skills in English, Mathematics and IT. The OAAA has prescribed outcome standards called the ‘Oman Academic Standards (OAS) for General Foundation Programs’ (GFPs) which every GFP in an HEI is expected to achieve before being accredited. The GFP standards clearly reflect the learning outcomes. “The standards also recognize that the higher education providers (HEIs) have the primary responsibility for providing high quality teaching and assessment of students.”

2 Learning Outcome Standards (LOSs) for English Language

The Learning Outcome Standards (LOSs) of the ‘Oman Academic Standards for General Foundation Programs’ (2008, p. 10) as regards English language include the four language skills viz., listening, speaking, reading, and writing and study skills. These standards expect students to:

- (a) Actively participate in a discussion on a topic relevant to their studies by asking questions, agreeing/disagreeing, asking for clarification, sharing information, expressing and asking for opinions.

- (b) Paraphrase information (orally or in writing) from a written or spoken text or from graphically presented data.
- (c) Prepare and deliver a talk of at least 5 min. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions.
- (d) Write texts of a minimum of 250 words, showing control of layout, organization, punctuation, spelling, sentence structure, grammar and vocabulary.
- (e) Produce a written report of a minimum of 500 words showing evidence of research, note taking, review and revision of work, paraphrasing, summarizing, use of quotations and use of references.
- (f) Take notes and respond to questions about the topic, main ideas, details and opinions or arguments from an extended listening text (e.g. lecture, news broadcast).
- (g) Follow spoken instructions in order to carry out a task with a number of stages.
- (h) Listen to a conversation between two or more speakers and be able to answer questions in relation to context, relationship between speakers, register (e.g. formal or informal).
- (i) Read a one to two-page text and identify the main idea(s) and extract specific information in a given period of time.
- (j) Read an extensive text broadly relevant to the student's area of study (minimum three pages) and respond to questions that require analytical skills, e.g. prediction, deduction, inference.

3 Learning Outcome Standards (LOSs) for the Level 3 English Language Unit (ELU) at the Foundation Program (FP) in Dhofar University (DU)

The LOSs stipulated by the OAAA are very generic and broad in nature and thus it is not convenient to have one level or one program to realize these outcomes. As a result, each HEI has its own FP structure depending on the needs of the learners. The FP at Dhofar University has prepared the learning outcomes for three levels viz., level 1, level 2 and level 3. These learning outcomes are specific in nature, and they are graded and distributed across the three levels. The learning outcomes relevant to this study are the ones for level three (see Appendix 1); these are directly related to the LOSs of the 'Oman Academic Standards for General Foundation Programs' by the Ministry of Higher Education (MOHE), Sultanate of Oman.

4 Assessment

Assessment is an important aspect of any teaching/learning program. Assessment is a process of collecting, synthesizing, and interpreting information on students' performance to measure their progress (Airasian, 2005; Harris & McCann, 1994). The main purpose of assessment is to improve student's learning. Oller (1987, p. 45) states that "within such a practical and comprehensive philosophy of language instruction and testing, every test becomes a natural rung in the ladder toward the instructional goal... and every instructional activity in which students participate becomes a language-testing activity". In other words, assessment assumes a significant role in the process of teaching and learning.

5 Outcomes-Based Assessment

An outcomes-based curriculum needs to have an associated assessment that is also based on the learning outcomes. This makes assessment meaningful and checks how far learners have realized the set learning outcomes. As such, assessment should be authentic and relevant to real life situations to make it effective. Malan (2000, p. 26) describes authentic assessment as competencies in contexts that closely resemble situations in which those competencies are required. Assessment then gets integrated with learning and enhances opportunities for learners to improve their skills. Assessment in Outcomes-Based Education (OBE) focuses on the achievement of clearly defined outcomes, making it possible to credit learners' achievements at every level (Government Gazette, Pretoria, 2003). Van der Horst and McDonalds (1997) state that assessment procedures should give a clear indication of what learners are learning, and thus learning/teaching must be aligned with assessment. Killen (2003) argues that validity is an important characteristic of good assessment. In other words, in Outcomes-Based Education it is important to train the learners to apply the knowledge or skills that they learn/acquire in the classroom in real life situations. Assessment in OBE should not be a mere reproduction of the classroom activities. In order to validate Outcomes-Based Assessment (OBA), relevant types of assessments such as, self-assessments and continuous assessments etc., are to be implemented along with conventional tests and exams. Orsmond and Gildenhuis (2005, p. 122) mention the following different types of assessment: baseline assessment, formative assessment, diagnostic assessment and summative assessment. Therefore, it is important to use various types of assessment to measure the achievement of learning outcomes of the learners. Pallapu (2004) opines that assessment in OBE is more than memorization; the student should be able to demonstrate the skills acquired. That is, assessment in

OBE should be aimed at measuring the learners' achievement of the learning outcomes. Pallapu (2004) further states that skills should also be assessed. Similarly, Olivier (2002) believes that different strategies should be followed when assessing learners so that skills, knowledge, attitudes, values and the learning process can be assessed. Taking all this into consideration, in order to make OBE effective we must align Outcomes-Based Assessment with Outcomes-Based Education.

6 Research Questions

The present research deals with the following research questions:

1. Does assessment need to adhere closely to the teaching/learning outcomes of the courses?
2. How does the implementation of the alternative assessments such as writing portfolios, continuous assessments and self-assessments bridge the gap that is left by the traditional paper based tests and exams?
3. What are the perceptions of the teachers and the learners toward conventional as well as alternative assessments?

7 Research Tool and Method

In order to get a comprehensive picture of the perceptions of the students and the teachers, questionnaires were designed on a 5 point Likert Scale in which 1 corresponds to 'strongly agree' and 5 corresponds to 'strongly disagree' with scale 3 corresponding to 'neutral' (see Appendices 1 and 2). The questionnaires were developed based on the learning outcomes of level 3. One hundred level three students and twenty level three teachers were selected for the present study as the researchers thought that the learning outcomes at level 3 correspond to the Learning Outcome Standards (LOSs) as specified in the General Foundation Program document. The teachers belong to various nationalities including both non-native and native English speaking countries. The questionnaires were administered randomly. For the final analysis, the five response categories of the Likert scale were merged into three categories of 'Agree', 'Disagree' and 'Neutral.' The findings were tabulated accordingly.

8 Findings

The items in the questionnaires require teachers and students to record their perceptions on how far the learners have achieved the learning outcomes by the end of level 3, which is the last level of the English Language Unit (ELU) in the Foundation Program. The responses of both students and teachers indicate that the learners were able to realize most of the intended learning outcomes. To record the perceptions of each category of respondents clearly, the findings of each category are presented below separately. It can be observed that the perceptions of both teachers and students were, by and large, the same as they work towards the same learning outcomes. It is interesting to note that both the teachers and the students have positive perceptions towards Outcomes-Based Assessment.

9 Discussion

The first research question is, “Does assessment need to adhere closely to the teaching/learning outcomes of the courses?” Language courses at ELU in the FP are outcome-based. The outcomes of the courses follow the LOSs of the GFP document of the Ministry of Education, Sultanate of Oman. GFPs in all the Higher Education Institutions should follow the guidelines of the MOHE. It is mandatory for the HEIs to prepare their courses based on the broad LOSs. Dhofar University, like any other HEI, has split the broad LOSs into very specific learning outcomes and distributed them to levels 1, 2 and 3. The learning outcomes at level 3 are directly aligned with the LOSs. Therefore, the realization of learning outcomes at level 3 implies the realization of the LOSs.

The second research question is, “How does the implementation of the alternative assessments such as writing portfolios, continuous assessments and self-assessments bridge the gap that is left by the traditional paper based tests and exams?” The English Language Unit at the FP in Dhofar University conducts traditional/conventional paper based tests and exams. There is a mid-semester assessment and a final assessment. These tests and exams as well as alternative assessments are designed based on the learning outcomes. The writing portfolio assesses the following learning outcomes that expect them to:

W1: ‘Use newly learnt vocabulary in new contexts’: The learners have to maintain a vocabulary log where they record the words they studied in their reading lessons, identify the parts of speech, guess/find their meanings and write sentences on their own.

W2: ‘Recognize the structure of short texts: introduction, body and conclusion by analyzing model texts’: The writing portfolio consists of model essays for them to understand the structure and the layout corresponding to the learning outcome.

W3: ‘Write clear, well-structured paragraphs to express opinions using suitable vocabulary’: The writing portfolio contains writing topics that students are expected to develop into essays to realize the outcome.

W4: ‘Use the writing process (i.e. brainstorming, outline, etc.) to write clear, well-structured and well-organized Compare and Contrast; Agree/Disagree; and Cause and Effect essays of about 250 words’: The writing portfolio expects the employment of the mechanics of writing by the learners in producing the essays falling under these genres.

W5: ‘write descriptions of graphs, tables and diagrams’: The portfolios provide a variety of graphs, charts and diagrams, so that the students understand and interpret the information presented there effectively.

Self assessment sheets are provided to the learners in the writing portfolio to evaluate their learning outcomes at the beginning, in the middle and at the end of the semester. Continuous assessments are conducted to test certain learning outcomes that cannot be tested through traditional tests and exams. The learning outcomes such as; “S1—Actively participate in a discussion”, and “S2—Respond to oral and written communication using suitable vocabulary” in speaking are tested through continuous assessments. Study skills, such as “SS1—Managing time and accepting responsibility” are also tested through continuous assessment.

The third research question is, “What are the perceptions of the teachers and the learners toward conventional as well as alternative assessments?”

This part of the paper deals with the most significant aspects of the research. For convenience, it is sub-divided as follows:

10 Outcomes-Based Education and Assessment

While 90 % of the learners indicated their awareness of outcomes-based education, only 50 % of their teachers confirmed their students’ view. The statement, “English language courses should be outcomes-based” (item 2) received 90 % agreement from teachers, while 80 % of the learners also agreed with the statement. William Spady (1994) believes that outcomes based education means clearly focusing and organizing everything in the education system around what is essential for all students to be able to do successfully at the end of their learning experiences. This means starting with a clear picture of what is important for students to be able to do, then organizing the curriculum, instruction and assessment to make sure that learning ultimately happens.

Item 3, “English language course materials should be tailored keeping the learning outcomes in mind,” was agreed by 82 % of the learners; similarly, 90 % of the teachers also agreed. This shows the importance of selecting the materials that are outcome-based as, according to Norman (2006), “Education that is outcome-based is learner centered, results oriented and founded on the belief that all students can learn” (p. 46).

Regarding assessments, for item 4, 82 % of the learners felt that English language assessments should be outcomes-based. In the same way, 90 % of the teachers too felt that English language assessments should be outcomes-based. Nearly, 70 % of both the learners and the teachers expressed their agreement with the statement, “Outcomes-based assessments alone help us understand whether or not we have achieved those outcomes” (item 5).

The statement, ‘Outcomes-based assessments motivate the learners to acquire the language skills better’ (item 6) was agreed by 84 % of the learners. On the contrary, only 60 % of the teachers agreed. However, Charles (1999, p. 244) feels that this approach has a strong influence on motivation of students as it encourages students to acquire knowledge and skills that are considered important in their lives.

Interestingly, about 72 % of both the teachers and the learners felt that outcomes-based assessments can evaluate the real learning outcomes achieved by the learners (item 7). This is covered by Malan (2000):

The ultimate purpose of assessment is to validate learning outcomes—be it for diagnostic, formative or summative purposes. The role of assessment in OBE is part and parcel of the aims of assessment in all its root models. OBE, however, highlights continuous and criterion-referenced assessment (p. 26).

11 Learning Outcomes—Reading

Skimming, scanning and deducing word meanings from the context are some of the important sub-skills of reading. Learners should master these sub-skills to become effective readers. 72 % of the learners agreed with the statement “I can skim a short passage (moderate difficulty) to identify the main ideas and supporting details,” (item 8). Similarly, 80 % of the teachers accepted that their learners are able to skim short passages to identify the main ideas and supporting details. Rayner and Pollatsek (1989, p. 447) gave considerable importance to skimming. According to them, students are often overloaded by too many books presenting too much information. Skimming helps them by saving their time, but people who are “unable to skim material would find [that] they spend their entire day reading”. Skimming also helps in understanding the main ideas in a text. Grellet (1996, p. 19) thinks that it is “a more thorough activity” because it “requires an overall view of the text and implies a definite reading competence”.

Teachers unanimously agreed that their learners can scan a short text to identify specific information (item 9). On the contrary, only 70 % of the learners agreed that they can scan a short text to identify specific information. Scanning is an important skill that helps the readers to search for the right information. Williams (1996, p. 107) states that scanning is “reading for particular points of information”.

More than half (58 %) of the learners felt they can deduce word meanings from context (item 10). This is compared to 50 % of the teachers who felt that the learners can deduce word meanings from context. Although deducing word meanings from context is one of the sub-skills supported by researchers who advocate incidental vocabulary learning, most of those surveyed are of the opinion that these learners are not able to guess word meanings from context. This is in agreement with Folse (2004), who identified learning new words from context as a “vocabulary myth,” arguing that for L2 readers in particular, the linguistic context itself may be too unfamiliar to be helpful.

“I can interpret text by answering relevant questions (comprehension; evaluation & inferences)”, (item 11) received 84 % agreement from the learners. Similarly, 70 % of the teachers agreed that their learners can interpret a text by answering relevant questions (comprehension; evaluation & inferences). Widdowson (1978, p. 94) has suggested various types of questions including open questions, right or wrong, multiple-choice options etc., for understanding meaning. However, Grellet (1996, p. 13 and pp. 21–24) has pointed out that these ‘question-types’ can be advantageous to the students only if they involve the students actively, that is, if the activities require the students “to think and reason in order to give answers or make a choice”.

12 Learning Outcomes—Writing

Vocabulary usage in context is an important aspect of writing. 88 % of the learners and 80 % of the teachers felt that the learners can use newly learnt vocabulary in new contexts (item 12).

A considerable number (80 %) of the learners expressed that they can recognize the structure of short texts (introduction, body and conclusion) by analyzing model texts (item 13). Likewise, 90 % of the teachers felt the same way. Acknowledging the importance of this ability, Tangkiengsirisin (2006, p. 4) feels that “Certain structural entities, e.g. Introduction-Body-Conclusion, are the major components of the texts, and students are taught to write with particular organizational patterns or modes (normally narration, description, and exposition)”.

To the statement, “I can write clear, well-structured paragraphs to express opinions using suitable vocabulary” (item 14), 82 % of the learners responded positively. In contrast, only 30 % of the teachers expressed that their learners can write well-structured paragraphs, while 40 % remained neutral and 30 % of them disagreed. The learners might be aiming at expressing ideas in a piece of composition, but the teachers might be expecting their learners to produce an error-free and fluent essay with good ideas. The teachers need to make the learners aware that a written composition should be fluent and also accurate with good ideas.

The statement, “I can use the writing process (i.e. brainstorming, outline, etc.) to write clear, well-structured and well-organized Compare and Contrast; Agree/Disagree; and Cause & Effect essays (250 words)” (item 15), on one hand received 72 % positive responses from the learners. On the other hand, 80 % of the teachers felt that their learners can use the writing process to write essays. A process approach to writing helps the learners to start writing essays. In this regard, Raimes (1985, pp. 230–231) mentions that the process approach to writing stresses: “generating ideas, writing drafts, producing feedback, and revising” in order to help make learners, “behaviour, and ultimately their products, more like those of skilled writers”. The writing process ultimately assists the learners to produce well-organized essays.

Somewhat surprisingly, only 64 % of the learners responded positively to the statement, “I can write descriptions of graphs, tables and diagrams” (item 16). However, 90 % of the teachers felt that their learners were able to describe graphs, tables and diagrams. This suggests that there is a need for teachers to provide timely feedback to the learners.

13 Learning Outcomes—Listening

Table 1 illustrates the responses of learners and teachers toward learning outcomes for listening skill from items 17–21 on the questionnaire (refer to Appendices 1 and 2 for a full list of items and response per cents).

Listening assumes a degree of prominence in English language learning as language learning begins through listening. As such, it is an important component in many international tests such as IELTS and TOEFL. Not only is listening comprehension important at the beginning stages of SLA, it appears to be crucially important for advanced level learners (e.g., those with TOEFL scores > 500) as well (Powers, 1985).

Table 1 Teacher and learner responses toward learning outcomes

Item no.	Teachers' responses			Learners' responses		
	Agree	Neutral	Disagree	Agree	Neutral	Disagree
17.	80	20	–	80	10	10
18.	80	20	–	66	20	14
19.	90	10	–	76	14	10
20.	80	10	10	78	10	12
21.	40	40	20	64	20	16

Table 2 Teacher and learner responses to items 22 and 24

Item no.	Teachers' responses			Learners' responses		
	Agree	Neutral	Disagree	Agree	Neutral	Disagree
22.	80	10	10	74	22	4
23.	90	–	10	72	18	10
24.	70	30	–	58	38	4

14 Learning Outcomes—Speaking

Table 2 presents the responses of teachers and learners to items 22–24 on the learning outcomes of the speaking skills (refer to Appendices 1 and 2).

The statement, “I can deliver a talk of at least 5 min on an assigned topic” (item 24), received only 58 % agreement from the learners. In contrast, 70 % of the teachers felt that their learners can deliver a talk of at least 5 min on an assigned topic. This is an area that needs to be attended to as the learners seem to be not confident enough regarding their abilities to give presentations.

15 Learning Outcomes—Study Skills

Study skills play a significant role in the academic success of the learners; effective study skills no doubt improve performance. This argument was supported by the learners as well as teachers as 70 % of the learners and 60 % of the teachers felt that the learners can manage time and accept responsibility (item 25). With regard to taking notes, 66 % of the learners and 60 % of the teachers agreed that the learners can take notes (item, 26). Another important learning outcome under study skills is giving presentations. 78 % of the learners and 90 % of the teachers expressed the belief that learners can give presentations.

16 Learning Outcomes—Traditional and Alternative Assessments

The teachers unanimously agreed that there are tests and exams during and at the end of the semester in the Foundation Program (item 28). 86 % of the students too agreed with it. However, only 66 % of the learners and 60 % of the teachers agreed that the learners score good grades in tests and exams (item 29). This indicates that the performance of the learners is still not at a high enough level. In an

outcomes-based education the focus is less on whether the learners have passed or failed the exam. The focus is primarily on what outcomes the learners have achieved and what outcomes they need more support with (Norman, 2006). Lorenzen (1999, p. 2) argues that, by its very nature, outcomes-based education eliminates the need for traditional assessment tools such as tests or grades. This shows that in an outcomes-based education the assessment should be focused primarily on the outcomes. The Foundation Program at Dhofar University uses both traditional paper based tests and exams and alternative assessments to make assessment outcomes-based.

The statement, "There are alternative assessments such as portfolios, self-assessments, continuous assessments etc." (item 30), was unanimously agreed with by the teachers, and 88 % of the learners agreed with it. The teachers unanimously agreed that the learners get good grades on alternative assessment (item 31). On the contrary only 66 % of the learners agreed that they scored good grades in alternative assessments. Chapelle and Douglas (1993) maintain that the discrepancies between tests and classroom activities can be resolved with the help of continuous assessment. Moreover, self-assessments have a significant role to play in assessing learners' knowledge and skills. Blanche and Merino (1989) found that self-appraisal exercises are likely to increase the motivation of the language learner. They also showed that people can assess themselves quite accurately, given the proper conditions, with the most accurate self-test items described being concrete linguistic situations that the learner can size up in behavioral terms (1989, p. 324). This proves that self-assessments help learners to motivate themselves and try to understand their strengths and weaknesses, which paves the way for them to achieve the learning outcomes of their courses.

17 Conclusions

From the preceding discussion, certain conclusions can be drawn. Perhaps most importantly is that small in language courses need to adhere closely to course teaching/learning outcomes. It was found that the English Language Unit in the Foundation Program at Dhofar University implements alternative assessments such as student self-assessments, continuous assessments and writing portfolio assessments along with the traditional/conventional paper based tests and exams.

With regard to outcomes-based education and assessment, a considerable number of learners and teachers have agreed that English language courses and assessment need to be outcome-based.

Regarding reading skills, both the learners and the teachers felt that the learners are able to apply skimming and scanning techniques while reading a text. However,

a much lower percentage of learners and teachers agreed that the learners are able to deduce word meanings from the context.

As regards writing, both the learners and the teachers agreed that the learners are able to use newly learnt vocabulary in different contexts and that they can recognize the structure of short texts. By contrast, only 30 % of the teachers opined that their learners can write clear, well-structured paragraphs to express opinions using suitable vocabulary. In addition, the majority of learners and teachers felt that the learners can use the writing process to write essays. 90 % of the teachers felt that their learners were able to describe graphs, tables and diagrams, but only 64 % of the learners felt that they have the ability to describe graphs, tables and diagrams.

The majority of teachers do not believe that their learners can distinguish fact from opinion. 58 % of the learners felt that they can deliver a talk on any assigned topic for 5 min, but 70 % of the teachers felt they could. Here again, the disparity between the learners and the teachers shows that there is a need for providing the learners with feedback on their short talks. However, the teachers did not agree unanimously. This shows that, according to teachers, there are still 30 % of the learners who cannot deliver a talk on an assigned topic for 5 min. Time management and taking notes are the other two areas where the teachers and the learners felt that the learners are still not up to the expected level.

Almost all the teachers agreed that the learners score better on alternative assessments compared to tests and grades. Learners also confirmed this perception.

The positive attitude of the learners and teachers toward Outcomes-Based Assessment indicates that the English language courses should be tested based on the course learning outcomes. This would help educators to focus on the real needs of the learners and further help the learners to apply their learning in real world situations.

18 Pedagogical Implications

Teacher workshops and seminars have to be conducted on Outcomes-Based Education and Outcomes-Based Assessment in order to help the teachers understand the nuances of the concepts and help to make the assessment more valid.

The low percentage of teachers agreeing that their learners are able to write essays shows that more assignments or changes in classroom practices are needed to enable the learners to improve this skill.

Alternative assessments such as writing portfolios, self-assessments and continuous assessment should be structured appropriately and thought should be given to their implementation. Appropriate rubrics have to be developed to validate them. Learners should be given projects/assignments that require them to use the English language in a real world context. This helps them understand how they are able to achieve and apply the learning outcomes of their courses. Finally, teachers should provide timely feedback to the learners to help them understand their weakness and strengths in all the language skills.

Appendix 1

Students' Questionnaire

Tick (✓) the correct options.

SA = Strongly Agree A = Agree N = Neutral D = Disagree SD = Strongly Disagree

Outcome-Based Education and Assessment:

No	Statement	A	N	D
1.	I am aware of learning outcomes.	90	6	4
2.	English language courses should be outcome -based.	80	12	10
3.	English language course materials should be tailored keeping the learning outcomes in mind.	82	12	6
4.	English language assessments should be outcome-based.	86	8	6
5.	Outcome-based assessments alone help us understand whether or not we have achieved those outcomes.	68	28	4
6.	Outcome-based assessments motivate us to acquire the language skills better.	84	8	8
7.	Outcome-based assessments can evaluate the real learning outcomes achieved by us.	72	18	10

Reading Skills:

No	Statement	A	N	D
8.	I can skim a short passage (moderate difficulty) to identify the main ideas and supporting details.	72	22	6
9.	I can scan a short text to identify specific information.	70	20	10
10.	I can deduce word meanings from context.	58	28	14
11.	I can interpret text by answering relevant questions (comprehension; evaluation & inferences).	84	8	8

Writing Skills:

No	Statement	A	N	D
12.	I can use newly learnt vocabulary in new contexts.	84	8	8
13.	I can recognize the structure of short texts: introduction, body and conclusion by analyzing model texts.	80	8	12
14.	I can write clear, well-structured paragraphs to express opinions using suitable vocabulary.	82	14	4
15.	I can use the writing process (i.e. brainstorming, outline, etc.) to write clear, well-structured and well-organized Compare and Contrast; Agree/ Disagree; and Cause & Effect essays (250 words)	72	14	14
16.	I can write descriptions of graphs, tables and diagrams.	64	20	16

Listening Skills:

No	Statement	A	N	D
17.	I can identify specific information from monologues, dialogues or lectures.	80	10	10
18.	I can identify context of monologues, dialogues or lectures.	66	20	14
19.	I can respond to spoken instructions.	76	14	10
20.	I can transfer spoken information to a table.	78	10	12
21.	I can distinguish fact and opinion from a spoken text.	64	20	16

Speaking Skills:

No	Statement	A	N	D
22.	I can actively participate in a discussion.	74	22	4
23.	I can respond to oral and written communication using suitable vocabulary.	72	18	10
24.	I can deliver a talk of at least 5 minutes on an assigned topic.	58	38	4

Study Skills:

No	Statement	A	N	D
25.	I can manage time and accept responsibility.	70	20	10
26.	I can take notes.	66	22	12
27.	I can give presentations	78	14	8

Traditional Tests and Alternative Assessments

No	Statement	A	N	D
28.	There are tests and exams during and at the end of the semester.	86	8	6
29.	I get good grades in tests and exams.	66	22	12
30.	There are alternative assessments such as portfolios, self-assessments, continuous assessments etc.	88	6	6
31.	I get good grades in alternative assessments.	66	28	6

Appendix 2: Teachers’ Questionnaire

Tick (✓) the correct options.

SA = Strongly Agree A = Agree N = Neutral D = Disagree SD = Strongly Disagree

Outcome-Based Education and Assessment:

No	Statement	A	N	D
1.	Students are aware of learning outcomes.	50	30	20
2.	English language courses should be outcome-based.	90	-	10
3.	English language course materials should be tailored keeping the learning outcomes in mind.	90	10	-
4.	English language assessments should be outcome-based.	90	-	10
5.	Outcome-based assessments alone help us understand whether or not the learners have achieved those outcomes.	70	10	20
6.	Outcome-based assessments motivate the learners to acquire the language skills better.	60	20	20
7.	Outcome-based assessments can evaluate the real learning outcomes achieved by the learners.	70	20	10

Reading Skills:

No	Statement	A	N	D
8.	Learners can skim a short passage (moderate difficulty) to identify the main ideas and supporting details.	80	10	10
9.	Learners can scan a short text to identify specific information.	100	-	-
10.	Learners can deduce word meanings from context.	50	30	20
11.	Learners can interpret text by answering relevant questions (comprehension; evaluation & inferences).	70	20	10

Writing Skills:

No	Statement	A	N	D
12.	Learners can use newly learnt vocabulary in new contexts.	80	10	10
13.	Learners can recognize the structure of short texts: introduction, body and conclusion by analyzing model texts.	90	-	10
14.	Learners can write clear, well-structured paragraphs to express opinions using suitable vocabulary.	30	40	30
15.	Learners can use the writing process (i.e. brainstorming, outline, etc.) to write clear, well-structured and well-organized Compare and Contrast; Agree/ Disagree; and Cause & Effect essays (250 words)	80	20	-
16.	Learners can write descriptions of graphs, tables and diagrams.	90	-	10

Listening Skills:

No	Statement	A	N	D
17.	Learners can identify specific information from monologues, dialogues or lectures.	80	20	-
18.	Learners can identify context of monologues, dialogues or lectures.	80	20	-
19.	Learners can respond to spoken instructions.	90	10	-
20.	Learners can transfer spoken information to a table.	80	10	10
21.	Learners can distinguish fact and opinion from a spoken text.	40	40	20

Speaking Skills:

No	Statement	A	N	D
22.	Learners can actively participate in a discussion.	80	10	10
23.	Learners can respond to oral and written communication using suitable vocabulary.	90	-	10
24.	Learners can deliver a talk of at least 5 minutes on an assigned topic.	70	30	-

Study Skills:

No	Statement	A	N	D
25.	Learners can manage time and accept responsibility.	60	40	-
26.	Learners can take notes.	60	20	20
27.	Learners can give presentations	90	10	-

Traditional Tests and Alternative Assessments

No	Statement	A	N	D
28.	There are tests and exams during and at the end of the semester.	100	-	-
29.	Our learners get good grades in tests and exams.	60	20	20
30.	There are alternative assessments such as portfolios, self-assessments, continuous assessments etc.	100	-	-
31.	Our learners get good grades in alternative assessments.	100	-	-

Appendix 3: Learning Outcomes of Level 3, ELU, FP, Dhofar University

Learning Outcomes – Reading Skills

- R1** Skim a short passage (moderate difficulty) to identify the main ideas and supporting details.
- R2** Scan a short text to identify specific information.
- R3** Deduce word meanings from context.
- R4** Interpret text by answering relevant questions (comprehension; evaluation & inferences).

Learning Outcomes – Writing Skills

- W1** Use newly learnt vocabulary in new contexts.
- W2** Recognize the structure of short texts: introduction, body and conclusion by analyzing model texts.
- W3** Write clear, well-structured paragraphs to express opinions using suitable vocabulary.
- W4** Use the writing process (i.e. brainstorming, outline, etc.) to write clear, well-structured and well-organized Compare and Contrast; Agree/ Disagree; and Cause & Effect essays (250 words)
- W5** Write descriptions of graphs, tables and diagrams

Learning Outcomes – Listening Skills

- L1** Identify specific information from monologues, dialogues or lectures
- L2** Identify context of monologues, dialogues or lectures
- L3** Respond to spoken instructions
- L4** Transfer spoken information to a table.
- L5** Distinguish fact and opinion from a spoken text

Learning Outcomes – Speaking Skills

- S1** Actively participate in a discussion.
- S2** Respond to oral and written communication using suitable vocabulary.
- S3** Deliver a talk of at least 5 minutes on an assigned topic.

Learning Outcomes – Study Skills

- SS1** Managing time and accepting responsibility
- SS2** Research skills
- SS3** Taking notes.
- SS4** Giving presentation

References

- Airasian, P. (2005). *Classroom assessment: Concepts and applications*. Boston: McGraw-Hill.
- Blanche, P., & Merino, B. (1989). Self-assessment of foreign language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313–340.
- Charles, C. M. (1999). *Building classroom discipline* (6th ed.). New York: Longman.
- Chapelle, C., & Douglas, D. (1993). *A new decade of language testing research*. Alexandria, VA: TESOL.
- Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor: University of Michigan Press.
- Government Gazette, Pretoria. (2003, November, 7). No. 25699.
- Grellet, F. (1996). *Developing reading skills: A practical guide to reading comprehension exercises*. Cambridge, UK: Cambridge University Press.
- Harris, M., & McCann, P. (1994). *Assessment*. Oxford: Heinemann.
- Kachru, B. (1997). World Englishes and English-using Communities. *Annual Review of Applied Linguistics*, 17.
- Killen, R. (2003). Validity in outcomes-based assessment. *Perspectives in Education: Assessment of Change in Education*, 21(1), 1–14.
- Lorenzen, M. (1999). *Using outcomes-based education in planning and teaching new information technologies*. Retrieved from <http://www.libraryinstructoin.com/obe/html>.
- Malan, S. P. (2000). The 'new paradigm' of outcomes based education in perspective. *Journal of Family Ecology and Consumer Sciences*, 28, 22–28.
- Ministry of Higher Education. (2008). *Oman academic standards for general foundation programs*. Muscat: Oman Accreditation Council
- Norman, P. (2006). Outcomes-based education: A PNG perspective. *Contemporary PNG Studies. DWU Research Journal*, 5, 45–47.
- Olivier, C. (2002). *Let's educate, train and learn outcome-based*. Pretoria: Benedic.
- Oller, J. W, Jr. (1987). Practical ideas for language teachers from a quarter century of language testing. *English Teaching Forum*, 25(4), 44–46 & 55.
- Orsmond, C., & Gildenhuis, C. (2005). *Further education and training, life orientation*. Johannesburg: University of Witwatersrand.

- Pallapu, K. S. (2004). *Automating outcome based assessment*. Retrieved from www.poly.asu.edu/ctas/dest/projects/pdf/SuseelPallapu.pdf.
- Powers, D. (1985). *A survey of academic demands related to listening skills*. Princeton, NJ: Educational Testing Service.
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, 19(2), 229–256.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Spady, W. (1994). *Outcome-based education: Critical issues and answers*. Arlington, VA: American Association of School Administrators.
- Tangkiengsirisin, S. (2006). Approaches to teaching second language writing. *Language Institute Journal*, 3, 1–26.
- Van Der Horst, H., & McDonalds, R. (1997). *Outcome-based education. A teachers' Manual*. Pretoria: Kagiso Publishers.
- Widdowson, H. (1978). *Teaching language as communication*. Oxford: Oxford University Press.
- Williams, E. (1996). *Reading in the language classroom*. Kuala Lumpur: Modern English Publications.

Dhofar University Documents

- Dhofar University. (n.d.). *Foundation program. A way to think about the OAC learning outcomes*. Salalah: author.
- Dhofar University. (n.d.). *Foundation program. Scope and sequence*. Salalah: author.

Assessment of EFL Through the Process of Problem-Based Learning

Melissa Caspary and Diane Boothe

Abstract The increasing numbers of English as a Foreign Language (EFL) and multiethnic students require engagement and exploration coupled with stimulating pedagogy across cultural and linguistic lines. Language learning and Problem-based Learning (PBL) are closely intertwined in their efforts to achieve success. This chapter focuses on meeting the diverse needs of learners and assessing their progress applying PBL methodology to EFL assessment. It offers educators creative and dynamic assessment strategies that include outcomes based on the PBL process and product. This pedagogical approach focuses on assessment tools and models that address unique teaching styles and key competencies while stimulating critical thinking and effective teamwork. Assessment tools are clearly identified allowing for individual differences as well as collaboration and group synergy in order to enhance EFL academic achievement.

Keywords Problem-based learning · Assessment · EFL

1 Introduction

This chapter discusses the process of applying Problem-based Learning (PBL) and how to relate that methodology to English as a Foreign Language (EFL) assessment. The theoretical foundations and underpinnings of PBL are explored in a focused and contrasting self-directed learning environment. Creative and dynamic PBL activities are linked to multiple strategies for continuous, summative, and formative EFL assessment. Rubrics that include positive outcomes based on a process and a product are included. PBL is incorporated into classroom assessment techniques (CATs),

M. Caspary (✉)

Georgia Gwinnett College, Lawrenceville, USA

e-mail: mcaspary@ggc.edu

D. Boothe (✉)

Boise State University, Boise, USA

e-mail: dianeboothe@boisestate.edu

portfolio assessments, peer assessments, self-assessments, reflective journals, writing samples, and authentic products that reflect individual and collaborative endeavors and meaningful learning. Opportunities for data gathering that support continuous improvement and connections to EFL digital and hybrid learning are also explored. An interdisciplinary approach that strengthens intercultural communication is an additional advantage of PBL and provides essential bridges to the 21st century classroom. Assessment tools and models are discussed in order to tap into unique teaching styles and opportunities for incorporating PBL into the EFL curriculum, i.e. flipping the classroom and developing a tool box of innovative activities to unlock the potential of PBL and nourish creativity. Furthermore, added value is derived from competency-based content area knowledge that is acquired during this process. A discussion of the benefits and challenges of EFL assessment, culminating activities, and value-added data provides an opportunity for students and educators to be bold and creative with a mutual sense of accomplishment for the PBL process. Cohesive plans for integrated professional development focusing on sustainability and building capacity address the growth metrics and structure that PBL assessment contributes to EFL mastery. A review of the existing literature as it directly relates to EFL and assessment is included, although it is certainly limited in its scope and breadth. This is further indication that this topic is relevant for EFL students, researchers and educators.

2 PBL Overview

Raising EFL student achievement through quality preparation and redesigning learning is at the forefront of academic pursuits. Implementing strategies that emphasize effective teaching and specific, practiced goals focusing on actively engaging students leads to academic excellence. Collaborative partnerships of EFL learners cultivating content expertise in multiple subjects while maximizing language skills will build strong connections and positively impact learning. Language acquisition is a demonstrated social activity that requires active participation (Scott, 1965). In the case of EFL education, Problem-based Learning (PBL) provides a competitive advantage in both EFL and content instruction, and focuses on experienced-based hands on learning. It has been shown to increase the integration of newly acquired concepts into existing knowledge structures and improve metacognition (Capon & Kuhn, 2004; Downing, Kwong, Chan, Lam, & Downing, 2009). When students are driving the problem posing and decision making, it has been found that these inquiry-based methods personalize the project, increase relevance, and create ownership (Johnson & Kean, 1992). A shared vision and evidenced-based learning goals coupled with collaborative action and an investment in sustainable assessment will help achieve positive results over time.

Because PBL is a total approach to education that includes a curriculum and a process, assessment of PBL activities and results is tantamount to the success of PBL best practices. Its roots stem from interdependence and constructionist theory,

cognitive development theory, and strategic learning theory and it is defined as “a curriculum development and instructional system that simultaneously develops both problem solving strategies and disciplinary knowledge bases and skills” (Finkle & Torp, 1995, p. 1). Psychologist Lev Vygotsky’s work has evolved into the Constructionist theory with strong underpinnings supporting the notion that learners gain knowledge as they make sense of experiences (Driscoll, 1994). Members of the medical field claim to have invented PBL, but educators in the content and language arena significantly developed it. The use of PBL is most appropriate for disciplines where there is an emphasis on applied knowledge (Mastascusa, Snyder, & Hoyt, 2011). Careful planning of innovative teaching and learning methodology is essential and that includes selected problems purposefully designed in terms of language learning as well as content knowledge tailored to students’ needs. The content is introduced in the context of real world problems in order for learner acquisition of critical knowledge. As students strive to solve problems and examine techniques and tools, they bring focused initiatives together in a manageable space for transformative learning.

Implementation of the PBL model requires structured planning and preparation. The selected problems that are purposely designed to actively engage students and address relevant problems including content and EFL instruction need to be determined prior to beginning the project. Because the roles of the instructor and students change, and the instructor becomes a facilitator encouraging students to assume an active role, all participants need to understand the lesson plan and expectations for the product, outcome and assessment. The specifics of the lesson design will include a hypothesis, authentic learning experiences for independent and active learning, group work, observations, and solutions to problems. The most common cycle to follow when introducing PBL is to begin to address a problem making observations and identifying learning issues posing a case for the hypothesis that they will propose. Questions are identified that will test the hypotheses and additional information is introduced as learning and project development will take place. As group dynamics evolve and participants interact, the use of English is encouraged and vocabulary is strengthened as new concepts are introduced and conversation takes place in the subject specific content and the collaborative strategizing develops to solve the problem. As new information is introduced, problems are analyzed and hypotheses modified in an effort to pose an innovative solution.

Assessment linked to this focused, contrastive and self-directed method is essential for measuring the end goal of what the curriculum is designed to achieve. Assessment methods tailored to learning strategies that are self-directed and encourage independent learning, collaboration and teamwork while strengthening EFL skills are essential for a world class education system that supports a shared vision and goals. Formative assessment connected to course goals and objectives with flexible timelines and quality feedback is essential and valuable for students and instructors. Success is derived by juxtaposing instructional content and access to multiple modes of learning and inquiry through a process that allows a diverse

array of learning activities strengthened by a high quality, well developed and appropriate assessment framework.

Quality assessment techniques are aimed at mobilizing EFL learners to practically apply their studies in real world collaborative settings. This enhances their experiential learning and preparation for the group dynamics of a changing workplace that is focused on rapidly evolving technological skills, a challenging economy, and expectations for collaboration coupled with innovation and teamwork. As students acquire language and subject matter knowledge, they become proficient in problem solving. All of these components improve and boost learning and preparation for careers and future opportunities.

3 Strategies for EFL Assessment

In order to address and benefit from the evaluative expertise of assessment, it is important to determine the end goal of what the curriculum expects to achieve. Creative and dynamic PBL activities directly linked to multiple strategies for assessment are parallel indicators that improve practice and are tantamount to success. These assessment strategies can be categorized into continuous, summative, and formative assessment enriching EFL learning and strengthening new horizons in innovation. A toolkit for best practices, action research and assessment will result in creativity and cooperation. This learning platform has the potential to positively impact pedagogical epistemology and curriculum scaffolding in building teacher and learner success. Each scaffolding activity becomes a steppingstone as valuable opportunities unfold and benchmarks in the learning process are achieved in both EFL and content. Instructional design and delivery of PBL and EFL best practices empowers students to enthusiastically embrace learning opportunities in both areas combining them into one learning experience and deriving success focusing on growth and achievement.

4 Rubrics

The use of rubrics to provide clear parameters regarding expectations for students coupled with PBL content is highly beneficial and successful in enhancing EFL pedagogy. Rubrics are powerful tools for teaching and assessment and allow students to clearly understand learning expectations. The PBL activities embody the authenticity of solving real world problems and working collaboratively with a dedicated team. As students navigate the purposefully designed problems and acquire the capacity to project solutions, they become immersed in EFL learning on one hand and gain content knowledge on the other. Rubrics will result in positive outcomes as students pursue both process and the product achieved through PBL and appropriate performance assessments. Rubrics can assist in monitoring student

progress and help to describe student achievement. As a tool for understanding knowledge transfer, rubrics are valuable means of determining ways to improve methodology for more effective teaching. Rubrics also provide a means of performing assessment efficiently. When working with an EFL student population, rubrics also clarify learning outcome goals when clear communication is critical for student success.

5 Classroom Assessment Techniques (CATs)

Classroom Assessment Techniques (CATs) are formative assessments of student learning based on simple, non-graded, anonymous, in-class activities designed to provide real-time feedback on the teaching-learning process as it is happening in the moment. CATs provide instant feedback to allow for individualization and modification that improves teacher-student communication and response. CATs are incredibly effective at taking the temperature of knowledge-transmission and classroom climate, and have been demonstrated to facilitate student-centered learning in the classroom (Soetaert, 1998). Numerous different CATs have been established and should be used for effective assessment in an EFL environment including the ‘Background Knowledge Probe’, ‘The Muddiest Point’, ‘The Minute Paper’, and ‘What’s the Principle’ assessment techniques.

The ‘Background Knowledge Probe’ is a short, simple questionnaire introduced before a topic to assess pre-conceptions in a content area. ‘The Muddiest Point’ is a request for a quick response to the most unclear concept introduced in a classroom session. ‘The Minute Paper’ is a reflective tool used to determine student conceptions about the most important and unaddressed issues in a course module or section. ‘What’s the Principle’ provides students with a few problems and asks them to state the principle that best applies to each problem (Angelo & Cross, 1993). Collectively, these CATs are used to highlight student misconceptions as well as those concepts that students find most important or unclear. These tools can be used to complement the PBL model and to define potential areas for improving EFL student learning.

6 Portfolio Assessment

Portfolio assessment affords students the opportunity to showcase accomplishments and to provide evidence of incremental knowledge and EFL skills that are strengthened and refined over time. As students assume responsibility for selecting examples of their finest work based on guidance from the professor and requirements for specific examples of various types of accomplishments, self-evaluation and selection lead to knowledge, enhanced perspectives and informal personal assessment. The opportunity to select and feature one’s own accomplishments

based on thoughtful reflection embodying authenticity will spark a more luminous snapshot of student learning and crystallize accomplishments. Evaluating and rediscovering their own accomplishments leads students to combine expertise in PBL and EFL with innovation and generates a unique approach to learning. As multimedia resources become the wave of the future, students are inspired to create a range of portfolios with numerous design features.

7 Peer Assessments

Peer assessments can be difficult to accurately attain as the surface veneer of friendship grows thin when balanced with the pressure of extraordinary innovation and accountability aims to establish sincere, honest and accurate responses. Peer assessments require clear and concise direction based on very specific guidelines. The popularity of this type of assessment is growing steadily yet must be cautiously understood by all who are participating and fostering relationships. Precise monitoring is also necessary to ensure that personalities are not intervening and redirecting desired results. The beneficial aspects of peer assessment are that students understand from the beginning of the assignment that they will be held accountable and responsible for their share of the work and that they will have the same responsibility for assessing the contributions of others. A variety of challenges can be avoided if very clear guidelines are provided at the outset of the lesson and all students buy into the process. In this case, there is even more reason to be very specific in providing directions and expectations so that one student is not pitted against another, and the process is devoid of any partialities. Analyzing classroom interaction data often focuses on teacher interaction and some studies have provided startling data that instructor dialog took up to 89 % of classroom verbal interaction time (Nunan, 1989). The PBL model specifically focuses on peer interaction and dialog in an effort to ensure that learning becomes student centered and beneficial. If EFL were the only consideration, then language would become the medium and content of instruction, so by combining PBL content and EFL acquisition, this challenge is alleviated particularly in the environment associated with peer assessment.

8 Self-assessments

Self-assessments, when done well, can have extraordinary results. Improving metacognitive skills can have profound impacts on student performance and learning (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). Students are required to tackle a major area that is often yet to be explored when the expectation is that they will provide careful analyses and assessment of their own unique accomplishments including EFL knowledge, content knowledge, and their overall

contribution to the PBL activity and assignment. Pertinent questions here include: Were they innovative and creative? Were they instrumental in seeking out and posing solutions to the problems? Did they contribute to the entire group on an equal or greater basis than others? What was their level of comfort in the process and the benefit derived from yet to be explored hypotheses? The optimum outcome is for students to be able to look at their own accomplishments and consider the contributions and aims to establish positive outcomes to the highest standards. The honesty during this aspect of assessment speaks to the positive relationship that the professor has established with the students and the standards set for self-assessment.

9 Reflective Journals

Reflective journals can have a tremendous influence on EFL and PBL skills because the learning is reinforced by writing and considering the syntax and content of the message that may be randomly scribbled during the PBL activities. Refinement of this writing is a valuable learning tool that provides thoughtful and captivating feedback and has the ability to underscore the key points derived from the PBL model. PBL “gives English Language Learners an opportunity to respond, read, write, and discuss oral and written English texts expressed in a variety of ways” (Nutta, Bautista, & Butler, 2011, p. 6). There is an expansive connection between journaling and language learning by comparing and even fictionalizing the world, and an invigorating journey is launched in PBL activities which includes experimentation and the resolution of world problems. Essential ingredients of PBL approaches and pedagogical resources impacting self-regulated learning connect EFL and PBL representing educational and personal growth.

This assessment method is full of expressive potential qualifying as a key contributor to PBL and EFL concepts. In order for it to work well and provide assessment that is vibrant and inspired, guidelines need to be carefully crafted and equally addressed by all students. The reflection should specifically address the process and the product of PBL and problem solving as well as English language learning acquisition and the roles of the participants.

10 Authentic Products

Comparative education research suggests that assessment techniques need to be applied to address the development of authentic products in PBL (Gijbels, Dochy, Van den Bossche, & Segers, 2005). This endeavor should be descriptive and include aspects that also focus on EFL accomplishments. Course design can change dramatically when cleverly engineered and irresistibly engaging products provide pathways for energizing opportunities. How did the group collaborate to develop a product or concept? Was the work engaging and did the group come up with a

Table 1 General factors that define methodology

General factor that define the methodology						
Continuous	Summative	Formative	Qualitative focus	Quantitative focus	Encourages collaboration	Independent skill building
	X	X	X	X	X	
		X	X			
X	X		X		X	X
	X		X		X	
X		X	X			X
X		X	X			X
X	X		X	X	X	X

unique solution? Did they carefully address each aspect of the PBL learning cycle making appropriate revisions and adjustments, seeking out information as they analyzed and modified hypotheses? Did EFL comprehension and vocabulary increase as they made observations and identified learning issues? What questions were identified to test hypotheses and pose solutions to problems? Did students conduct discussions in English to increase comprehension and verbal English skills? Spontaneously intriguing conversation and discovery often result from innovative design and an atmosphere that encourages fast-paced discovery and reinvents learning. The ease of conversation related to authentic products has the potential to produce a constellation of thrilling ideas tempting students to discover authentic products from a mix of cultures with a delightful array of creative opportunities. Students become eager to work collaboratively and there are added twists enriching the learning and increasing the options offered by a traditional lesson. Authentic assessment often becomes the most rewarding and celebrated assessment method when guided by a clear set of expectations and students are accountable yet given the leeway to innovate and incorporate a multitude of options and vibrant ideas. A whole new world of learning motivates students in an atmosphere that specifically reflects discovery and the innovation they are capable of achieving. High impact learning and cultural influences often blend focusing on the task at hand (Table 1).

11 Data Gathering and Analysis

PBL represents a multilevel approach to learning that emphasizes active rather than passive learning where the traditional roles of the student and teacher change and students assume increasing responsibility for data gathering and analysis that supports or refutes their hypotheses. During this process, it is essential to continue incorporating EFL strategies and English dialog and discussion of the relevance, coverage, and complexity of the data gathered. Critical and analytical thinking are encouraged to support problem solving that can take many directions. In this way,

opportunities abound for EFL development. As challenges are addressed and language learning problems tackled, EFL students will use strategies that are encountered in life and careers. Students acquire knowledge leading to proficiency in problem solving and language acquisition, and this prepares them to become practitioners.

Data gathering can result in the substantial improvement of the problem addressed, and it is an important aspect of assessment as it relates to PBL and EFL. Electronic assessment significantly expands horizons and its implementation has grown phenomenally as a way to complement other forms of assessment. Quantitative or qualitative methods can be used to investigate the cause of a particular situation. Extensive discussion will result based on data gathered and findings intentionally considered over time. Learning goals, objectives and assessment methods become part of the assessment plan that uses a clear scoring rubric. Writing samples may be incorporated to gather data, scoring elements of effective writing and summarizing results. Statistically, clear strengths and weaknesses with areas for improvement and a plan moving forward should be presented. Data can also be gathered focusing on student presentations of PBL projects. One aspect may be verbal communication assessment results focusing on presentation skills i.e. a list of important factors is designed including grammar, vocabulary, rapport, voice quality, eye contact, body language, command of material, organization, and the use of visual aids. Data is gathered and assessed on items using a ten-point scale where higher numbers indicate better performance.

Certainly, students are working with diverse projects and will need wide latitude with their projects. As a result, the qualitative aspects are crucial in terms of descriptive knowledge, strategic analysis and availability for EFL students to participate with respect to PBL goals and EFL competency. Well-designed data gathering plans will add credibility and validity to research findings (Hendricks, 2006). It is desirable that data gathering and analysis will lead to actionable recommendations clearly tied to the PBL situational analysis.

12 Connections to EFL Digital and Hybrid Learning

As we address PBL challenges and tackle language learning strategies that are encountered in life and careers, we must consider the inclusion of digital technology as it relates to education and new technologies. Students acquiring a strong knowledge-base in digital technology in the information environment become proficient in problem solving and language acquisition. There is a rapid evolution in the information age and PBL is enhanced by the use of laptops and tablets, mobile learning devices, as well as rapidly deployed software and apps. Blended learning and an integrated approach have become a successful pedagogical model and perspective for the future of education. These are further supported by gaming systems and social media that can impact EFL learning and assemble traces of humanity to the learning environment. Information Communications Technology

(ICT) interfaces with data gathering and an emphasis in analytics as it assumes a significant role leading and supporting PBL and EFL endeavors of the future. These modern tools for teaching are causing teaching methods and assessment to change significantly to the point that learning platforms are designed to maintain student grades and accumulate data for their educational records. Today's students are struggling to balance digital connectedness and personal isolation. This is why PBL is such a powerful factor in increasing the collaboration and teamwork essential for success in the workplace and helps transition learners into practitioners where students claim responsibility for their learning. Further transition from authentic materials to the live materials, podcasts and instructive interactive videos allows cultural learning and inspires a new range of assessment opportunities and activities by offering imaginary solutions and contextualized problems. Education games are often internationally acclaimed and can be incorporated into PBL activities to strengthen language learning through interaction. They are often the key to success when students are not challenged by traditional methods and support learning by motivating students and engaging them in the learning process.

13 Interdisciplinary Approaches

Intercultural communication and relationships are forged by interdisciplinary approaches and effective multicultural communication is synonymous with the contemporary world. There are many benefits to be gained from blending disciplines and enriching learning through innovative, unique twists to PBL and EFL methods. An integrated approach that activates multiple dimensions is critical for content acquisition (Banks, 1993). Complex and vibrant activities inspire students and assessment takes creative directions with unanticipated results that perpetuate the collaborative team member inside many students simultaneously immersed in EFL. Weaving interdisciplinary content with intercultural communication skills increases relevance and invites students to engage in intellectual contributions (Lee, Poch, Shaw, & Williams, 2012). Furthermore, the triangulation of examining student success in PBL and EFL from various standpoints maximizes the opportunity to assess learner and group accomplishments from a variety of perspectives (Brown & Rodgers, 2002). The flexibility to triangulate assessment and take an interdisciplinary approach will motivate EFL students in a variety of directions and assessments can be tailored to fit with the PBL and EFL activities designed to fit the interdisciplinary topics. Interdisciplinary collaboration often improves students' understanding and increases their involvement. In many science, technology, engineering, and mathematic (STEM) disciplines, the primary form of assessment is testing. By encouraging a more problem-based methodology, education and the understanding of knowledge transmission can be assessed on a number of different levels. A learning path that encourages interactive participation provides an

approach to teaching and learning that positively influences creative as well as formative techniques and enhances student centered trends in language and content teaching.

14 Integrated Professional Development

It is crucial for educators to have the opportunity for continuous professional development in both PBL and EFL learning strategies to improve teacher quality and positively impact student outcomes. A component of professional development must address the unique assessment strategies directly related to the aspects of this style of learning as well as the design of methodology and practice. Educators should continually ask themselves what is being measured and what the desired results are. PBL will serve as a strong tool for English language acquisition if used effectively. Professional development will encourage and motivate educators to purposefully design PBL activities and serve as facilitators as the learning becomes student centered rather than teacher centered. Self-directed learning with a team approach requires professional development to introduce methodology and pedagogy that fits the curriculum while also addressing EFL learning strategies appropriate for flexible inter-disciplinary settings. Professional development will also support more effective use of time in and outside of the 21st century classroom. Educators will be better prepared to differentiate teaching in order to meet the unique needs of students, PBL groups, and design appropriate assessment methods. In the case of blended PBL and EFL learning methods, professional development is essential to design progressively appropriate scaffolding pedagogy that takes into consideration the diverse technology skills of the educator and students strengthening the relationship between them while developing and accessing resources needed by both. When flipping the classroom, students may arrive more engaged and prepared to focus on the PBL activities, yet professional development is essential for educators to develop the critical skills they need to utilize this model. Challenges continue to occur and underprepared students underscore the need for appropriate professional development opportunities. During group professional development activities, educators have the ability to reflect on past practices and evaluate their own teaching. Relevant opportunities to communicate, share experiences, and capture new strategies for future endeavors are gained.

15 Conclusion

PBL and EFL heralds a bright future for both language learning and preparing students to become practitioners in an innovative workplace where collaboration and teamwork are valued. Implementation of this innovative constructivist multi-dimensional model engages and motivates all students, including those from

underserved populations. Assessment accompanying this highly cognitive methodology needs to recognize an experiential learning environment dedicated to problem solving and language acquisition. Learning takes place in critical stages and the useful combination of PBL and EFL needs to be adapted appropriately for success. As these techniques are incorporated with the emergence of additional technology tools and skills, it is important to effectively monitor and assess learning to challenge and motivate students as educators, reinvent learning, and establish a framework for success. When students are empowered and achieve global awareness of opportunities for effective learning and purposeful assessment, they contribute to excellence and create strong relationships that will serve them well throughout career and life experiences. As Gardner (1984) recommended long ago, it is crucial to create positive change by focusing on educational excellence and equity.

Collaboration and positive group dynamics serve to bring siloed initiatives together. While the assessment methods are not without controversy, the concepts are cutting edge and innovative. A balanced approach is key to providing educators with the autonomy to responsively source learning, enhance and establish a network of learners where discovery with a whole new world of learning evolves from a mix of cultures, subject specific content and language acquisition. Effective and valuable assessment appropriately adapted for learners will support and enhance this model and improve the educational experience for students. Let your journey begin providing support and flexibility for unique teaching methodology launching your students into an exhilarating phenomena that will serve them well in their future endeavors.

References

- Ambrose, S., Bridges, M., Di Pietro, M., Lovett, M., & Norman, M. (2010). *How learning works*. San Francisco, CA: John Wiley and Sons.
- Angelo, T., & Cross, P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). San Francisco, CA: (Jossey-Bass) John Wiley and Sons.
- Banks, J. (1993). The canon debate, knowledge construction, and multicultural education. *Educational Researcher*, 22, 4–14.
- Brown, J., & Rodgers, T. (2002). *Doing second language research*. Oxford, UK: Oxford University Press.
- Capon, N., & Kuhn, D. (2004). What's so good about problem-based learning? *Cognition and Instruction*, 22, 61–79.
- Downing, K., Kwong, T., Chan, S., Lam, T., & Downing, W. (2009). Problem-based learning and the development of metacognition. *Higher Education*, 57, 609–621.
- Driscoll, M. (1994). *Psychology of learning for instruction*. Needham Heights, MA: Allyn & Bacon.
- Finkle, S., & Torp, L. (1995). *Introductory document*. Aurora, IL: Illinois Math and Science Academy.
- Gardner, J. (1984). *Excellence: Can we be equal and excellent too?* (Rev ed.). New York: W. W. Norton.

- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*, 27–61.
- Hendricks, C. (2006). *Improving schools through action research: A comprehensive guide for educators*. Boston, MA: Pearson Education.
- Johnson, J., & Kean, E. (1992). Improving science teaching in multicultural settings: A qualitative study. *Journal of Science Education and Technology, 1*, 275–287.
- Lee, A., Poch, R., Shaw, M., & Williams, R. (2012). Engaging diversity in undergraduate classrooms. *ASHE Higher Education Report, 38*, 1–132.
- Mastascusa, E., Snyder, W., & Hoyt, B. (2011). *Effective instruction for STEM disciplines*. In Weimer, M. (Ed.). San Francisco, CA: John Wiley & Sons.
- Nunan, D. (1989). *Understanding language classrooms*. London: Prentice-Hall.
- Nutta, J., Bautista, N., & Butler, M. (2011). *Teaching science to English language learners*. New York, NY: Routledge.
- Scott, C. (1965). Teaching English as a foreign language. *The English Journal, 54*(5), 414–418.
- Soetaert, E. (1998). Quality in the classroom: Classroom assessment techniques as TQM. *New Directions for Teaching & Learning, 1998*, 47–55.

The Perception of Assessment as a Multilayer Dimension in the Armenian EFL Classroom

Marine H. Arakelyan

Abstract The study sets out to investigate how teachers and learners of English in Armenia perceive the multivoice notion of assessment as well as the role they ascribe to the practice of assessment in the teaching/learning process in a language learning environment. The specific interest behind the study was to explore the extent to which the two stakeholders of teachers and students view assessment as a triggering force to build future class content on the information elicited from the assessment of each preceding set of materials. A further effort was made to distinguish between the perception of teachers with and without prior exposure to assessment as a separate instruction unit to highlight the importance of its inclusion in the curriculum. With a view to eliciting qualitative data, questionnaires with nine open-ended questions were administered to students, both school-aged and at the undergraduate level, as well as to two cohorts of teachers, i.e. teachers who have completed their studies in MA TEFL programmes, and those with no explicit instruction in the assessment of English language learners. The analysis of the data revealed an outstanding feature of participants' views regarding the role of assessment. The responses of the teachers with previous exposure to assessment in language learning as part of their studies were in line with the notion of 'assessment for learning', whereas those by the other group of teachers and students indicated assessment and motivation for learners as being analogous.

Keywords Assessment · Evaluation · Motivation · Language learning · Grading

1 Introduction

The purpose of the current paper is to investigate how teachers and learners in Armenia perceive the ongoing, dynamic notion of assessment in the teaching/learning process, the extent to which they think assessing and being assessed impact their

M.H. Arakelyan (✉)
Sultan Qaboos University, Muscat, Oman
e-mail: marinesqu@gmail.com

day-to-day practice, the role they ascribe to assessment and some closely related facets of the issue. I put in the center of my perception the fact that assessment is a discipline of continuous inquiry that has developed in its own right as a branch of applied linguistics and is a very important and powerful area which should avoid and challenge the stereotyped, fixed and routinized way of understanding. It is worth mentioning that a variety of labels is used for students' achievement expectations. Some call them *goals* and *objectives*. Others refer to *scope* and *sequence*. Still others label them *proficiencies* or *competencies*. More recently, we refer to *standards* and *benchmarks*. These terms all refer to the same basic thing: what we want students to know and be able to do.

Whichever concept is considered to be an umbrella term for the students' achievement expectation, it is closely related to the boundaries of assessment perception. Many specialists in the field have proposed different definitions of assessment and have set forth various motives behind employing assessment in education. In the same vein, the focus of this paper has been on eliciting the attitude of local teachers and learners towards external and internal factors of plurality of views on using assessment in their teaching/learning process and to explore what their perceived reasons are that render assessment an integral part of the whole scope of education. It should be noted that the responses given by both of the cohorts of teachers and learners have been analyzed in light of the notion "assessment of learning" versus "assessment for learning". In addition, their views on the purpose of assessment were compared to those prescribed by Shohamy (2001). I do agree that one of the most challenging tasks for language instructors is finding effective tools to determine what and how much their students are actually learning. It is evident that students learn from the assessment itself as it is continuous and every activity has assessment of some type built into it. Thus, when assessment is an integral part of the learning process, students' learning improves. Hence, instructors need to think carefully about what kinds of knowledge and skills their multiple tools of assessment allow students to demonstrate.

The reason why I have been interested in revealing teachers' and learners' understanding of assessment is related to the fact that it has not received much attention in today's education in Armenia as a specific aspect of the teaching/learning process, and hence local universities/institutes preparing future teachers or specialists in the field of education do not have assessment courses integrated into their curriculum. Rather, assessing is assumed to be a "naturally occurring" component in the teaching process with no need to separately focus on it. Just as it seems to be assumed that anyone appointed to a teaching post in education can automatically teach, it is also implicit that they should be able to assess students' work. As a result, many teachers wield their red markers for the first time without ever having participated in any organized training in how to assess any specific skill of the learners within the subject area based on different elaborated criteria for the specific level. Many are embarrassed at the notion of even asking for any guidance, yet are quite threatened at the responsibility attached to the multiple tools of assessing. This concern has fueled my interest in contrasting the perceptions of assessment as held by teachers with exposure to special advanced

assessment courses and by those with no training. A further objective in the present study has been to compare the understanding of teachers and learners as the two stakeholders in assessment practice in the setting.

2 Theoretical Framework for the Study

Probably since the time the first test or the first grade was given, controversy has surrounded their use. For instance, some argue that grades “dehumanize” education and establish distrust between teachers and learners. Others say that grading and comparing students leads to harmful anxiety and minimizes self-esteem for those who receive poor grades. Even those who acknowledge the importance of assessment and evaluation often condemn current practices for the emphasis on testing basic skills out of context and the excessive competition that results. Regardless of the criticism and controversy rooted in the perception of this topic, the process of assessing and evaluating learners has persisted and continues to fulfill highly crucial tasks in the teaching/learning process that are vital to the effective organization and “smooth” flow of educational practice everywhere. Stiggins (2005) manifests a new belief related to the instructional decisions contributing the most to student success which are, in fact, not made by adults. Rather, the decisions that contribute the most to determining and illustrating student success or failure in the process of learning are made by the students themselves. The students decide whether the learning is worth the risk and how many attempts are required to achieve the desired learning outcomes. The students decide on the issue based on whether they believe they are smart enough to learn it. And they decide these things based on their own interpretation of their personal record of academic success.

Therefore, whatever else we do, we must help students believe that success in learning is possible for them and worthy of the effort they make. If we cannot do that, we cannot help them believe that they are capable learners. According to Stiggins (2005), students who participate in the thoughtful analysis of quality work to identify its critical elements or to internalize valued achievement targets become better performers. To make their decisions effective, students need continuous access to understandable descriptive information about their own improvement as readers, writers, problem-solvers, decision makers, listeners and interlocutors. This will lead to the greatest potential value of classroom assessment when teachers open the process up and welcome students in as full partners.

As Slavin (1991) signposts, tests and grades are needed to tell teachers, students and parents how students are doing in school. In light of this, teachers can use tests to see whether their instruction has been effective and to find out which students need additional help. Students can use them to see whether their strategies are paying off.

In hindsight, parents need grades to find out how their children are doing in school; in other words, for parents, grades serve as the one consistent form of communication between school and home. Grading and tests are really essential to

learning since research on the use of tests finds that students learn more in courses that use tests than in those that do not (Bangert-Drowns, 1986). However, even with the acknowledgement of the important role of assessment and evaluation, we assume that there are certain problems, and one of them is not evaluation per se but inadequate or inappropriate use of evaluation. As Shohamy has pointed out, “we need to maximize the role of tests as a means for obtaining information relevant to the improvement of learning, while at the same time minimizing their power and control” (2001, p. 141). The other problem with student evaluation is that we expect one grade or score to serve many purposes, yet an evaluation that is optimal for one use may be inappropriate for another. With a view to summarizing the uses that student evaluation is at the service of, we could come to the following definitions: incentives to increase student effort, feedback to students, feedback to teachers, students and parents. In making decisions about the assessment process, it will be useful for teachers to keep in mind the benefits that accrue to learners from sound assessment practices. Good assessment can be achieved through the gathering of information about learners over time, and through a combination of methods. A similar view has been set forth by Spolsky—“rather than expecting some simple mechanical device to translate the complex data of individual language proficiency into a single measure, language testers too would benefit from intelligent and responsible ‘interpretations drawn from patterns evident among the combined measures’” (1998, p. 10, as cited in Shohamy, 2001). Of course, these things do not happen on their own. It is up to the teacher to develop his or her awareness of assessment, to encourage learner awareness, and to make the process as effective as possible. Assessment should have what Tierney et al. describe as “a working relationship with teaching and learning”, and “students should view assessment as an opportunity to reflect upon and celebrate their effort, progress and improvement, as well as their processes and products” (1991, p. 21 as cited in Hedge, 2000).

3 Methodology

3.1 Participants

Participants were both students and English teachers in an Armenian education context. The students engaged in the study included school age and undergraduate learners. Sampling students of various age groups (13–22 years-old) was done with the aim of getting more in-depth descriptive data on the perception of assessment in the SL/FL classroom across an enlarged age bracket. The total number of participating students reached 10 in the present research.

There were sixteen teachers participating in this study, eight of whom were MA TEFL teachers who took an assessment course during their C-TEFL program and are currently doing Advanced Assessment in an MA TEFL program. The other

eight were secondary school teachers who had not taken any courses in assessment. The objective behind the selection of the sample of two groups of teachers was to examine any differences of understanding in this ongoing, multivoice inquiry and reveal the plurality of views on assessment between these two groups of teachers. The inclusion of students in the study was done to compare the views of teachers and learners on the same issues. All the participants voluntarily participated in the questionnaire completion procedure.

4 Methods

The current study used a descriptive/qualitative method as we believed that a qualitative approach could help us gain insight into how the understanding of assessment by teachers and learners functions on the ground, i.e. in the Armenian ESL classroom setting. Whatever the reasons attached to assessment by policy makers, external evaluators, testing authorities, etc., teachers and learners shape their own understanding of assessment which heavily impacts how they treat it on the insufficient bases within current hierarchy in assessment. With a view to better comprehend why grading and evaluation takes a certain form in a classroom, we have to go to the conceptual basis that triggers this very form. This basis is closely linked to the beliefs and assumptions held by those who are directly engaged and affected by the outcomes of the activity in question–assessment.

The instruments were questionnaires administered to both students and English teachers. Two different questionnaires were constructed; one was designed for student respondents and the other one was designed for teacher respondents. Both the student and the teacher questionnaires consisted of 9 open-ended questions to which participants were asked to provide their opinions based on their own experience. The same questions were included in both of the questionnaires with only some adjustments to suit the purpose.

5 Procedure

Participants were informed about the study, its goal, procedures and length. Data collection was conducted through the questionnaires described above. The teacher questionnaires were either mailed to participants or hard copies were directly handed to teachers. The students were provided with hard copies of the questionnaire. Half of them completed the questionnaires in English, while the rest answered the questions in Armenian. These were later translated by the researcher.

6 Data Analysis

When asked to give their personal definition of the term “assessment”, the teachers with no previous exposure to an assessment course (for the purpose of this paper, we will be using the terms “non TEFL-er” to refer to teachers with no formal training in assessment, and “TEFL-er” to those with previous exposure to an assessment course) gave more or less homogenous replies. It should also be noted that these teachers confused the notion of “assessment” with that of evaluation, as evidenced by responses to the first question. Three of these teachers see assessment as a way of evaluation or making judgments about some work done, whereas two of them view assessment as a device to both punish and motivate learners. Others consider assessment to be a contributive force in an effort to reach the goal in the teaching/learner process. This last point is somehow in line with what we referred to as “assessment for learning”, while the answers provided by the rest of the teachers are indicative of their understanding of “assessment of learning”.

With regard to the responses given by the “TEFL-ers”, these reveal a broader understanding of the concept of “assessment”. Assessment is not regarded as having motivational or punitive functions only; rather, exposure to the assessment courses during their two-year study has helped them shape a sound perception of what assessment could mean in a classroom context. Some have stated that assessment is “self-development” both for teachers and learners, and others have argued that assessment is a diagnostic device to check where teachers and learners have reached on their road to the set objectives and how to better organize the “journey for the rest of the road”.

Regarding student participant responses, we have observed that they liken the notion of “assessment” to that of grades, and they consider grades as the appreciation and reward for their efforts inside and outside the classroom. Some respondents even pointed out that though they study for themselves and not for the sake of grades, they see grades as quite a necessary part of study at school since it is only through grades that they get an idea of how well they can perform.

As can be seen from the discussion above, there is difference between what “TEFL-ers” and “non TEFL-ers” believe about assessment. This is what I had assumed and what instigated my genuine interest in doing the current research. As to the comparison of the students’ definition of assessment with that of the “non-TEFL-er” teachers, though the former group used the word “motivation” and the latter “reward for efforts” in their responses, it is believed that the views held by both are similar in their essence. The teachers think that grades motivate students to study better. Similarly, if students say that grades are a sign of appreciation of their efforts, they also consider that being graded acts as a stimulus for their further learning. However, when comparing the students’ perception with that of the “TEFL-er” teachers, we can observe difference in their responses. Teachers ascribe more “attributes” to assessment, whereas students’ responses allow us to judge that they view grades as incentives for their learning only.

The second question of the questionnaire evoked short answers by the “non-TEFL-ers” where they have mainly stated that assessment is indeed important for the teaching/learning process. However, most of them have failed to specify the reasons for its importance, and only two have stated that it is important to keep the students “alert”. On the contrary, “TEFL-ers” see the importance of assessment in that it provides valuable feedback of how the teaching/learning process goes, where there are gaps, and how to better address them, at the service of promoting learning. However, they have also mentioned that assessment is important inasmuch as it is done in a non-threatening manner. As to the students, they again reinforced their view of “assessment for motivation” in their replies and have not gone beyond this. Insufficient information provided by them and the “non TEFL-ers” gives us very little ground to make any comparison with the sample of the “TEFL-er” teachers.

The role of assessment by the “non TEFL-ers” is described as encouraging learners to do better in all learning cycles, while “TEFL-ers” see the role of assessment in the provision of key information for making instructional adjustments, e.g. to improve teaching methodology, to introduce new teaching materials that are more relevant to learners’ needs, etc. Besides, some of them mentioned that the role of assessment lies in revealing learners’ strengths and eliminating any possible weaknesses that may become known as a result of assessment. As to the students, they perceive the role of assessment in their learning as having only a motivating function. A mere look at the answers given by the three groups makes it clear that “non TEFL-ers” and students ascribe an “encouraging and motivating” role to assessment, while “TEFL-ers” see assessment as an instrumental tool in their day-to-day practice in the work with their learners.

To summarize the responses for the preceding two questions, I could claim that the perception of the students and the “non TEFL-ers” on the purpose of assessment has little to do with those proposed by Shohamy (2001) where she puts forward *prediction, placement, categorization, acceptance and/or rejection, provision of feedback, following progress, motivation, establishment of discipline, power exercise, accountability and research conduction* as the objectives behind the employment of assessment in the educational domain. On the contrary, the views of “TEFL-ers” are in line with the various purposes assessment may be utilized for.

In response to the question of whether assessment contributes to their teaching practice or not, the “non TEFL-ers” stated that it helps to arouse the learners’ interest in learning. One teacher has stated that assessment is of no help in her teaching without giving any explanation for such an attitude. The contribution of assessment to teaching practice is perceived by “TEFL-ers” as helping to reveal information about salient features in the EFL classroom. In light of the results of the assessment, they constantly revisit their previous performance and the underlying beliefs/judgments to make it more suitable for the given situation. With respect to students, the analysis of their answers reveals some interesting findings. Though all of them agreed that assessment helps them keep on learning, two of the answers were of much interest. One of the respondents stated that, if she is given a grade that she thinks is higher than her response should have gotten, she feels responsible for “deserving” that grade and hence studies hard. Another respondent mentioned her

concern about being assessed, since she thinks she only “studies for grades”. Even though she comprehends a text after having read it once, she keeps reading it as many times as necessary for reaching “the level of an excellent mark”. Again, more or less similar answers were noticed by the “non TEFL-ers” and students, while “TEFL-ers” demonstrate a deep understanding of how assessment can contribute to the teaching process.

The next question related to whether assessment interferes with teaching practice elicited different answers among the “non TEFL-ers”. One of the respondents mentioned that the main “impediment” is the grading system currently used in Armenia: the grading scale ranges from 1–5 giving no possibility of correctly grading the student who “falls in between”. Another interference as perceived by these teachers is that assessment is a time-consuming activity. A further reply concerns the fact that assessment interferes with the teaching practice inasmuch as teachers and students become enemies: students always want higher grades, whereas teachers tend to give much lower grades. This last point is highly relevant to the situation in Armenia as assessment in schools has been described by some as a “corruption tool” in the hands of the teachers.

Many of the “TEFL-ers” think assessment does not interfere with their teaching. This can be accounted for by the fact that they perceive assessment in its real sense and hence apply assessment at the service of facilitating, and not hindering, student learning. As to the students, they interpreted the term “interfere” differently from what was initially intended. They look at the word positively and mention that interference of assessment is associated with motivation only. Only one of the students said that she gets disappointed whenever the teacher gives her a mark that is lower in relation to her response. As with the analysis of the previous questions, it becomes evident that the opinions of “non TEFL-ers” and students match to some extent. On the other hand, “TEFL-ers” do not see “serious” interfering features embedded in good assessment.

When including the question “Do you think you grade your students fairly?”, I was aware of the fact that the respondents would give only positive answers to it. My expectation proved to be true for both of the groups of teachers. However, some have mentioned that they may sometimes be subjective in their grades assigned to students, but only in favour of the latter. They explain this as a means of trying to motivate students to perform at their best during coming classes. There is only one answer given by a “non TEFL-er” that differs from the rest in that the respondent “confessed” grading students whom she is not familiar with more fairly than those whom she has known for a considerably longer time. She accounts for such a biased approach by mentioning that the personal characteristics of a student affect her ability to be impartial while grading. The majority of the students think they are graded fairly. Only two of the students reported about teachers being subjective in assigning grades.

The answers provided by the “TEFL-ers” indicate that all of them support the holistic approach to assessment when the grade reflects more aspects of the student’s work than a single performance. This is due to the fact that they are well aware of the fact that a “cumulative” look at a learner’s progress gives more a

realistic picture of his/her knowledge, whereas single-performance assessment has a number of faults. The same approach to assessment holds true for the “non TEFL-ers” as well, and all of them have mentioned about grading their students with a number of factors being considered by them. The students can be equally divided into two groups; the first group believes that their teachers grade them based on single performance, while the second group thinks they are graded on the basis of previous performance and some personality features they are most liked for. Therefore, we can state that both groups of teachers were homogenous in their answers with half of the students sharing such an opinion.

The “TEFL-ers” share a common view on the 8th question in that they all think that a grade/score expressed in a *number* cannot be indicative of the actual knowledge a student may have. Rather, they highlight the fact that students’ real knowledge is demonstrated throughout the entire process of teaching/learning, and a numeric symbol cannot be representative of this ongoing process. With regard to “non TEFL-ers”, it should be noted that half of them are inclined to think like “TEFL-ers”, while the next half mentioned that the grades they assign to their students do reflect their knowledge. Those who do not think that grades can reflect the actual knowledge possessed by the student refer to time constraints as being the main reason. They claim that, since they ask only some questions for the purpose of grading a student, they cannot arrive at a complete understanding of how much this student knows beyond the questions asked. Here we see that both groups of teachers think that a single grade cannot represent a student’s actual knowledge; however, they rationalize their answers by setting forth quite different reasons for their opinion. As to the students, some of them think that grades do reflect their knowledge. The other students have stated that, since some teachers do not grade fairly, these grades are not a real representation of what they know.

The “TEFL-ers” see assessment as an indispensable part of the education process and think that assessment is a prerequisite for the effective organization of education. It is worth pointing out that they consider assessment as the driving force behind educational goals. Their responses communicate their understanding of assessment as promoting the teaching/learning process rather than hindering it. Similarly, the “non TEFL-ers” consider assessment as an essential component and state that, without any assessment, no education can be organized efficiently. However, they do not rationalize their answers. One respondent among the “non-TEFL-ers” indicated that there could still be education without any assessment if schooling was not compulsory, and if the enrollment was on a voluntary basis with the learner’s wish being the reason for engagement in schooling. The analysis of students’ responses has demonstrated that, again, they confuse grading with assessment. They have mentioned that without grades education could not be well organized since they need grades to study. Without any grades, they will lose their incentive for learning, and the whole process of teaching/learning will end up in a complete failure.

7 Conclusion

It is believed that teaching is a public affair, and we get all sorts of feedback regarding how well or how badly we teach—even without deliberately seeking feedback. The expressions on students' faces, the attendance at our classes, and the level of students' performance all help us to adjust our teaching techniques. This is what is referred to as “assessment” in the literature dealing with this specific aspect of teaching. Though it is believed by many practicing teachers that assessment and evaluation are targeted at revealing the learning endeavors or at motivating learners to strive for better accomplishments, it is worth highlighting that we should see more to assessment and evaluation; they are done to mark the progress of the joint efforts of teachers and learners in their collaborative acts of moving forward. If we believed in the transformation mode of teaching, then we would also accept the view that assessment is a device to check where students have arrived in their learning efforts. But I do have a firm stand in that the teaching/learning process is accomplished through the equal participation of learners and teachers in the classroom and, therefore, assessment is a perception of the final destination where the learning outcome takes the journey to its end. Admittedly, the teacher is in the leading position of that teaching/learning journey, but, without the passengers, the journey would not have taken place at all. This leads to more tangible perception related to the concept of assessment for learning being appropriate in all situations that helps to identify further steps to build on success and strengths as well as to revisit the point of departure to get rid of drawbacks and correct weaknesses.

In light of this, given the importance of assessment, it is probably the aspect of our profession that should be scrutinized in a more careful and professional way. Even with the best of intentions, external examiners and moderators can only contribute a limited amount to the processes of assessment. Yet, the primary responsibility for assessment continues to rest with teachers as every separate method is good when it is used by a teacher involving all the students by turning them from passive listeners into active participants. It is worth stating that some student-teacher conferences serve as structured or unstructured audits of student achievement, in which the objective is to talk about what students have learned and have yet to learn. In this way, teachers and students talk directly and openly about levels and scopes of student attainment, comfort with the material the students are mastering, specific needs, interests, and desires and/or any other achievement-related topics that contribute to an effective teaching and learning environment.

One of the disadvantages of the whole aspect of assessment is the fact that, very often, students receive very little or no feedback about their performance which negatively affects their further learning. Another drawback is poorly set exams that stimulate surface learning because, after taking these exams, students merely get ready for upcoming exams in another subject. On the other hand, the techniques for taking exams are considered to be too important. More specifically, exams aim at

measuring how good the test takers are at answering exam questions, rather than how well they have acquired the skills or can demonstrate learning outcomes.

I do believe that we, as teachers, should understand that language is an open socially-embedded construct, divergent, rather than convergent, in nature, and if we once begin to tolerate the ambiguities and inconsistencies which can exasperate the more dogmatic types of learners and shake their stereotypical ways of understanding, we will find that our own teaching and/or learning will benefit enormously. Indeed, this socially-embedded view of language should reflect on socially-embedded views of testing. It is in its own right an independent course of enquiry with longitudinal understanding which has problems because it has been characterized by a lack of cohesive understanding of test specifications. Thus, we should look at assessment as educational necessity, as a reflective process. In addition, there is a need to develop testing competence that involves being able to formulate and implement advanced practices that are designed to provide a more definitive understanding of how and why language learners have shown the outcomes they have.

It is worth mentioning that one of the most useful benefits of assessment can be feedback gained by students on their performance regarding skills they are intended to develop, and their understanding of theories and concepts. Assessment is often the major driving force which leads students to serious studying. In general, both positive and negative assessments should, in my opinion, be made available to the learner as honestly as possible; mainly because this is what learners feel, and say, they want. However, it is essential for such assessment to be given in an ambiance of support and warm solidarity, so that the learners feel that the teacher's motive is honestly to promote and encourage their learning, not to put them down. The concern of negative assessment is not the assessment itself, but rather the accompanying implications of aggression on the side of the assessor and humiliation on the side of the assessed—which can, and should, be eliminated (Ur, 1996).

To be able to “get the best” out of assessment practice, teachers need to be deeply taught about the principles and priorities of assessments, as it is not a part of the teaching skill that automatically gets into the teaching repertoire of a teacher. Teachers come from a variety of backgrounds with a diversity of beliefs about assessment. This is readily reflected in the way they view assessment and hence assess their students. To minimize the detrimental effect that assessment can have on the learners' “fate”, the inclusion of special assessment courses in the curriculum of higher educational establishments should come at the top of their agendas. This will contribute to having future teachers shape a sound understanding of assessment, which, in turn, will positively affect their performance in the classroom. If we succeed in having good teaching and good assessment go hand in hand, then we have little to concern about the welfare of our society.

References

- Bangert-Drowns, R. L. (1986). Review of developments in meta-analytical method. *Psychological Bulletin*, 99, 388–399.
- Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford, UK: Oxford University Press.
- Shohamy, E. (2001). *The power of tests*. Harlow: Pearson Education.
- Slavin, R. E. (1991). *Student team learning: A practical guide to cooperative learning*. Washington, DC: National Education Association.
- Spolsky, B. (1998). *What is the user's perspective in language testing?* Paper presented at the Colloquium, The State of the Art in Language Testing: The Users' Perspective, National Foreign Language Center, The John Hopkins University, Washington DC, June 15.
- Stiggins, R. J. (2005). *Student-involved assessment for learning*. Columbus, OH: Pearson Prentice Hall.
- Tierney, R. J., Carter, M. A., & Desei, L. E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon Publishers.
- Ur, P. (1996). *A course in language teaching*. Cambridge, UK: Cambridge University Press.

Part III
Skill-Area Assessment

EFL Assessment: Assessment of Speaking and Listening

Seetha Jayaraman

Abstract Assessment is a measure of the level of teaching and learning achieved. It has a heightened emphasis in the case of learning a language which is not used as a Second Language (L2) or language of communication in the learners' communities. In an English as a Foreign Language (EFL) learning situation, exposure to spoken language is either minimal or lacking completely outside formal instruction environments. The case under examination in this chapter is the EFL learning situation of Arabic speaking undergraduate students in Oman's Dhofar region. The study aims to investigate the methods and criteria involved in testing the English speaking proficiency level among a group of Arabic speaking undergraduate students for whom English is a foreign language (FL) and hence use it sparingly in real life situations. The reasons for this limited use vary from: personal preference in using Mother Tongue (MT), comfort levels, lack of relevant vocabulary, and the need to be able to communicate fluently. Regardless of the objectives of learning and using English in its spoken form, and the opportunity to listen to the different varieties of English, either at the stage of formal instruction or in informal situations, the users' ability to speak the language always involves a gap between the standards to be attained and the standard attained at different stages of their learning. This chapter seeks to shed light on how well the standards are achieved among the participants.

Keywords Assessment · Evaluation · Listening · Perception · Production · Proficiency · Pronunciation

Seetha Jayaraman (✉)
Dhofar University, Salalah, Oman
e-mail: seetha.jay@gmail.com

1 Introduction

Listening and speaking are two aspects of the same skill and hence need to be treated, taught and assessed using similar criteria, viz., the ability of perception or reception and the ability of production respectively. Effective training in perceptive oral skills should lead smoothly to an equally effective productive oral skill. However, the prevalence of local languages as the medium of communication has reduced the scope for cultivating good spoken English among the subjects being studied. Nonetheless, at the university level, proficiency in spoken English forms a major module of instruction, involving testing and qualifying criterion to achieve the required standard at different levels.

Today the world has effectively decreased in size on account of the rapid growth of telecommunications (and the role of media and developments in audio-visual aids), and the need to listen and respond to spoken English has become crucial. Increased travel and tourism far and near makes the need to understand and react to spoken English very important. Furthermore, increasing job opportunities at home and abroad necessitate the importance of good spoken English skills among job seekers. Hence the skills of listening and speaking should be measured and quantified through appropriate assessment tools. The chapter presents techniques in the evaluation of listening and speaking skills in an EFL classroom, as a part of formal and informal assessment of proficiency in the target language.

2 Listening

Language relates primarily to the ability to understand and speak. The development of other skills of reading and writing are comparatively easier once the speaking skill is acquired. Listening and speaking are the key aspects of aural-oral skills. As expressed by Baruah (2001), the following elements make learning listening skills successful.

1. Recognizing the characteristics of English speech sounds in isolation and in combination.
2. Distinguishing such sounds from similar sounds in the Mother Tongue (MT).
3. Understanding the lexical meanings of words in context and the grammatical meanings of structures.
4. Understanding the meaning conveyed by stress and intonation patterns.
5. Grasping the intent and mood of the speaker and the theme of the discourse.
6. Anticipating words and structures from the context for understanding speech at normal conversational speed.
7. Guessing the meanings of unfamiliar words from the context.

Listening is not an independent activity. In a way, we can say that listening is the first step to learning to communicate. Our ability to speak closely depends on our

ability to listen carefully. Therefore, in the process of communication, listening is the input element and speaking is the output. Listening is an integration of different language learning skills. There is a difference between hearing and listening. That is, listening is a conscious act and so the message is retained in the mind longer, while hearing is an unconscious act and dissipates from the memory. Hence, it is important to develop listening to be an efficient communicator.

As we know, language is made up of sounds, words and structures. Following this idea, to be able to communicate orally, one must have the ability to distinguish between the contrasting sound units in the target language. The normal tendency of a learner, when they come across sounds which do not exist in the mother tongue, is to substitute the sound with the one closest in his mother tongue. These unfamiliar sounds pose production problems. Therefore, learning English as a Foreign Language (EFL) should begin by listening to the characteristic English speech sounds, including stress and intonation patterns.

Listening is done for two purposes: (1) for pleasure, as in the case of literature, like listening to poetry recitation, conversations, news bulletins, media programs, debates, lectures and the like; (2) for knowledge, as in the case of listening to intellectual discourses, seminars, and interviews with experts in different fields. Regardless of the objectives, listening remains a fundamental component of language learning. Therefore, attention must be paid to the variety of listening activities a learner is exposed to, and which ones he should focus on. Some useful activities to practice in listening are,

- understanding the meaning of words
- following instructions, guidelines and directions
- interpreting stress and intonation meaningfully
- understanding the importance of listening in learning to speak fluently.
- listening to media programs like news readings, interviews and group discussions.

Most of the programs we are exposed to through audio-visual media are designed for entertaining or gaining knowledge. A similar tendency can be noticed in the classroom activities for listening and speaking. Listening is often taken for granted as part of a speaking lesson. In order to produce utterances which are intelligible (even through imitation), in terms of stress, rhythm and intonation, an additional feature called 'juncture' contributes significantly to speaking. *Juncture*, in simple terms, is the time gap between words or phrases, which helps interpret the meaning or the intent of the speaker. An example of juncture is the time gap between the two sets of words '*I scream*' and '*Ice cream*'. The time gap in uttering the two words determines the intent of the speaker or can cause misinterpretation of the meaning by the listener. Having looked at the objectives and the needs of listening as a perceptive skill to master speaking, we can also understand the role of testing listening, along with speaking at the end of a language learning session.

Testing listening is a pre-requisite of a course in learning speaking. A listening assessment makes a learner aware of the areas of his learning which need to be

focused on. It may or may not be their weakness, but it is a gap in one's learning through attention and conscious effort that he has to concentrate on.

Listening can be tested to:

1. obtain specific information
2. get the gist of the content
3. follow instructions correctly
4. draw inferences from contexts with verbal clues.

The testing items can be used for perception or for comprehension. Either way, the tests of listening through aural or visual techniques encompass the overall communicative ability through listening, which is the passive skill necessary for effective speaking. These tests are useful for remedial and special courses. Special courses are designed for weaker students and there are no formal syllabus or exam schemes.

3 Speaking

When languages evolved out of meaningful sounds, they were spoken words and not written. In the process, the human ear was important in listening to the sounds and sound combinations, pauses and unspoken words. More meanings emerged from word groups phrases and spoken sentences, with reference to varying situations (actions, plans, experiences etc.).

In learning EFL, acquiring sounds, sound clusters and correct pronunciation is important. In fact, good pronunciation is the evidence of good education and culture, reflecting sometimes the social status of the speaker. The English language has some unique sounds and sound features such as stress and intonation. There are other advanced features of pronunciation like elision, assimilation and so on, which are different from the local languages and have to be learnt correctly, if one has to attain proficiency and fluency in spoken English.

4 The Importance of Speaking in Second Language (L2) Acquisition

The main purpose of teaching speaking skills is to improve the students' communicative competence and the ability of learners to express themselves accurately and appropriately in different situations. In order to develop their speaking abilities in another language, L2 learners need to have an adequate vocabulary, pronounce words correctly, use word and sentence stress, organize their thoughts in a meaningful and logical sequence, access and use the language quickly, and confidently master the syntax.

Developing the ability to speak in a second language involves the development of communication skills. Oral language and oral processing skills are different from reading and writing skills. Speaking is an interactive process of constructing meaning that involves producing, receiving and processing information. It can mean to pronounce correctly (linguistic competence) or to understand and produce language effectively (communicative competence).

Lee and Van Patten (2003) stress that L2 learners must develop interactive skills and the ability to negotiate meaning. Both these skills require a learner to speak correctly and to be understood correctly by their interlocutor. This concept of communicating meaning via expression, and interpreting and negotiating meaning in a given context has also been reiterated by Savignon (2005). These opinions regarding expression and interpretation of meaning in language in its oral form include correct pronunciation patterns in the L2. To this effect, the learners must take the initiative and responsibility to work towards good pronunciation.

A preliminary observation of a group of undergraduate and graduate students learning EFL who belong to the Dhofar region shows that the learners wish to communicate in the English language, provided that they can get positive results instantaneously. This means that they expect to be able to comprehend and communicate in English fluently, without being at a loss for the vocabulary needed for their active conversation. The results of an assessment of the learners' speaking skills used in the classroom and in the language laboratory are evidence to this attitude. This chapter presents the observations and issues related to students' perception and performance in relation to listening and speaking skills. Evidence is supplied via the assessment techniques used to evaluate their proficiency levels, and the challenges faced in the process of assessing these skills are outlined.

5 Evaluation

Primarily, the distinction between 'assessment' and 'evaluation' must be made clear. Evaluation is a process to attach value to the learner. After selecting the 'what' (curriculum) and the 'how' (methods) in teaching and learning, the evaluation process begins. Thus, evaluation, is a continuous process, and assessment is a phase in this process used to make valid judgments about the learners' progress.

Evaluation is done along specific criteria to make a judgment about the student's progress and performance in general, in relation to the prescribed plan and the learning outcomes envisaged. It helps the teachers in planning and designing specific instructional programs to identify the learning goals.

Evaluation is done using certain techniques and tools. Some of the techniques that can be used for evaluation include: observation, student self-assessment, daily practice assignments, quizzes samples of student work, tests, holistic rating scales, projects, oral and written reports, performance reviews and portfolio assessments.

The common characteristics of a good evaluation are:

- It is a continuous process and helps to measure educational achievement and improve it.
- The technique of evaluation and the tools employed must be valid, reliable, objective and practicable.
- It includes both academic and non-academic subjects.
- It is a procedure for improving the product.
- It is used to discover the needs of an individual and design learning experiences.
- It is used to fulfil educational purposes.
- It is a complex process.

For instance, a listening task for a formal assessment must be context-based to promote interaction and stimulate motivation to develop the skill further among the participating learners. The task should be made authentic and focused on listening through negotiating meaning. It should be constructed to guide the listener through overall comprehension and must be designed to be purposeful, have clear procedures and a tangible outcome. It must also lend itself to evaluation of performance.

A valid test for listening can adopt the bottom-up approach (through word-sentence recognition, listening for morphological endings), top-down approach (through listening and identifying the topic, understanding the meaning) or consist of interactive tasks (through listening to a list and categorizing words or following directions).

6 Evaluation of English Language Proficiency in Dhofar University—Criterion Level

The learners' first step on the path to the graduate courses begins at the Foundation Program. Since English is the medium of instruction, the approach to teaching of English at the university is systematic, meaningful, and purposeful. The English language courses offer the incoming students (with low proficiency in English) an intensive program to help them pursue their studies in the major of their choice through the medium of English with the aim of immersing students in the English language.

The policy of placing the students in the placement test is given in the following table:

- 0–39 Level-1
- 40–55 Level-2
- 56–70 Level-3
- 70 and above—Major.

Students who score 70+ on the English Test and those who produce a certificate of IELTS with a band of 5 or an equivalent certificate are exempted from studying in the Foundation Program. Such students begin the credit courses associated with their Major.

The Learning Outcome Standards (LOS) for the English Program are as follows:

a. Listening

- Follow spoken instructions in order to carry out a task with a number of stages.
- Listen to a conversation between two or more speakers and be able to answer questions in relation to context, relationship between speakers, register (e.g. formal or informal).
- Take notes and respond to questions about the topic, main ideas, details and opinions or arguments from an extended listening text (e.g. lecture, news broadcast).

b. Speaking

- Actively participate in a discussion on a topic relevant to their studies by asking questions, agreeing/disagreeing, asking for clarification, sharing information, expressing and asking for opinions.
- Paraphrase information (orally or in writing) from a written or spoken text or from graphically presented data.
- Prepare and deliver a talk of at least 5 min. Use library resources in preparing the talk, speak clearly and confidently, make eye contact and use body language to support the delivery of ideas. Respond confidently to questions.

A student's academic performance is assessed throughout the semester using various instruments: homework, exams, research papers, projects, practical work, research, etc. The student has the right to receive continuous feedback about his/her performance. The instructor must complete a mid-term performance assessment to give students a chance to withdraw from the course before the end of the withdrawal period and to help faculty advisors better advise students during the pre-registration period (if implemented). Students who score sufficiently high scores in the Level 3 examination are admitted to the university programs (year 1).

Some of the different types of tests which are used to test specific skills in English language (e.g. speaking and communication) are:

1. Production and Recognition tests: Production tests measure creativity. For example, examinations in speech and/or writing. Recognition tests tend to rely on identification of specific items, like multiple choice and listening tests.
2. Language sub-skills tests and Communication skill tests: Testing communicative competence involves testing for grammatical discourse, socio-linguistic and elocutionary competence as well as strategic competence. It requires the learner to use the language naturally for genuine communication and to put authentic language to use within a context. It should be direct and should test the learner in a variety of language functions.

7 Assessing Listening Skills

Testing language skills—Listening: The following test-types can be used to test listening:

1. Phonetic perception
 - Phoneme discrimination tests.
 - Tests of stress and intonation.
 - Listen and draw tests (oral instructions carried out by the learners).
 - Multiple choice questions in listening (a short-talk/passage/directions and answering questions).
2. Take an aural comprehension exercise.
3. Complete the sentence when a choice is given (ex: It's raining. I'll have to take a (raincoat/textbook/envelope).
4. Answer questions according to time or direction (ex: (LuLu Supermarket) where did you go last evening?)
5. Identify the central theme or the nature of a talk when listening to a news-broadcast (ex: social/scientific/education).
6. Play the role of a listener/a speaker in a dialogue.
7. Indicate whether the two sounds are the same or different.
8. Indicate minimal pairs of words (ex: send/lend; friend/fiend; true/through).
9. Identify the stressed syllable (I/never/wanted to/work/for money).
10. Indicate the intonation pattern, i.e., rising or falling intonation (ex: 'Did you go to school? 'Where did you go?)
11. Follow a request or a command (ex: Go to the board, write your name and then erase it).

In administering any of the above tasks, the greatest challenge faced by the instructors is in making the learners distinguish between sounds involving the same place of articulation like /p/ and /b/ or /f/ and /v/. Consequently, the identification of minimal pairs at the perception level is affected both in classroom activities and during testing. This leads language instructors to believe that prolonged training and extremely cautious and focused efforts on the part of the learner is needed. This is possible only through willing and motivated participation and drive by the learners.

8 Assessing Spoken Skills

As indicated by Lee (2000, p. 35) the test of an oral task can be structured along the following criteria:

1. Identify a desired information outcome
2. Breakdown the topics to sub-topics

3. Create and sequence concrete tasks for learners to do
4. Build in linguistic support.

In the traditional testing method, learners are usually assigned a specific topic as an open-ended question and are asked to talk about it for a stipulated time. But, in the communicative approach to testing, the topic is presented as an exchange of information task either in the form of a dialogue, interview or a role-play (see Appendices 1 and 2 for sample tasks).

Testing the students' command of spoken English is important in judging their overall language competence. But this is not given the importance it needs, perhaps due to the difficulty to administer the test and interpret the results with concrete evidence of the learning level attained, unlike in the case of reading or writing. According to Cyril Weir (1983, cited in O'Sullivan, Weir, & Saville, 2002), the following factors are to be considered while testing spoken language.

1. Operations
2. Performance Conditions
3. The expected level of performance

By Operations, it is meant that the functional areas of the spoken language are being tested, for instance, informational or interactional language aims involve presentation, description, comparison, narration, instruction and so on. Whereas, interactional function involves the active use of spoken language in routine face to face conversations, telephonic conversations, interviews, meetings or lectures and so on. It also involves the reciprocal ability to use the receptive and productive skills to negotiate meaning, rephrase to impact clarity, react, role-play and the like.

Performance Conditions is a parameter of assessment used for testing speaking abilities. For example, the utilization of the time allotted, carrying out uninterrupted speaking on any given topic, presence of an audience, familiarity with the situation etc.

The level of performance directly relates to speaking ability and focuses on the fluency, appropriateness (style or register), effectiveness (as per the needs of the situation such as reasoning), coherence (organization, sequencing, connecting information, re-ordering in case of gaps in speaking or presentation), accuracy (using appropriate vocabulary, stress, tone and intonation), correctness in terms of intelligibility and grammar, articulation of sounds and sound sequences/clusters) and the ability to employ a wide range of vocabulary (substituting synonyms, meanings and paraphrasing).

There is a widespread belief by some teachers and language specialists alike that assessment of a spoken language should be based on specific tasks, which can assess the learner's individual speaking ability. The following are some specifications, along with examples of effective tasks to assess spoken language skills:

Specifications

- A test of speaking should not be of more than 15 min' duration.
- The task must contain different items for different elements of the task being tested.
- In case of fluency testing, a pre-speaking time range of 1–5 min must be allotted to compose and organize the ideas to speak on a given topic.

Tasks

- Use pictures to stimulate ideas.
- Use themes or an imaginary situation as prompts—these can help generate ideas or encourage giving opinions. The prompts can contain a controversial topic or idea to test the ability to present an argument or counter-argument.
- Story-telling from a given series of picture windows.
- Guess the ending of the story from a series of pictures to check the imagination.
- Connect and interpret information presented in disconnected pieces.

Assessment can also involve individual or group performances of the kind where the learner initiates, controls and moderates a group discussion or develops an idea individually presented to him in a group, on a variety of topics in different imaginary situations. This also brings out the spontaneity of ideas and the effective use of relevant vocabulary in a given context. Although this is more a test of logic and timeliness in conversation (as they are unstructured conversations), they help to assess the test taker's selection of appropriate vocabulary and mastery of intonation. This will test the linguistic competence in terms of language skills, speaking skills in terms of articulatory capacity of the speaker, and the communicative competence in terms of the logical arguments and information presented logically and relevantly in the responses. In this sense, the testing of the spoken skills is purposeful both in evaluating spoken skills and communicative language skills. This type of testing is more suitable for job interviews and testing the ability to interact with groups or, participate in team activities, in addition to testing the confidence levels of the speakers in a group. Another advantage of this method of testing is that, it is less time-consuming when the objective is to test a large group of learners, where the learners' performance can be tested both individually and collectively at the same time, and the judgment of the assessment would be more reliable, since the performance is based on authentic and meaningful situations.

A more common testing task is role play by a pair of learners in a given situation. A variation which can be introduced is to improvise the given situation using a register/language variety or degree of formality which is unusual in the given situation, for example, "the use of informal register and the mock Parliament session".

Among the criteria to be used in testing spoken language, we can include,

- Appropriateness—limited responses, inappropriate vocabulary, lexis, socio-culturally inappropriate usage, use of slang.
- Adequacy—vocabulary, limited, gaps, lexical inaccuracies.

- Grammatical accuracy—broken sentences, bits of information, incomplete sentences, ungrammatical utterances.
- Intelligibility—Pronunciation, stress, rhythm and intonation, use of appropriate tone.
- Fluency—Using bits of information, halting, hesitation, gaps in using cohesive devices like connectors or conjunctions.
- Relevance of content—Coherence, relevant, limited or sufficiently covering the task.
- Gestures—used to supplement or to substitute language output. While supplementing language with gestures has a positive effect on the listener, it can reveal the lexical incompetence or inadequacy of the speaker.

Judgment based on any or all the above criteria using any of tools and techniques mentioned must conform to the conditions of reliability and validity. The results of the assessment are reliable when they are measurable in terms of grades, scores, or empirical scales. They are valid only with adequate evidence of the tape-recorded or recorded criteria used for testing, along with a detailed description of the tools used for the procedure.

The group of learners being studied is being tested on their competency in speaking through tasks consisting of pre-assigned topic, prompts, interpretation of graphic information and oral presentation using audio-visual aids. Their performance is assessed using criteria as stipulated in the rubric (See Appendices 3 and 4 for the rubrics of a speaking test), which also details the allocation of the relative weight of marks for each criterion.

The following sample tasks can help in evaluating the speaking skills of the learners.

- (1) Repeat sentences of varying lengths.
- (2) Take one of the roles in the dialogues.
- (3) Answer questions either when specific instructions are given or without a cue.
- (4) Ask direct questions when an indirect statement is given (ex: ask me how I got to school this morning).
- (5) Transform sentences according to the direction given.
- (6) Formulate questions on a page.
- (7) Describe what they see in a picture.
- (8) Give a summary of a story they have read.

Overall, it is observed that the learners are able to perform reasonably well on a pre-assigned topic, through preparation and rehearsing. Interestingly, they are more enthusiastic and are observed to be more confident with oral presentations using visual techniques, which cover up their limitations of vocabulary and their grammatical incompetency. But where speaking involves face-to-face interaction either with the peers in the presence of the moderator, or a direct conversation with the instructor, they are halting, hesitant, self-conscious and do not meet the expected standard.

9 Conclusion

To conclude, listening and speaking are the important components of good communication. A good speaker must be a good listener. Intent listening results in good perception and equally good comprehension. This is the key element in developing communication skills, especially good conversational skills. The term skill implies an ability, which needs regular and continuous practice. English language proficiency testing systems across the globe are attempting to evaluate this ability in order to make inferences regarding the proficiency levels of English users. Its development and assessment, both formative and summative, have therefore never been more important.

Appendix 1: Speaking Task Sample Test (1) Level 3

Task 1

1. Prepare a **dialogue** on a familiar topic with a partner. It must demonstrate the following features:
 - Initiate conversation
 - Asking and answering questions
 - Asking for clarification
 - Concluding conversation
2. **Research and talk about a current news item.** It can be from the newspaper, magazine or the internet. You must give a 2-minute talk explaining the main points of the item you have chosen. Your talk must demonstrate the following features:
 - Provide the reference
 - Show clear understanding of the topic
 - Use vocabulary that the class can understand
 - Be able to answer questions seeking clarification

- Paraphrase successfully
1. Give a **presentation** of 5 minutes about a career you are interested in pursuing or the major you will take at the university. It must demonstrate the following features:
 - Evidence of research
 - Effective use of visual aids (charts, posters, pictures, power point)
 - Clear information and ability to answer questions related to topic.

Appendix 2: Speaking Task Sample Test (2) Level 3

We begin speaking with people by greeting them and asking questions about them. In this first part you will ask questions about yourselves such as:

- Yourself
- Your town/city
- Work/study
- Your family
- Your house
- Your free time
- Your reasons for learning English
- Your plans for the future

Task 1

Ask questions to get the following information when you meet for the first time:

1. A new student or classmate -What he or she is doing?

_____?

2. Your new teacher -About himself or herself and the background?

_____?

3. Your friend-Where is he or she is from?

_____?

4. A Tourist -Which places would he or she like to visit?

_____?

Task 2

Complete the following conversations between two speakers A and B.

1. A _____?

B _____ I don't like it very much. I think it is really difficult, especially grammar.

2. A _____?

B _____ I often go out with my friends, but sometimes I enjoy just reading in my room.

3. A _____?

B I usually go to Muscat. But next holiday I want to travel to Dubai.

4. A _____?

B I have one brother and two sisters.

5. A _____?

B I m studying Math, I.T. and English in the Foundation Program.

Appendix 3

Student’s Name: _____ Level _____ Section _____

Total Marks _____/30

SPEAKING TEST RUBRIC-I

THINGS WE LOOK FOR		
Pronunciation	All words are clearly understood	4
	Most words are clearly understood	3
	Some words are clearly understood	2
	Not many words are clearly understood	1
Intonation	Interesting to listen to	3
	Sometimes interesting to listen to	2
	Boring to listen to	1
Use of Grammar	Meaning is clear	3
	Most meaning is clear	2
	Confusing	1
Fluency	Can keep speaking	3
	Stops and starts with pauses	2
	Only gives 1-2 word answers	1
Developing confidence	Appears confident	3
	Trying to be confident	2
	Makes listener uncomfortable	1
Content	Excellent ideas	4
	Inte resting	3
	OK ideas	2
	Not interesting	1
Task Fulfilment(Paraphrase graphic information)	Excellent interpretation of graph	4
	Very good attempt at explaining graph	3

	Good attempt at explaining graph	2
	Confused explanation of graph	1
Vocabulary (Use of target language)	Demonstrates excellent use of vocabulary	3
	Attempts to use target vocabulary	2
	Limited attempt to use target vocabulary	1
Use of visual aids	Interacts well with visual aids	3
	Attempts to interact with visual aids	2
	Doesn't incorporate visual aid	1
Things that were good:		
Things that I can work on to improve my grade:		

Appendix 4: Speaking Test Rubric-II

1	Pronunciation and clarity	1 2 3 4 5
2	Fluency of presentation	1 2 3 4 5
3	Body language and confidence	1 2 3 4 5
4	Development of ideas	1 2 3 4 5
5	Quality of information	1 2 3 4 5
	Total	

Additional comments

Instructions:

Circle the number which best reflects the student’s performance.

Key to evaluation scale:

5= excellent; 4= good; 3= acceptable; 2= needs improvement; 1= poor

Appendix 5: Mid-term Speaking Test Topics

1. In the teacher's office

You had something important to discuss with your teacher. You went to his office during his office hours. Act out the conversation that took place between you.

2. In a shopping centre

You wanted to buy something from the Lulu Centre. Enact the exchange that took place between you and the shop assistant.

3. At home

You had an important thing to discuss with one of your family members. Report the exchange that took place between both of you.

4. In the cafeteria

One day you sat in the DU cafeteria with a friend or a classmate of yours and you had an interesting chat. Tell us what took place between you two.

5. In a restaurant

You invited a friend to some dinner in a good restaurant. The waiter brought you the menu. Tell us the exchange that took place between you and the waiter.

6. Over the phone

You received a call from a friend. The conversation that took place between you was interesting. Enact the same conversation with a friend.

References

- Baruah, T. C. (2001). *The English teacher's handbook*. New Delhi: Sterling Publishers.
- Lee, J. (2000). *Tasks and communicating in language classroom*. New York McGraw-Hill.
- Lee, J., & Van Patten, B. (2003). *Making communicative language teaching happen* (2nd ed.). New York: McGraw-Hill.
- O'Sullivan, B., Weir, C. J., & Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing*, 19(1), 33–56.
- Savignon, S. (2005). *Communicative competence: Theory and classroom practice*. New York: McGraw-Hill.

Assessing Entry-Level Academic Literacy with IELTS in the U.A.E.

Christopher Morrow

Abstract While the IELTS exam has earned a reputation for its usefulness in many international educational contexts, it does not appear to fully address the specific needs and issues of students and universities in the United Arab Emirates. IELTS was introduced in 2004 at the main national university, United Arab Emirates University, to complement existing assessment tools in the English Foundation Program. Since then, its influence on instructional and assessment practices has grown at that university in ways that do not seem to be completely congruent with the needs of low-level English learners who are about to begin studies in English-medium universities. Evidence for this comes from the claims IELTS makes about the test, the national results of UAE students, research on English-medium instruction in the UAE, and survey results from English lecturers in the UAE University Foundation Program. This chapter argues that the advanced linguistic demands of the IELTS exam, its equal weighting of scores from the four skill-based sections, and the exam's general communicative orientation are not well-suited for making valid and reliable decisions about the readiness of Arab students to begin college-level studies in English. The disproportionate effect of testing in UAE educational programs makes it imperative that a more appropriate means of assessing students is selected or developed. Alternative assessment options are considered which address these issues more directly and efficiently.

Keywords IELTS exam • Language proficiency testing • Validity research • Washback • English-medium instruction • Admissions testing

C. Morrow (✉)
U.A.E. University, Al Ain, UAE
e-mail: gomorrow@gmail.com

1 Introduction

The rapid introduction of tertiary-level English medium-instruction (EMI) in the United Arab Emirates (UAE) has had major effects on its colleges and universities. The impact this policy shift has had on content learning, Arabic literacy, and English language development has been problematic in several ways, especially since overall levels of English proficiency are quite low (McLaren, 2011). Much of the legitimacy for this policy change is linked to the IELTS testing system as a means of determining students' readiness for EMI in colleges and universities. Unlike most other countries, UAE colleges and universities usually accept a low score of overall Band 5 as evidence of readiness for EMI (Gitsaki, Robby, & Bourini, 2014). It is unclear if this demanding and sophisticated test is appropriate for this purpose considering its high-level texts, its limited ability to address local cultural perspectives, and its effects on classroom instructional practices (*viz.*, washback). It appears that IELTS is being used in ways that are not in total accordance with the realities of undergraduate study in EMI programs in which students have ongoing language development needs. Survey data from English lecturers at United Arab Emirates University (UAEU) is summarized at the end of this chapter to identify some apparent discrepancies between the IELTS exam requirements and students' developmental needs.

This chapter focuses on the use of the IELTS exam at UAEU, the country's main national university (which is referred to as a 'test user' in this chapter). While prospective students normally take at least one course in the UAEU's Foundation Program (FP), a growing number are achieving the minimum requirement (IELTS Band 5) on their own and skipping all or part of the FP English course sequence (Moussly, 2012). This development puts more pressure on institutions to assess students accurately, both by choosing an appropriate test and by setting appropriate minimum requirements in the specific sections of the tests. IELTS has had a disproportionate influence on English teaching throughout the region recently; however, its usefulness is unclear relative to the types of English proficiency needed for success at institutions like UAEU.

2 Entry-Level Testing

Language tests for admissions are crucial for setting the norms of EMI and determining its viability. If education is inherently dependent on language and communication, then achieving adequate levels of English at entry strongly affects the results of EMI (Doiz, Lasagabaster, & Sierra, 2013). Although many UAEU students are still learning intermediate-level academic English, the university typically follows a late immersion approach to education which occurs almost completely in English. At a similar university, Rogier (2012) reported that teachers found the weak writing and listening skills of their students forced them to modify

their instructional and assessment methods which limited their overall effectiveness. As the leading national university, UAEU has programs like linguistics, philosophy, and medicine which are inherently demanding both from cognitive and linguistic perspectives. Several researchers at UAEU have described the obvious difficulties that this causes when students have low levels of English achievement (e.g., McLaren, 2011; McLean, Murdoch-Eaton, & Shaban, 2013). My own teaching in English language education is particularly constrained by students' weak reading skills so I am sympathetic toward colleagues who resort to summarizing as much as possible for students in PowerPoint presentations in order to make sure students comprehend essential course content. Technology offers some new options (e.g., audio and video recordings of lectures), but I consistently have students who lack the necessary literacy skills to manage the course requirements. Ironically, students with the lowest English skills often choose the humanities and education even though these fields have some of the highest linguistic demands.

3 Searching for a Proficiency Test

Prior to 2004, when the IELTS was introduced at UAEU, undergraduate students were nearly all graduates of its own English FP. High school students were placed in one of three levels in the FP by means of a national standardized test called the Common Educational Proficiency Assessment (CEPA). Typically, a very small group of about 5 % got high CEPA scores and were allowed to sit for a TOEFL in order to skip the FP English track completely. The vast majority spent 1 or 2 years in the FP program where they were evaluated using many types of assessment activities which emphasized the achievement of specific English course outcomes. Evidence for students' readiness for EMI came from the FP's conventional tests and assignments, so those who passed the program's highest level of English (Level 3) were deemed eligible to begin their undergraduate coursework. These types of assessments resembled many of the projects and tests they would face later on, so they were authentic in the sense that they were based on classroom activities and academic requirements. This authenticity was valuable, but measurement errors were possible at the two extremes of student performance. The first was that strong students may have lacked appropriate opportunities to show they were qualified to bypass the FP. The second, and more serious one, was that weak students sometimes got through the FP after failing and repeating several times even though their true language proficiency levels were still inadequate. Implementing a reliable external exam seemed like an objective way to address the assessment needs of very weak and very strong students.

4 Introduction of IELTS

TOEFL and IELTS were the leading candidates for becoming university entrance exams at UAEU in 2004. The paper-based TOEFL available at that time was relatively academic in nature and used only multiple-choice items to assess listening, reading, grammar, and structure. In contrast, IELTS incorporated all four skills and adopted a more communicative approach to language proficiency in line with the Common European Framework of Reference. This may have appealed to administrators since it gave students credit for their functional and oral abilities, skills which were not assessed well by the normal FP instruments. Furthermore, IELTS was already widely used throughout the region for university admissions purposes, and several teachers in the FP at UAEU had experience as raters of its speaking and writing sections. Given the intense focus that the government placed on improving English, there was reason to believe in 2004 that standards would eventually rise and IELTS requirements to enter EMI courses would reach levels that were similar to the rest of the world.

EMI had already been commonplace at UAEU for a few years in 2004. The first cohorts to take the IELTS in 2004 only needed to reach a benchmark score of Band 4.5, and 85 % of those who passed Level 3 attained this low requirement. While most other UAE public colleges soon set their minimum score at Band 5, UAEU did not raise its requirements to Band 5 until 2011. By that time, about 60 % of students were normally reaching the overall score of Band 5 (Morrow, 2005).

By setting Band 5 as a minimum score, public colleges and universities in the UAEU are catering to students who are clearly at the low end of the spectrum of IELTS candidates globally. IELTS publishes score results on their website which show how poorly UAE students tend to perform on an international level. Globally, only 10 % of all candidates received scores of Band 5 or lower in 2012 on the Academic version of IELTS (IELTS, 2012). In contrast, 72 % of candidates in the UAE scored in that low range. Similarly, just 12 % of UAE candidates obtained Band 6 or above in 2012, but 76 % of the world-wide cohort achieved that level (IELTS, 2012). These figures cover every candidate who sits for an IELTS in the UAE for any reason; consequently, they are not necessarily an accurate reflection of the nation's overall English proficiency level. Nevertheless, they suggest that UAE students may not be the main target group for the IELTS test. Although IELTS has a well-earned reputation for test quality, the relevance of available validity evidence to low scorers has not been conclusively established. Most testing programs address this issue by producing an array of exams which target successive levels of proficiency. IELTS may function well across several levels but it is very unlikely that it can measure proficiency across all nine levels with equal levels of accuracy.

5 Early Costs of the IELTS

Implementing the IELTS exam on such a large scale required a great deal of support from UAEU staff, lecturers, and administration because of its burdensome financial and logistical issues. At a current cost of \$239 (USD) per candidate, the exam is relatively expensive and time-consuming. Examining hundreds of candidates often takes several days because of the speaking interviews. It is common for the FP to lose 1 week of instructional time at the end of a term to compensate for the fact that IELTS is only available intermittently (typically about three times each month). Preparing students to sit for the test consumed a fair amount of class time in Level 3 since IELTS tasks employ a wide array of question formats. Test preparation became a dominant aspect of many Level 3 classes because students needed to face two major exams at the end of the school term: the standard course exams immediately followed by IELTS. Finally, UAEU was forced to follow the IELTS protocols for handling the results. This included making students wait about 10 days for verified scores to arrive from abroad. These burdensome procedures appeared justified at the time as a way of identifying a small group of low-achieving students who were truly unprepared for EMI despite their success in the FP courses. Interestingly, however, most of those students had actually failed major parts of the FP exams; their weaknesses were identified by existing assessment practices, but they ended up passing the course because of high scores on other assignments.

6 Early Benefits of the IELTS

There were some immediate benefits which seemed to justify IELTS as a benchmark test. For the first time ever, the actual achievements of UAEU students could be compared to others worldwide. Of course, the test challenged even the best students in the FP and this increased their motivation to study much more than the relatively easy FP tests they were used to. By including a speaking section, IELTS helped document a skill area which had been largely ignored previously. FP administrators were able to use IELTS results to check the concurrent validity of their own tests and assessment instruments. Perhaps the greatest benefit of the IELTS for the whole nation was a logistical one: any UAE student could visit their local IELTS testing center and find out if their English was adequate to attend the main national university without attending the English FP. Outsourcing assessment to an international organization brought benefits but unfortunately the costs associated with this practice were considerable.

Of course, the challenges of EMI at UAEU have little to do with the IELTS exam per se; instead, they are readily explained by the fact that low scores are accepted for admission to EMI programs. The current minimum requirement, overall Band 5, is categorized by IELTS as a “Modest” user of the language. This is defined as follows: *The Modest user has partial command of the language, coping*

with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field (IELTS, 2013). From the perspective of IELTS, no academic courses should be attempted by anyone who does not have at least IELTS Band 6.5. Similarly, IELTS claims that no training courses should be attempted by anyone scoring below IELTS Band 5.5, and even then the courses should be less linguistically demanding. It is possible that these rigorous guidelines may be directed primarily at institutions in English-speaking countries where instructional accommodations for language learners are few. If so, UAE colleges might contend that the linguistic or academic demands of their programs were lower so a Band 5 is adequate. Nevertheless, the exact nature of the linguistic demands in the EMI programs of UAEU are poorly understood (Rogier, 2012). The education students that I teach appear to struggle to read 20 pages a week for a standard course, especially when more technical or abstract language is used (e.g., theoretical discussions of language learning). Belhiah and Elhami aptly described this dilemma in their extensive survey project in which they found that, “The current EMI situation leaves much to be desired with students struggling to learn the subject matter due to their low-proficiency in English” (2014, p. 1).

7 Argument-Based Test Validity

Although the IELTS is widely-recognized as a valid test for the purpose of college admissions, traditional notions of validity as a characteristic of a test itself have been recently revised by many in the field of language testing. Leading figures in this area have proposed that test validity needs to be seen as encompassing all phases of the testing process, including the interpretation of scores and the consequences of subsequent decisions based on those scores (Bachman & Palmer, 2010; Weir, 2005). In the traditional model, reaching a minimum score in IELTS represented a certain level of English proficiency that was sufficient evidence for acceptance and rejection decisions. Argument-based approaches, however, call for additional types of direct and indirect evidence to confirm the soundness of these interpretations and decisions. For example, Weir’s socio-cognitive framework divides such validity investigations into two main areas (Weir, 2005). The first, context-based validity concerns the correspondence of the test items and tasks to the larger domain of target language use (viz., undergraduate EMI). The second, theory-based validity, deals with the soundness of the linguistic and cognitive processing that students engage in. For example, rapid or expeditious reading is a major challenge for low-level candidates on the IELTS exam, but the importance of this kind of reading in current theories of reading and the EMI context of UAEU is unclear.

8 Validity Threats

A good test must have an appropriate mix of tasks and items so that valid decisions can be made based on scores from the tests. Testing experts identify two main threats to test validity: the inclusion of items assessing irrelevant knowledge and skills (construct-irrelevant variance), and the inadequate use of appropriate tasks and items (construct underrepresentation) (Fulcher & Davidson, 2007). A possible example of the first type of threat in the UAEU context is the fact that the weight given to speaking in the IELTS score formula is equal to the other three language skills. Speaking was rarely assessed at all in the FP before IELTS was introduced in 2004, presumably because the other three skills were much more highly valued for academic purposes. Relatively high speaking proficiency scores frequently raise the overall IELTS band scores for some students with low academic literacy. The IELTS website reports that UAE candidates do much better as a national group on the speaking section than the other sections: the speaking mean score for the country is 5.3 but the other skills have means of 4.6, 4.7 and 4.8 for listening, writing, and reading respectively (IELTS, 2012). Orality is highly valued in Arab culture and conversational English predominates in many cities in the UAE (O'Sullivan, 2004; Troudi, 2009). Therefore, many young people have opportunities and incentives to develop their speaking, with the possible exception of females from rural areas. Consequently, students with good speaking skills seem to benefit on IELTS exams in ways that may not be consistent with the FP emphasis on academic literacy. Furthermore, the spontaneous type of speaking assessed in the IELTS interviews seems relatively rare in EMI contexts. Academic discussions and presentations usually occur in contexts where participants have had many chances to organize their speech relative to course readings and topics.

Of course, ignoring speaking does not agree with current conceptions of communicative competence, but there are sound reasons for proposing that the underlying knowledge and skills it requires are quite different in nature than the other three skills. Gu (2014) analyzed TOEFL iBT scores using structural equation modeling and found that two latent components accounted for the results: the ability to speak and the other three skills. Including speaking scores in a language test may interfere with measurement of the second, and arguably more important in the present context, latent variable. Although effective speaking is highly valued in fields like business, giving it equal weight with the other three skills appears inappropriate for academic discourse in the UAEU context. Other UAE public institutions (e.g., Zayed University) have partially addressed this issue by specifying minimum skill area scores for the IELTS and insisting that none of the four Band scores fall below 4.5. This policy is an appropriate way to prevent students who are very strong in one skill (usually speaking) to compensate for severe weaknesses in other areas. However, the expense and inconvenience of testing speaking needs to be reconsidered if speaking scores contribute relatively little to the measurement of academic language proficiency.

A second possible example of construct irrelevance in the IELTS is the inclusion of very difficult reading items that are far beyond the levels of UAEU students. These items are relevant to the broad claims IELTS is able to make about very high levels of proficiency, but they are a source of frustration to UAEU candidates since they encourage students to resort to various guessing techniques. The IELTS website states that a Band 5 in reading on the Academic IELTS is normally achieved by getting at least 15 out of 40 questions correct (IELTS, n.d.). UAEU passers can easily succeed by only answering thirteen items correctly to get Band 4.5 in reading. Of those thirteen, it seems plausible that some guessing techniques can help students succeed on five or six of the items, especially the ones that only have three possible choices (i.e., True/False/No Information). In other words, low-level IELTS candidates may only appropriately demonstrate their actual reading ability by answering about ten out of forty questions correctly. Even though the question formats on IELTS usually minimize the effects of guessing, relying on such a small number of items to assess reading is obviously unwise considering its importance in the EMI context. The skills and strategies needed to improve the accuracy of one's guesses on the difficult IELTS items probably have little to do with normal academic reading processes (Haladyna & Downing, 2004). Nevertheless, successful use of guessing strategies could have a major impact on one's score when the proportion of difficult questions is so large. The widespread popularity of private IELTS language institutes in the UAE indicates that certain kinds of testwiseness are highly valued here; several of my students have told me that guessing was a key to their success. The key issue in this context is not the use of guessing methods per se but the fact that they probably have a disproportionate effect on results given the very small proportion of items that candidates need to answer correctly.

Another possible threat to the context-validity of the reading exam is the type of texts it uses. IELTS reading passages are characterized by a lack of titles, section headings or graphic and typographic aids that are considered essential for top-down processing in reading comprehension. This demanding kind of plain text allows testers to use many item formats which assess students' ability to identify main ideas and text structure. However, such reading passages are the direct opposite of the user-friendly formats which are so common in popular textbooks. Learning to refer to common text features is considered an essential literacy skill for native speakers because they make the reading process more efficient and effective when reading to learn is the focus. Therefore, the ability to read plain text passages may be viewed as a form of construct-irrelevant variance in the UAEU context.

A similar example is the forms of background knowledge needed to succeed on IELTS reading and writing tasks. On one hand, good background knowledge is considered an essential component of literacy so reading comprehension suffers when students lack basic knowledge of major historical, social and scientific trends. On the other hand, certain topics and tasks seem to present unusual challenges for the UAE population because they reflect issues and concerns that are rare in this society. In her study of UAE students sitting for the IELTS reading test, Freimuth (2014) found that the cultural themes related to religion and social values seemed to

interfere with students' reading comprehension. English is increasingly being viewed as an international language in the UAE so the importance of using background knowledge from Western cultures is questionable. The Middle East is an important market for the IELTS program; nevertheless, Arab students have different sorts of background knowledge than students from Asia and elsewhere. Fulcher and Davidson (2007) consider this kind of differential item functioning to be a source of construct-irrelevant variance, and the ability of IELTS to account for this while serving the needs of many different regions of the world is questionable.

9 Construct Underrepresentation

In the case of low-level reading and writing on the IELTS exam, separating the two types of validity threats is not that easy. As mentioned in the previous example, plain text reading passages are clearly very rare in college texts but the processes involved in reading them are somewhat germane to some kinds of academic reading. Similarly, the use of many difficult, high-level reading tasks on the test is not an obvious error, but it is associated with a corresponding reduction in the number of items that are accessible to students with lower comprehension levels. The IELTS reading test would obviously serve the UAE population better if it did not underrepresent lower-level tasks and items. It is very likely that IELTS relies mainly on challenging skills (e.g., reading for inferences) to assess those above Band 7. This seems very appropriate, but it may result in low-level items being underrepresented on the exams. UAE professors like this author recognize the difficulties faced by students of modest ability when performing high-level reading skills. IELTS reading tasks draw on complex combinations of lexical knowledge, syntactic knowledge, discourse knowledge, etc. If they did not, the test would have to be three times longer so that each of these areas of knowledge could be individually assessed at all the relevant levels. This test design, however, leaves low-level items in a minority on the test even though UAEU professors are probably more interested in this aspect of reading skill.

The IELTS reading test is technically impressive because of its ability to discriminate at so many levels of proficiency in just 40 questions. The IELTS writing test has fewer objectionable elements than the reading test, but it also has fewer positive aspects. The main IELTS writing prompt is a simple statement eliciting an essay of 250 words that is worth about two-thirds of the writing score. Opinion essays are commonly used for this section, and students must rapidly produce concepts and language that are relevant to the given topic. The specific wording used in this simple prompt statement is of utmost importance since it sometimes contains a phrase or a concept that is unfamiliar or confusing to many UAEU students. Of course, this undermines the reliability of the test in serious ways that other researchers have recognized (e.g. Gebril, 2009). However, my major concern at this point is the way the two writing tasks underrepresent the other forms of writing that are characteristic of undergraduate EMI programs. Undergraduates are

rarely asked to write using only their personal background knowledge and linguistic knowledge. Moore and Morton (2005) investigated this issue by analyzing 155 writing assignments at Australian universities. They concluded that IELTS tasks have some resemblance to the university essay as a genre, but they are mainly non-academic in nature. The most common academic genre, writing from sources, was used as a task type in older versions of the IELTS but was abandoned in the new versions of the test. Integrated writing tasks which combine reading and writing are clearly more authentic than opinion essays when it comes to undergraduate studies.

Green (2006) compared two types of writing courses to investigate this issue: IELTS preparation courses and English for Academic Purposes courses. He found many similarities between the two, but the former usually avoided writing from sources in favor of genres demanded by the IELTS test: descriptive writing (for Task 1), and evaluation and hortation (for Task 2). He found that IELTS classes emphasized reproducing information from graphs and from memory while EAP classes were more cognitively demanding since they stressed integrating information from sources. Both types of writing can be challenging for low-level students, but the absence of writing from sources is a troubling form of construct under-representation in the case of IELTS.

10 Validity Research

To their credit, IELTS has sponsored many studies examining the validity of the exam, most of which lend support to its common usage: selecting high-intermediate level and advanced students for western universities. Nevertheless, the amount of validity evidence that pertains to its specific applications in the UAE is relatively small. It is necessary to collect validity evidence in order to confirm the suitability of a test for a specific purpose, and Davies (2011) emphasizes that this is the responsibility of test users, not researchers or test developers. As informed test users, UAEU administrators realize how weak a Band 5 candidate can be, but they have done relatively little to investigate the effect of poor English skills on the classroom discourse and academic achievement of their EMI students.

This author found only one validity study that specifically addressed the reading processes of lower level candidates. Weir, Hawkey, Green, and Devi (2012) assessed the cognitive reading processes used by 352 subjects by using verbal retrospective techniques. Due to the unusually large number of subjects involved, the researchers were able to validly compare weaker readers (Band 5 and below) with intermediate and stronger ones. The results generally support IELTS validity by noting many more similarities than differences between reading processes and strategies reported by the Band 5 candidates and the other groups of candidates. For example, all groups claimed that their most common strategy was: *quickly match words that appeared in the question with similar or related words in the text*. More evidence like this is needed to ensure that low proficiency candidates in the UAE

are being assessed appropriately. Weir (2005) outlines several useful techniques for collecting such data. They include detailed examinations of the reading and writing tasks required by universities in order to establish their similarities to test tasks. At UAEU, humanities students are often seen carrying around thick anthologies of Western literature which seem far beyond their normal reading levels. The specific reading tasks associated with such texts need to be fully analyzed in order to determine how well they correspond to IELTS reading tasks.

11 Hijacking the Curriculum?

Before the IELTS was introduced in 2004, getting ready for undergraduate EMI was the obvious mandate of the FP at UAEU. Coordinators and instructors consulted occasionally with others (especially colleagues in the English for Academic Purposes program) about the linguistic and academic needs of students, and they also relied on their professional judgment when planning courses and tests. Unfortunately, these local curriculum development processes began declining in importance after 2004 when educational progress was largely defined in terms of IELTS scores. Preparing for the demands of the IELTS tasks became a major focus of the final level of English in the FP, and the normal coursebook was replaced by an IELTS preparation book called *Focus on IELTS Foundations* (O'Connell, 2007). Higher-level secondary students who only need one term in the FP may be ready for this approach, but weaker students probably suffered when test preparation became the focus of classes rather than dealing with the obvious deficits in their global English development. From an educational perspective, it may be plausible that some IELTS preparation activities can be integrated with normal language development activities. Nevertheless, there are far more positive ways to achieve balanced language development using methods that are consistent with best practices in language education, academic literacy training, and formative types of holistic assessment (Pilgreen, 2007). The IELTS exam seems poised to become a key assessment tool for English programs in secondary schools across the UAE but evidence suggests that an emphasis on learning test-taking strategies for the exam may divert students' attention away from balanced language development and preparation for EMI (Gitsaki et al., 2014).

The FP at UAEU attempts to serve underprepared, at-risk students while at the same time challenging those who are ready for some fundamental aspects of academic literacy. Rather than using a test like the IELTS as the arbiter of success, educators need to analyze the exact nature of EMI practices and tasks in various departments and programs of UAEU. In the last few years, I have served on two university-wide committees addressing the challenges of helping undergraduates to continue developing their English in the mandatory English for Academic Purposes courses. Both groups concluded that academic literacy was the major need of most students, and one made a policy shift to replace a module on oral presentations with an increased emphasis on academic writing. Although faculty members from

several departments participated in these committees, some resisted the idea that professors were responsible for helping students improve their English. Only a few from the humanities (e.g., literature) have consistently recognized that ongoing language development is a key aspect of their educational mission.

12 Parallel and Converging Assessment Systems

Since its implementation in 2004, the IELTS has had a major effect on how English is taught and tested in the FP. When it was first adopted, FP exams and IELTS exams seemed to complement each other since the former stressed achievement while the latter assessed proficiency. In recent years, however, FP tests have begun imitating many features of the IELTS test to the extent that the current FP writing exam is a virtual copy of the IELTS writing exam. The FP reading exam still has sections for grammar and vocabulary that represent achievement more than proficiency. Nevertheless, it has become difficult for FP instructors to see their role as extending much beyond that of serving as IELTS test preparation specialists. As educators with graduate degrees, they are qualified to help students develop a wide repertoire of educational skills and strategies. However, their current program narrows that focus almost exclusively to IELTS exam performance. Since instructors in the highest level of the English FP are responsible for both IELTS preparation and course exam preparation, I surveyed them to determine their views of the two types of assessment instruments. Most of the instructors had been hired in the past decade and had only known the dual exam system currently used by the FP. Others, however, recognized that the FP had enjoyed more freedom in the past to teach and test in ways that were more consistent with the needs of future EMI students.

13 Reading Teachers' Views of IELTS

I first surveyed reading teachers to determine their global attitudes toward the IELTS exam as an appropriate assessment tool given the normal demands of the Level 3 curriculum, the instructional needs of the students and the future demands of EMI. These frequencies appear in Table 1. Twelve out of 15 of the reading teachers in Level 3 responded to the survey. Their responses revealed their mixed views of the usefulness and appropriateness of the IELTS reading test and its washback. While 50 % endorsed a statement regarding the alignment of the Level 3 curriculum and the IELTS reading exam, more than half disagreed with a statement about the relevance of the IELTS reading texts and tasks to the future academic work of their students. A total of 42 % of respondents felt that preparing for the IELTS reading test had positive effects on their students but 25 % disagreed with that statement.

Table 1 Teachers’ global opinions about IELTS reading

	Strongly agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Strongly disagree (%)
The IELTS requirements correspond closely to the stated course objectives of Level 3 with regard to reading	8	42	42	0	8
Reading instruction in Level 3 is heavily influenced by the types of tests used to assess reading	25	75	0	0	0
The IELTS reading exam seems relevant to the types of reading texts and tasks my students will face in their future studies	8	33	0	42	17
Overall, activities to prepare for the IELTS reading exam have a positive effect on my students	0	42	33	25	0

In order to investigate their views of specific aspects of the reading exam, I used a seven-point Likert scale in which the extreme points on the scale were labelled either as “Very Appropriate” or “Very Inappropriate” (see Table 2). Respondents were relatively positive about the topics covered on the IELTS reading section, with 40 % indicating that they were appropriate. They were evenly divided, however, concerning the effect of the time constraints on the exam: 40 % felt they were appropriate and 32 % judged them as inappropriate. With regard to the appropriacy of the linguistic levels of the test, the opinions were more strongly critical of the IELTS. Fifty-eight percent judged them to be inappropriate while only 33 % considered them relatively appropriate.

Exam preparation is such a major aspect of the Level 3 course that it is essential that these preparation activities have their own language learning benefits beyond mere testwiseness. To examine teachers’ views about this, I asked respondents to compare the instructional usefulness of IELTS activities to conventional Level 3 test preparation activities (see Table 3). The teachers showed a clear preference for

Table 2 Teachers’ views of specific aspects of the IELTS reading exam

	Very appropriate	⋈	<	Neutral	>	⋈	Very inappropriate	Weighted averages
Weighting	+3	+2	+1	0	-1	-2	-3	
Topics covered	0 %	17 %	33 %	33 %	8 %	8 %	0 %	0.43
Time constraints	0 %	17 %	33 %	8 %	25 %	17 %	0 %	0.08
Linguistic levels	0 %	25 %	8 %	8 %	33 %	25 %	0 %	-0.25

Table 3 Teachers' comparisons of the usefulness of IELTS and Level 3 reading exams

	IELTS is far more useful	≪	<	Equal	>	≫	Level 3 is far more useful	Weighted averages
Weighting	+3	+2	+1	0	-1	-2	-3	
Student engagement/motivation	17 %	8 %	17 %	8 %	17 %	25 %	8 %	-0.07
Vocabulary learning	0 %	17 %	0 %	25 %	25 %	17 %	17 %	-0.76
Discourse awareness	0 %	0 %	25 %	8 %	33 %	17 %	17 %	-0.93
Thinking skills	8 %	0 %	8 %	42 %	0 %	25 %	17 %	-0.69
Preparation for the faculties	0 %	9 %	9 %	27 %	18 %	27 %	9 %	-0.72

the usefulness of Level 3 activities except concerning the area of “Student Engagement/Motivation”. The demands of IELTS readings seem to motivate students in ways that are useful even though their instructional value is unclear. Opinions were evenly divided between those who considered the IELTS more engaging and those who considered the Level 3 assessment tools more engaging.

14 Writing Teachers' Views of IELTS

Similar online surveys were completed by writing teachers in the Level 3 program and they revealed a somewhat more positive attitude toward the IELTS writing exam than the reading teachers had toward the reading exam. Nine out of 20 writing teachers responded to the survey and they saw general alignment between Level 3 and IELTS (see Table 4). Almost all the writing teachers (89 %) recognized a direct correspondence between the demands of the IELTS writing test and the Level 3 writing curriculum. Very few writing teachers had negative attitudes toward either the relevance of IELTS writing to university writing or the effect that IELTS writing had on their students. When asked to judge the specific aspects of the IELTS writing exam, teachers were relatively positive about the Expected Grammar Levels and the Expected Vocabulary Levels (see Table 5). Their opinions about the appropriacy of the Topics and the Time Constraints were very mixed, however. On the whole, these opinions were still slightly positive, but a significant number of teachers thought the writing exam was inappropriate in these two respects.

Although the teachers were slightly positive overall about the IELTS writing exam, they did not compare it favorably to the Level 3 writing assessment tasks (see Table 6). When asked about specific aspects of the instructional usefulness of both exams, a majority of respondents judged them to be equal. However, a few teachers clearly favored the Level 3 writing tests, especially with regard to the development of thinking skills and sentence structure. The only aspect of the IELTS that was

Table 4 Teachers' global opinions about the IELTS writing exam

	Strongly agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Strongly disagree (%)
The IELTS requirements correspond closely to the stated course objectives of Level 3 with regard to writing	33	56	0	0	11
Writing instruction in Level 3 is heavily influenced by the types of tests used to assess writing	78	22	0	0	0
The IELTS writing exam seems relevant to the types of writing texts and tasks my students will face in their future studies	22	33	22	0	22
Overall, activities to prepare for the IELTS writing exam have a positive effect on my students	25	50	13	13	0

Table 5 Teachers' views about specific aspects of the IELTS reading exam

	Very appropriate	≪	<	Neutral	>	≫	Very inappropriate	Weighted averages
Weighting	+3	+2	+1	0	-1	-2	-3	
Expected vocabulary levels	11 %	22 %	33 %	22 %	11 %	0 %	0 %	0.99
Expected grammar levels	12 %	25 %	25 %	25 %	13 %	0 %	0 %	0.98
Question formats	11 %	11 %	33 %	33 %	0 %	11 %	0 %	0.66
Topics covered	11 %	11 %	22 %	22 %	33 %	0 %	0 %	0.44
Time constraints	22 %	11 %	11 %	22 %	11 %	22 %	0 %	0.44

clearly more useful for teachers than the Level 3 test was that of “Student Motivation/Engagement”. It appears that the challenging aspects of the IELTS reading and writing tests motivate students in some positive ways.

When asked to comment on the advantages and disadvantages of the writing test, teachers had more negative comments than positive ones. For example, a few teachers questioned how authentic the writing test was in light of the faculty requirements. Several complained that much of their course was spent teaching to the test, and the improvements students made were too superficial. For example, one commented, “It’s more about making a band 4 writer look like they are actually better than genuinely improving their abilities as a writer of English”.

Table 6 Teachers' comparisons of the usefulness of IELTS and Level 3 writing exams

	IELTS is far more useful	≤	<	Equal	>	≥	Level 3 is far more useful	Weighted averages
Weighting	+3	+2	+1	0	-1	-2	-3	
Student engagement/motivation	22 %	11 %	0 %	56 %	11 %	0 %	0 %	0.77
Discourse awareness	0 %	11 %	0 %	67 %	11 %	0 %	11 %	-0.22
Preparation for the faculties	0 %	0 %	0 %	89 %	0 %	0 %	11 %	-0.33
Sentence structure	0 %	0 %	0 %	78 %	11 %	0 %	11 %	-0.44
Thinking skills	0 %	11 %	0 %	56 %	11 %	11 %	11 %	-0.44

In many respects, these findings were similar to those of Lewthwaite's study of IELTS writing washback which was also conducted in the FP at UAEU (Lewthwaite, 2007). He found very positive impressions of IELTS washback from both teachers and students in the English FP. The weakest area of congruence he found was similar to the one identified here: the relevance and usefulness of Task 2 for academic language learning and preparation for the faculties. In his study, many students were neutral on this issue and only 6 out of 16 teachers strongly endorsed his statements to that effect. His data were collected when the IELTS had been in place for just 3 years; since then, based on the present findings, the Level 3 course seems to have yielded to more test-driven pressure.

15 Conclusion

Thus far, there is little conclusive evidence that adopting the IELTS testing system has actually raised levels of achievement in the FP or UAEU in general. The IELTS exam appears to motivate students to study more and try harder in comparison to the traditional FP tests because of its unusual level of difficulty. Nevertheless, it seems quite unsuitable for normal language development activities given the nature of its advanced reading texts and its difficult time constraints. The IELTS writing exam may have fewer negative aspects than the reading exam, but its positive features are not clear in comparison to locally developed tests. The most serious problem with the writing test, in my opinion, is the uneven quality of the prompts for the UAE contexts. There is no doubt in my mind that a locally-produced writing exam could emphasize tasks and questions that are specifically tailored to the needs of UAE students and include relevant rhetorical forms, appropriate types of scaffolding, and suitable background knowledge requirements. Helping students prepare for crucial reading and writing tests should be good opportunities for

ongoing language development and academic study, not just ways to prove what a student has already achieved.

15.1 Pedagogical Implications and Future Trends: An Agenda for Assessment

Even though test-driven schooling is the norm in the Gulf region, there is no reason why good tests cannot accompany high-quality teaching and learning. There are several excellent tests available for students at the Band 5 level that could be used for UAEU students in ways that would probably be more valid and reliable than the IELTS exam. UAEU and UAE secondary school systems should be more realistic about their assessment needs and consider an exam system like the Cambridge First Certificate in English (FCE) if they want an external program. The FCE targets the CEFR B2 Level that corresponds to IELTS Band 5 through Band 6.5. Its linguistic content appears far more accessible than the IELTS, so teachers will be able to combine useful language practice with test preparation in ways that are rarely possible with an IELTS reading exam. Low-level students have a fundamental need for large amounts of relatively accessible input, which Nation (2007) calls “meaning-focused input”. This need is much greater than that of devising coping and guessing strategies for material which is far beyond the testees’ levels. Although the UAE has a large pool of expertise in language testing, they may prefer to coordinate with international testing organizations when creating a national test for university admissions. Alternatively, the UAE can look towards standardized, locally-produced English tests such as are being developed in nations like Italy and Mexico, where international experts are collaborating with local institutions in order to ‘co-certify’ tests and ensure their quality (Newbold, 2012).

Although external exams bring a certain amount of prestige, a more preferable approach on many levels would be for the nation to develop its own entrance exams that reflect the distinctive realities of both secondary schools and tertiary educational processes in this unique bilingual setting. The UAE Ministry of Education has been successfully producing and administering a testing system known as the CEPA exam which is a reliable way to place secondary students in FP English programs. Producing a similar customized test to replace IELTS is not a difficult task considering the fact that all the public universities accept Band 5 as the key qualification for EMI. A test that focuses on one or two key proficiency levels is much easier to produce than a test that covers nine levels. Rather than testing all students with time-consuming speaking and writing tests, the UAE may wish to adopt a more economical approach that divides assessment into two or three stages. For example, results from initial tests of listening and reading could be used to identify three groups: those who are ineligible for higher education, those who need further tests (e.g., writing), and those who could be accepted into higher education directly based solely on their superior levels of listening and reading. If institutions

like UAEU began setting minimum scores in specific skill areas based on analyses of EMI tasks, a test would be needed which allows candidates to retake only those sections in which they were unsuccessful. Ideally, such a test would offer results that would specify the components of students' performance. In the case of reading, it would be useful for students to know how they performed on skill areas like reading for main ideas, reading for details, and reading for inferences. The current IELTS system offers none of this flexibility or diagnostic information and it is not addressed to the local circumstances of the UAE. The great expense and time associated with sitting for an exam like the IELTS needs to be reduced so that many students can monitor their progress through high school, the FPs, and beyond.

Australian universities seem to be leading the way when it comes to recognizing the limitations of entrance test results when addressing the continuous language development needs of non-English speaking students. For example, Dunworth (2010) claims that universities put far too much faith in results like IELTS scores and neglect the realities of academic discourse. She calls for an "institutional process to link the measures that universities accept (on entrance exams) to the lived experience of the tertiary classroom" (Dunworth, 2010, p. 6). To some extent, institutions like the Higher Colleges of Technology in the UAE have done this by integrating language development into normal college courses and using the IELTS at the end (rather than the beginning) of the undergraduate program. Other universities should follow similar approaches since ongoing language development is vital to students' success. Ironically, very few of the graduates of my own program can meet Abu Dhabi's new requirement of IELTS Band 6.5 for government primary teachers because our curriculum does not specifically address English training. Instructional solutions are urgently required in programs like my own, but appropriate and flexible assessment systems will be needed to accompany and support them.

References

- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice*. Oxford, UK: Oxford University Press.
- Belhiah, H., & Elhami, M. (2014). English as a medium of instruction in the Gulf: When students and teachers speak. *Language Policy, 14*(1), 3–23.
- Davies, A. (2011). Kane, validity and soundness. *Language Testing, 29*(1), 37–42.
- Doiz, A., Lasagabaster, D., & Sierra, J. M. (Eds.). (2013). *English-medium instruction at universities: Global challenges*. Clevedon, England: Multilingual Matters.
- Dunworth, K. (2010). Clothing the emperor: Addressing the issue of English language proficiency in Australian universities. *The Australian Universities' Review, 52*(2), 5–10.
- Freimuth, H. (2014). *Cultural bias on the IELTS examination* (Ph.D. Thesis). Rhodes University, Grahamstown, South Africa.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London, NY: Routledge.
- Gebriel, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing, 26*(4), 507–531.
- Gitsaki, C., Robby, M., & Bourini, A. (2014). Preparing Emirati students to meet the English language requirements for higher education: A pilot study. *Education, Business and Society. Contemporary Middle Eastern Issues, 7*(2/3), 167–184.

- Green, A. (2006). Watching for washback: Observing the influence of the international English language testing system academic writing test in the classroom. *Language Assessment Quarterly*, 3, 333–368.
- Gu, L. (2014). At the interface between language testing and second language acquisition: Language ability and context of learning. *Language Testing*, 31(1), 111–133.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17–27.
- IELTS. (n.d.). *How is IELTS marked?* Retrieved from http://www.ielts.org/institutions/test_format_and_results/how_is_ielts_marked.aspx.
- IELTS. (2012). *Test taker performance*. Retrieved from <http://www.ielts.org/researchers/analysis-of-test-data/test-taker-performance-2012.aspx>.
- IELTS. (2013). *IELTS guide for educational institutions, governments, professional bodies and commercial organisations*. Retrieved from http://www.ielts.org/PDF/Guide_Edu-%20Inst_Gov_2013.pdf.
- Lewthwaite, M. (2007). Teacher and student attitudes to IELTS writing tasks: Positive or negative washback? *UGRU Journal*, 5, 1–16.
- McLaren, P. (2011). *English medium in the United Arab Emirates: Serving local or global needs?* Unpublished doctoral dissertation, (Ed.D.). University of Exeter, Exeter, UK.
- McLean, M., Murdoch-Eaton, D., & Shaban, S. (2013). Poor English language proficiency hinders generic skills development: A qualitative study of the perspectives of first-year medical students. *Journal of Further and Higher Education*, 37, 462–481.
- Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes*, 4, 43–66.
- Morrow, C. (2005). *An analysis of the course sequence in UGRU English* (Unpublished Report). Al Ain, UAE: Undergraduate General Requirements Unit, United Arab Emirates University.
- Moussly, R. (2012, May 27). Majority of Emirati students are unprepared for university. *Gulf News* (United Arab Emirates). Retrieved from <http://gulfnews.com/news/uae/education/majority-not-prepared-for-university-1.690616>.
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1, 1–12.
- Newbold, D. (2012). Global institution, local examination. In D. Tsagari & I. Csépes (Eds.), *Language testing and evaluation, Volume 26: Collaboration in language testing and assessment* (pp. 127–142). Frankfurt am Main: Peter Lang AG.
- O’Connell, S. (2007). *Focus on IELTS foundation*. London: Pearson Longman.
- O’Sullivan, A. (2004, September). Reading and Arab college students: Issues in the United Arab Emirates’ Higher Colleges of Technology. In *Proceedings of the First International Online Conference on Second and Foreign Language Teaching and Research: Beyond Borders*. Symposium conducted at First International Online Conference on Second and Foreign Language Teaching and Research.
- Pilgreen, J. (2007). Teaching the language of school to secondary English learners. In J. Lewis & G. Moorman (Eds.), *Adolescent literacy instruction: Policies and promising practices* (pp. 238–262). Newark, DE: International Reading Association.
- Rogier, D. (2012). The effects of English-medium instruction on language proficiency of students enrolled in higher education in the UAE (Ed.D.). Exeter, UK. Retrieved from <https://ore.exeter.ac.uk/repository/handle/10036/4482>.
- Troudi, S. (2009). The Effects of English as a medium of instruction on Arabic as a language of science and academia. In P. Wachob (Ed.), *Power in the EFL classroom: Critical pedagogy in the Middle East*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Weir, C. (2005). *Language testing and validation*. Houndmills, Basingstoke: Palgrave Macmillan.
- Weir, C., Hawkey, R., Green, A., & Devi, S. (2012). The cognitive processes underlying the academic reading construct as measured by IELTS. In L. Taylor & C. Weir (Eds.), *IELTS collected papers 2: Research in reading and listening assessment* (Vol. 2, pp. 157–189). New York: Cambridge University Press.

The Development, Validation and Use of a Test of Word Recognition for English Learners

David Coulson and Paul Meara

Abstract Word recognition is a basic aspect of vocabulary skill, and a critical skill in fluent reading. Native speakers of English can recognize single words in about one tenth of a second. Learners are somewhat slower, but this difference is difficult to measure without sensitive equipment. This chapter describes how we developed a test of word recognition for EFL learners, called Q_Lex. In our approach, words are hidden in nonsense letter strings and this slows recognition speed to a level that personal computers can easily measure. Learners are assessed on the basis of native speakers' reaction time norms. We describe the development and validation of this tool and the measurement principles that underlie it. Especially, we emphasize how we sought to improve its reliability. Finally, we describe an experiment with Q_Lex to investigate learners at different levels of proficiency.

Keywords Vocabulary · Word-recognition assessment · Reaction-time

1 Introduction to Word Recognition in Second Language Learning

Rapid word recognition skill is essential to reading. Hulstijn commented, “Learning to apply reading strategies should not take precedence over establishing a core of automatically accessible lexical items” (2001, p. 266). Yet, reliable, practical assessment for this skill is a major challenge. The relationships that develop between vocabulary sub-skills, such as word recognition and performance are dynamic and unpredictable. If performance in second languages developed in direct proportion to the effort spent memorizing words, research would be much less challenging and much less interesting. Instead, much time must be spent developing

D. Coulson (✉)
Ritsumeikan University, Kyoto, Japan
e-mail: coulson@fc.ritsumei.ac.jp

P. Meara
Swansea University, Swansea, UK

and validating new tests. This chapter will argue that the simpler tests are, the less likely they are to cause trouble. This is because confounding factors cast doubt on the results of even apparently straightforward tests.

The basic findings in L2 word recognition date back to some very early work by Cattell (1886, 1945). In his 1886 paper, Cattell reported a detailed study of two non-native speakers of German. His key finding was that the time to recognize single letters was only slightly shorter than the time needed to recognize whole words. From this, he inferred that individual letters are not perceived in word recognition. This became known as the “word superiority” effect. He also reported that word-recognition in an L2 (German) is slower than word recognition in an L1 (English), and that for single words, the difference is in the region of 10 ms. This finding has turned out to be surprisingly robust. More recent research has not significantly improved on this work, despite the advanced technologies that are available to modern researchers. Cattell’s work relied on an astonishingly innovative use of clockwork and electrical circuitry, which could measure reaction times accurately to about 2 ms.

Cattell’s second study (posthumously reported in 1947) used only a stopwatch to measure reaction times. In this study, he recorded reaction times for hundreds of words. He reasoned that since reaction times are not consistent, the accurate measurement of a few instances does not necessarily provide reliable information which can be generalised. With this simpler method, he found that the speed with which subjects could read words in sentences depended on how well they knew the language. Cattell’s observation that L2 speakers’ reaction times to words are slower than L1 speakers’ and that they could be used to track the degree of ability in the L2 were prescient, and remain highly relevant today.

2 Assessment of Second Language Word Recognition

Although a number of people have written about the need for a practical test of word recognition ability in an L2, there has not really been much progress in this area (see, for example, Daller, Milton, & Treffers-Daller, 2007; Milton & Fitzpatrick, 2014). Daller et al. (2007) accord fluency a central role in lexical processing as one of three components that describe a three-dimensional “lexical space” that is defined as learners’ **breadth** of lexical knowledge, the **depth** of this knowledge and the **fluency**, or ability to access the vocabulary appropriately. However, they state, “It would probably be true to say that we have no widely used or generally accepted test of vocabulary fluency. More research in this area is needed (p. 9).” A standardised test of word recognition ability is an obvious candidate here.

The main problem that faces researchers in this area is that, even with modern technology, it is not easy to measure accurately the very small differences that we expect to find when we compare native speakers and learners on a word recognition task. The standard approach in the very extensive literature on laboratory studies of

L2 word recognition (e.g. Akamatsu, 2008) is the **lexical decision task**, where subjects are presented with a string of letters, and are asked to push a button to indicate whether the string is a word or not. Sometimes, the stimulus strings are preceded by a **prime**, a briefly displayed letter string which can subtly affect the way the main stimulus word is read. For example, a prime like FRUIT makes it easier for people to decide that APPLE is a word. Laboratory studies typically manipulate prime types and stimulus words to show that L2 primes are less effective than L1 primes, and that L2 stimulus words are more difficult to process than L1 stimulus words. Work of this type typically relies on very large and tightly controlled stimulus sets, and this makes it difficult to use the methodology with low-level learners. The method also relies on specialist computer software, such as DMDX (Forster & Forster, 2003), or e-prime (Schneider, Essman, & Zuccolotto, 2002) but this software is expensive and can be difficult to work with outside the confines of a well-equipped laboratory, or well-controlled testing situation.

There have been only a few attempts to develop practical word recognition tests which might be usable outside the laboratory, and they have not been very successful. Below, we will review some of the most important of these.

Laufer and Nation (2001) investigated the relationship between vocabulary size, word frequency and fluency (of word recognition). Their aim was to measure the speed with which subjects match target words at various frequency levels with their meanings. To do this, they made the Vocabulary Recognition Speed Test (VORST), a computerized version of the standard vocabulary size test, Nation's Vocabulary Levels Test (VLT). In the VLT, each item consists of three target words and a set of six definitions (which may be single words), and subjects are required to indicate which of the six definitions best fit the three target words. In VORST, the items are split up so that only a single target word is presented alongside the six other words or phrases.

A sample item from VORST (Laufer & Nation, 2001):

1. apply
2. elect
3. jump
4. manufacture
5. melt
6. threaten

The software records the time from the appearance of an item until a choice is made. Two more items are subsequently displayed with the same block of six words as choices. Bizarrely, once the three selections are completed, test takers are given the chance to amend their answers. If a new choice is made, the new response time replaces the first latency. The mean response time for each block of six words and the mean response time in each word frequency level are recorded. Laufer and Nation (2001) claim that subjects with larger vocabulary size generated faster reaction times. Further, they claimed many words have to be acquired before some of this vocabulary becomes available for automatic recognition.

Many aspects of this test were unsatisfactory. The VLT is a complex tool that was originally created specifically to assess vocabulary size, so re-deploying it for assessing lexical fluency is very questionable (e.g. Kamimoto, 2004). In addition, the length of the six definitions is not consistent—low-frequency target words generally require longer definitions, so the central issue of word recognition speed is compromised. The decision to allow subjects to amend their answers at the end of each block completely invalidates the test as measure of lexical accessibility. Overall, VORST fails to meet Stanovich's requirements for a "clean test" (1982, p. 487). That is, in the business of measuring very specific psycholinguistic phenomena such as the time needed for word recognition, it is important to minimize the number and effect of complicating factors.

A more promising approach seems to be Harrington (2006). Searching for a relationship between size of vocabulary and speed of Word Recognition, Harrington devised a task which closely resembles the classical lexical decision task. Target items consisted of real words and pseudo-words, with subjects being asked to indicate if they know the meaning of the items or not. Unlike the standardized Yes/No test on which this study is based (Meara, 2005) subjects' reaction times were recorded. Harrington found that as the frequency of presented words decreased, accuracy decreased and reaction times increased. Harrington also calculated each subject's Co-efficient of Variation (CVRT). Segalowitz et al. have shown (e.g. 1998) that CVRT gets smaller as the processing of lexical items becomes more automatic, and this would lead us to expect that more proficient subjects would exhibit smaller CVRT on the stimulus set, a finding that is partially supported by the data. This is a superficially appealing approach to assessing lexical fluency, but we feel some caution is necessary. CVRT depends on very accurate time measurements, and it is not clear to us whether it is appropriate to use it with the Classical Yes/No task. This task is rather more complex for L2 learners than it is for L1 speakers, and it is not clear to us how CVRT measures will be affected by this. This is a concern shared by Eyckmans, Van de Velde, van Hout, and Boers (2007) who rejected the use of a computerized version of the Yes/No test for fear that a time constraint could lead to biased responses.

An approach which avoids many of the problems mentioned above is to be found in Prinzmetal and Silvers (1994), who developed a low-tech approach to word-recognition, which does not rely on advanced technology. In one of their studies, subjects were presented with a set of three words, a stimulus word and two other items; one identical to the stimulus word, the other differing from it by a single letter. The subjects' task was to read aloud the word they thought matched the stimulus. The difficulty of this task could be varied by making the stimulus word difficult to read, for example by showing it in a small font and/or in low contrast, as shown below.

went want want

The subjects were judged on the number of items they correctly identified. An advantage of words over non-words was found. This shows that useful data about word recognition skill can be obtained without recourse to sensitive testing equipment. As a result, these researchers moved away from the measurement of

recognition latencies and became concerned with assessment of word-recognition performance. As far as we know, this approach has not been tried out with L2 learners, but it seems to us to have some promise as a low-tech assessment tool.

Our own work has also taken a low-tech approach to measuring word recognition. Meara (1986) created a test methodology which could easily be implemented on the rather limited home computers that were available at the time. This test presented words hidden in a string of twenty letters such as:

weolsulusimpletggihha.

The test takers' task was to find the embedded word (here: **simple**) as quickly as possible, and the time taken to achieve this was measured. There are some critical features to this methodology. Firstly, the hidden words are very quickly identifiable by people who know them well, so English native speakers have little trouble finding the concealed words, and typically there is little variation in the time taken to do this. Meara reasoned that this lack of variation in the native speaker data might form the basis of a standard to assess learners by. Secondly, this approach generates much slower reaction times than we find with standard lexical decision tasks, and this makes it possible to deliver the task on ordinary computing equipment that can be used in a classroom. Thirdly, the method seems to exaggerate the differences between native speakers and learners, and this makes it much easier to identify nonnative speaker-like performance in learners. Finally, the available technology made it possible to monitor the performance of non-standard learners of an L2. In a significant departure for the time, Meara contacted a large number of L2 learners who were following a BBC TV course in Spanish (*Dígame*), and sent them some specially designed computer programs on cassette tapes which allowed them to do the necessary tests at home in their own time. Meara's results showed that, generally, recognition of Spanish words hidden in letter strings became faster as the learners progressed. However, reaction times did not speed up gradually with exposure: rather words seemed to shift suddenly from a pattern of slow reaction times to a pattern of faster ones. This shift was not seen for all words, and this hinted at a dynamic mechanism in which various outcomes are possible, including delayed progress and even loss of access for some words.

Meara's initial work was not taken up at the time, probably because the technology developed rather rapidly at this time, and quickly made his delivery mechanism obsolete. Nevertheless, we believe that the general approach still has much to recommend it. The idea resurfaced in the 1990s, when Meara worked on a revision of his original work which became known as *Q_Lex*—one of a series of tests that Meara developed for the EU's *Lingua* programme. The work reported in the next section is a further investigation of the *Q_Lex* approach.

Direct measurement of word-recognition speed remains a specialist endeavour. As seen, research tends to focus on elaborate research designs, although simple measures may reveal equally rich patterns of lexical skill, as shown by Cattell with his stopwatch and Prinzmetal and Silvers with their large and small words. Complex design inevitably has an impact on the validity and reliability of lexical tests.

3 Researching a Test of Recognition of Embedded Words

Q_Lex is one of several tests devised by Meara. The principle behind these tests is to provide quick and easy evaluation of lexical performance. The tests were designed to be simple enough to deploy in ordinary learning situations, and robust enough to give a reliable snapshot of learners' ability. The tests were generally short enough to be completed in a few minutes, but in spite of their brevity, they usually tested significantly more words than other readily available tests did. Typically, they involved high frequency words as stimuli, on the grounds that this allowed the same tests to be used with learners at different levels of proficiency. Typically, too, the tests were designed so that the performance of learners could be meaningfully compared with the performance of native speakers. The version of Q_Lex described in this section is Q_Lex v3.0. It was written in Delphi 4, and was designed to run as a stand-alone program on the Windows platform.

4 Test Design

In Q_Lex, 50 high-frequency 6-letter words are hidden in a 15-letter string as shown below:

pajlchanceacdut

These items are shorter than those used in the Dígame project, mainly because high frequency English words are shorter than their Spanish counterparts. Items are displayed on a personal computer. They appear in 20-point Arial bold font. The task for the test taker is to identify the hidden word in each item as quickly as possible. A timer starts with each presentation, and learners click a mouse to stop the timer and record their time when they have identified the hidden word. Once they do this, the program displays a set of four additional words: one word is the target word, while the other three are words which are similar to the target word. Test takers have to identify which of the four words they had seen hidden in the 15 letter display. This additional check is used to confirm that they had actually correctly identified the target words. Test takers' performance is judged against a native-reader standard as described below.

5 Masking String Design

A significant issue with this methodology is the construction of the longer strings in which the target words are hidden. These masking strings are not just randomly selected letters. Rather, a procedure is followed which allows the difficulty of items to be controlled. The masking strings used in Q_Lex 3.0 are "1st-order approximations to English", which reflect the frequency of different letters in English, and

Table 1 A comparison of items presented in strings of varying difficulty

Target	Zero-OA	1st-OA	2nd-OA
Leave	fiwleavemtsnt	lyleavekicbof	retleaveicter
Night	zqpwnightuemp	slenightrabyg	dirsnightunwi
Large	tsyjhlargegql	heclargenyiti	medbilargefou

their probability of being preceded and followed by another letter. The methodology for constructing these strings is based on work by Miller (1963) and is described in more detail in Appendix 1.

Some examples of target words hidden in zero-order approximations to English, 1st-order approximation to English and 2nd-order approximations to English are provided in Table 1. In general, hidden words become harder to identify as the masking strings become more English like. However, some care needs to be taken with higher-order approximations because these masking strings will often include short sequences of letters which make up real words that are not the intended target word. Items of this sort were excluded from the stimulus items used in this study.

6 The Use of Native Speaker Recognition Times to Create Norms

50 6-letter words from the JACET 8000 list (Aizawa, Ishikawa, & Murata 2005) were selected. They had a mean rank order of 1484. They were embedded in 1st-order approximation strings. Native-English speakers took the test, and then data was gathered to make the norms. 18 university graduates participated. Each of the fifty test items were displayed on a personal computer screen. Subjects clicked on a button to display each item in turn, and clicked the same button again as soon as they saw the hidden word. This action stopped the timer, and displayed the multiple choice screen that allowed subjects to confirm their answers. For **pa-jlchanceacdut** the multiple choice options were: **chance chalet change and chapel**. The three distractors were selected for their orthographic familiarity to the test item. The aim was to prevent test takers from selecting an answer based on memory of a few letters from the string, and to promote searching for the whole word.

12 of 18 subjects recognized all 50 items correctly with the others missing between one and three. The mean reaction time was 22.8 s with a standard deviation of 7.42, indicating that some subjects varied quite a lot from the mean. There was no evidence of acceleration during the test. In other words, the test was sufficiently simple for initial reaction times to items to be similar to those towards the end of the test. The subjects took the test again 6 months later and most recorded very similar recognition speed. The correlation co-efficient between the two tests was 0.85.

For each item, a norm score was created. These norms were calculated as follows: for each target word, the mean reaction time of the 18 native speakers was added to twice the standard deviation of these reaction times. 95 % of the native speakers' scores are faster than this norm score.

7 English Learners' Scores on Q_Lex

7.1 *Preliminary Pilot Studies*

50 first-year university students majoring in English took part in a pre- and post-test design, with a gap of 2 weeks. Each test started with five practice items, and then the fifty normed items described above were presented. Test takers scored one point for each item that they recognised correctly within the native-speaker norm, giving a top score of 50 points.

The purpose of the 2-week gap was to assess if Q_Lex could show consistent scores in learners who should still have the same level of proficiency as in the first administration. The results showed a very marked increase in scores over this short period. The average score in test 1 was 14.6 points and in test 2 it was 26.4 points. In addition, the test showed a significant practice effect, in that target words at the end of the test were recognized significantly faster than words at the beginning of the test. We had not anticipated either of these effects.

In a second pilot study, the target words were embedded in 2nd-order approximation strings which makes them slightly more difficult to detect. The prediction was that better concealed words would be visible to learners who know them better. Contrary to expectation, this led to a worse outcome in terms of reliability across the two tests. However, other aspects of the results were noteworthy. We compared the scores of groups of new first-year students embarking on an English degree, and other students entering other majors, who would not be studying English. In Japan, all students have to study English until the end of high school, and all of our students had to pass the university entrance examination which assessed English ability. At first, the scores of these two groups were comparable. A year later, the average scores of the English major group had increased significantly, while those of a non-English major group had significantly fallen. This suggested that Q_Lex might be sensitive enough to measure changes in both lexical acquisition and attrition.

Our concern was that while Q_Lex seemed to be good at detecting large shifts in the performance of groups, it did not appear to be good at detecting smaller shifts in the performance of individuals. Some improvement in the assessment of individuals was achieved by conducting a Rasch analysis on the test items, selecting those that performed best and discarding the strings that discriminated badly. This post hoc approach was only partially successful, and did not seem to offer a solution to reliability problems with the current version of the test.

Nevertheless, as a tool for measuring the skill of recognition in groups of learners, other tantalizing results also emerged. One was the relationship of word-recognition skill and scores on the reading section of TOEIC (a common standardized proficiency test in Japan). 73 individuals took TOEIC, and a moderate correlation (0.50) between Q_Lex and reading scores was found in the lower-scoring half of this group. Conversely, the top half had almost a non-existent relationship (0.05) with word-recognition skill. To investigate this further, a technique called the ‘moving window of correlation’ (Verspoor, de Bot & Lowie, 2011) was used to investigate this relationship. This involves incrementally sampling the scores of groups of five students through the group. The result showed a clear, but quite uneven, downward slope from left to right. In other words, whilst the group showed a decreasing correlation, no individual was certain of showing this feature. This is consistent with a dynamic view of second language development, where we do not expect individuals to develop in predictable stages. Rather, development is messy and unpredictable. Nonetheless, research tells us that poorer readers are less skilled at automatic word decoding, and they try to make up for this by relying on global, top-down skills (Grabe, 2009, p. 28). The fact that the correlations with word recognition decrease among stronger readers reminds us that word-recognition ability is not a sufficient condition for effective reading (Koda, 2005). These Q_Lex results appear to map this facet of reading skill. So one-way Q_Lex might be used to identify a *deficit* in word-recognition proficiency, rather than as a test of the presence of this skill in learners.

Overall, it seems that the unpredictable variation in scores for any learner between two tests is not some fatal weakness of the test, but rather a description of an usual feature of learners’ developing skills to be unstable and erratic.

8 Further Attempts at Improving the Reliability of the Test

We felt that the content validity of Q_Lex could be improved so that the initial scores would better reflect the ability of learners. We still had one more option on the strings for item masks: to use zero-order approximation instead of more English-like strings. Our aim was to reduce the amount of variability in scores, due not to the natural course of development, but rather unreliable facets of the test design. We therefore introduced three further changes into the test design:

- (a) A new set of five-letter words was selected. These had higher mean frequency than the words used before. Shorter words are also more likely to be a single syllable in length, and this helps reduce the occurrence of recognition based on word parts. (For example, the word “reduce” cannot be used in Q_Lex, since test takers might recognize “red” and stop the timer on that basis.) These items were placed in shorter 13-letter strings. We felt the combination of shorter words in zero-order approximation strings was likely to promote whole-word

(sight) recognition, and, as a result, improve the content validity of Q_Lex. An example of a new item is: **jjo vxz emptyjh**

- (b) In the earlier versions of the test, students could keep searching for the item for as long as they needed. On occasion, this took over 10 s and far exceeded the recognition time norms. This led to frustration among test takers. In the new version of Q_Lex, as each item is displayed, the software displays a timer which counts down from its norm value (typically about 2 s). When the counter reaches zero, the test screen automatically changes to the answer screen to display four choices. The countdown is displayed on the screen and decreases in steps of 100 ms. This format change was more acceptable to the test takers, and considerably reduced the total amount of time needed to finish the test.
- (c) Finally, the method of measuring test takers' recognition time was improved. In earlier versions, an on-screen start/stop button was clicked to display each item. Now, the test was started by clicking this button, but the reaction time was recorded by pushing the keyboard space bar. This change allowed for much more accurate measurement of reaction time (see Appendix 2).

To investigate this new format, 66 words were selected. They had a mean rank order of 833 in the JACET word list. 20 native speakers took the test, so the norm values necessary to test learners could be calculated. Initially, the time limit on the new counter function was set to 2 s. They answered 91 % of the items correctly. Their mean reaction time was 925 ms ($SD = 358$ ms). This was much faster than on earlier versions of the test. The test appeared very easy and uncomplicated so this reassured us that for learners, this version would be a more valid test of word recognition.

106 female first-year university students took the version with the new norms. Their mean score was 36.0 (54.5 %). The test showed good reliability by the Kr-21 method (0.92). Non-scoring responses were largely the result of the target word being timed out, with a relatively small number of incorrect identifications (an average of 7 items). Thus, it seemed that the use of 5-letter words resulted in shorter, easier strings that provide more reliable results. We felt we had a test that seemed to be a transparent test of learners' word recognition skill.

9 Rasch Analysis and the Creation of Equivalent Forms of Q_Lex

With the new format and items, learners achieved higher scores, but we were still not sure if these would be consistent over two tests. As explained, this is important for assessing genuine change in word-recognition ability. In earlier rounds of investigation, the learners had always taken the same version of the test (albeit, usually months apart). One way to deal with this issue is to create two parallel versions of the same test, and use them in a split-half design. This would reveal the

performance of groups of students on different, but equivalent, tests to the one they took first.

The 66 items were examined with Rasch analysis. The range of infit meansquare was from 0.73 to 1.19. Based on this, six items were removed. The remaining 60 items were split into two 30-item sets. One set had a mean hit rate of 55.1 per item (the number of hits by the 106 subjects mentioned above). It had a mean value of infit meansquare of 1.02. The figures for the other set were 56.7 hits per item and the infit meansquare was 0.98. Further shuffling of the items resulted in **Form A**, with an infit meansquare value of 1.00, and **Form B** with a value of 0.99. The mean number of hits per items was 55.8 in each set. We predicted that the two sets should produce similar scores in initial and second tests.

Two groups of subjects took each set. Group 1 had 47 students and Group 2 44 students. Re-tests were conducted 5 weeks after the initial test session. All the test takers were native speakers of Japanese learning English as a foreign language at university level. Test 1 mean score was 14.7 points for Form A and 15.6 for Form B. There was no significant difference between them. The mean scores of groups on both forms in Test 2 were practically identical (Form A, 18.5 points; Form B, 18.6 points). On inspection, it turned out that for the higher-scoring half of the group, the change in scores across the tests was 0 %. Form B was judged to be slightly more reliable. In a further study, learners took Form B twice with a gap of only 2 weeks. The results showed that there was an increase in scores of about 10 %, a significant improvement on earlier versions. We would have preferred to see a smaller change but we had to satisfy ourselves that some score increase is inevitable due to test habituation.

Overall, results suggested that the test came as close as possible to providing a reliable initial score for many learners. However, the number of items in Version B was only 30, compared to 50 in the original version of Q_Lex. This might be criticized as being less representative than our stated goal. In fact, due to their higher frequency, the 30 words in each set represent a better coverage of the first thousand words of English. In the final part of the chapter, we will report on a longitudinal investigation of learners' lexical development using Form B of Q_Lex.

10 Assessing Learners' Skill with Q_Lex

Our aim has been to develop a standard test for the practical assessment of English learners' word-recognition skill. The key feature of Q_Lex is that the test is operationalized as the number of items recognized within native-speaker norms, rather than actual speed of recognition. As described above, this is a practical solution to the difficulties of exact measurement. We have presented evidence that this test has good concurrent validity. That is, the scores of learners reflect their ability at that moment, and therefore we can gain insights into how this ability changes over time. Since the course that the students followed included an intensive vocabulary learning programme, we also address the relationship between

vocabulary-size growth and accessibility speed of high-frequency vocabulary, hypothesized by Laufer and Nation, and Harrington, as reviewed above. Further, we can investigate whether there is any change in items successfully answered on the first test. That is, to what extent the accessibility of learners' common words shifts over time. This should reveal more about the dynamic nature of vocabulary knowledge. We will report on a group of first-year university learners at discrete proficiency levels.

We had three research questions:

- (a) How does the word-recognition of learners change over time?
- (b) How does word recognition develop in response to vocabulary learning?
- (c) How consistent are learners over time in responding to the same items?

11 Method

42 first-year university students took part (34 females, 8 males). They were studying English for five, 90-min sessions per week in an academic preparation course. They were from three proficiency levels. 15 were in the 'advanced' class, 14 in the 'intermediate' and 13 in the 'basic' class. There was also a control group of students who were studying in the preparation course, but they did not follow the vocabulary course. Their proficiency was similar to the intermediate and advanced groups.

All participants (experimental and control group) took a test of the first five thousand words of English, called X_Lex, (Meara & Milton, 2003) at the start and the end of the investigation. Over 9 months, the experimental group students studied an online vocabulary learning system. Following a test to estimate their vocabulary size, the website selected words to match their estimated level. Students had to spend one hour a week learning vocabulary. All four groups (three experimental, one control) took Q_Lex twice with an intervening period of 30 weeks.

12 Results

Figure 1 shows the change in Q_Lex scores between the two test events. The groups scored at fairly similar levels at the outset. The scores of all the experimental groups increased between test events. An ANOVA showed that at Time 1 there was no significant difference between the groups, and likewise for the results at Time 2. The control group recorded the same score (15.3 points) at the end as at the start.

Figure 2 shows the X_Lex scores. The results show that the Basic group made the greatest gain at this 5 k level. The control group showed a slight fall in vocabulary size. The online vocabulary system reported that the average gain was 1109 words ($SD = 778$).

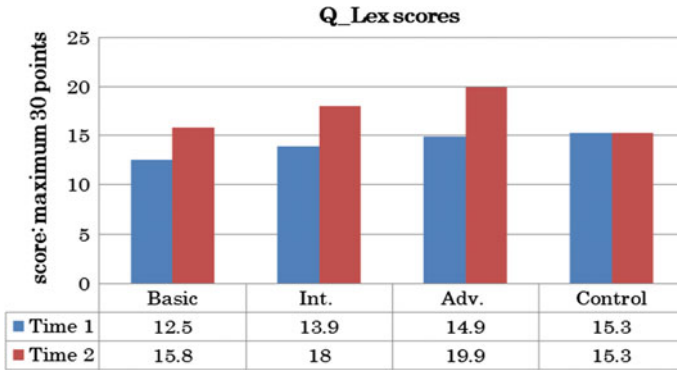


Fig. 1 The pre- and post-scores on Q_Lex

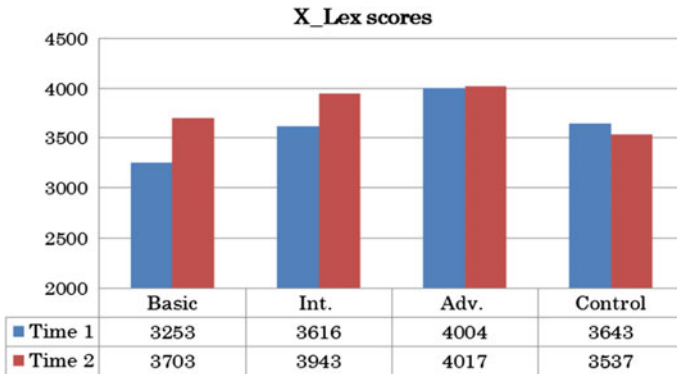


Fig. 2 The pre- and post-scores on X_Lex

Figure 3 shows the change of Q_Lex scores. Scores that were initially low tended to show medium to large gains whereas initially higher scores led to post-test scores that fell in a narrower range. The scores did not cluster by proficiency level.

There was a high level of consistency (68 %, $SD = 16.7$) in correct responses between tests. Further, consistently answered items appeared to have better accessibility. Figure 4 shows the change in reaction times. The left side shows the reaction times for items that were answered correctly in both tests. The right sides shows those answered correctly in only one test. In both cases, reaction times fell significantly (at the 0.05 level). However, items answered both at Time 1 and 2 had initially faster latencies compared to those which were missed later ($t = 4.52$, $p < 0.001$).

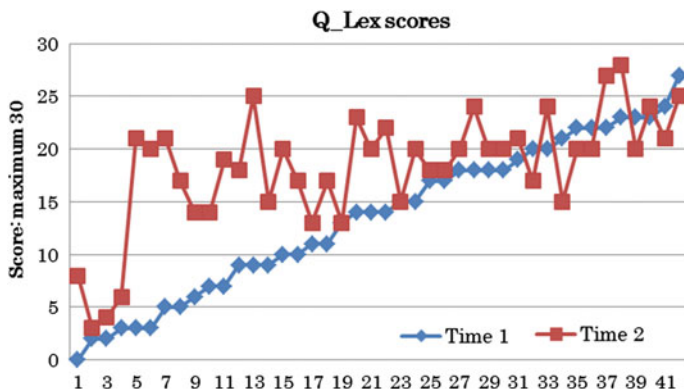


Fig. 3 The change in individual scores on Q_Lex

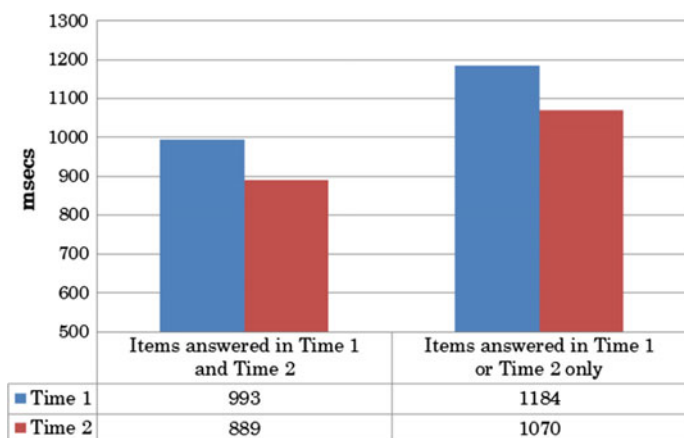


Fig. 4 Consistency in answering times

13 Discussion

The results of our investigation revealed that the word-recognition ability of groups of students who are engaged in a full-time English programme, with an explicit vocabulary learning component, improved their Q_Lex scores over one academic year. Conversely, the group not involved in the word-learning activity (but who followed the same academic preparation course) showed no gain in Q_Lex scores. That is, Q_Lex appears able to reflect changes in groups, based on proficiency and learning-activity differences. What conclusions can therefore be drawn regarding Q_Lex and what it reveals about the state of lexical development of learners? We had three research questions:

- (1) Concerning the change in word-recognition ability over time, all the experimental groups showed significant gains in Q_Lex scores. This test apparently reflects changes in lexical accessibility, which are weakly linked to general proficiency. The Advanced group made the biggest gain (5.0 points) and then the Intermediate group (4.1 points) and then the Basic group (3.3 points).
- (2) The evidence for the effect of vocabulary learning on word-recognition ability is not clear. Gains in vocabulary size were confirmed by the X_Lex results (Fig. 2) in which we can see gains at the 5 K level for the Basic and Intermediate groups. (The lack of progress for the Advanced group probably reflects the fact the system was giving them much lower-frequency vocabulary to learn. In addition, X_Lex tests only the first five thousand words of English, a level which may not have stretched the more advanced subjects.) A weak correlation was seen between the reported gains in vocabulary size on the on-line system over 9 months and gains in Q_Lex scores ($r = 0.25$).

When we look at these results by proficiency, a differentiated pattern appears. The correlation between the number of words learned and Q_Lex scores was 0.64 for the intermediate group and 0.47 for the Basic group. The Advanced group showed a negative correlation (-0.28). This might indicate that the students who learned more frequent vocabulary on the online system (the Basic and Intermediate groups) extended their Q_Lex score, whereas with the Advanced Class who studied much more infrequent vocabulary, there seems to be a negative effect on their Q_Lex scores.

As mentioned, these results do not support the idea that increasingly large vocabulary size leads to better accessibility on the high-frequency vocabulary items of Q_Lex. In particular, this result does not match the claims of Laufer and Nation (2001) that a larger vocabulary leads to greater accessibility. However, these results do match the finding by Miralpeix and Meara (2014) that there is no consistent relationship between vocabulary size and accessibility skill. They also claim the relationship is not random, and that also appears to be reflected in the data. The result lends some support to the idea that accessibility might be an independent dimension of vocabulary knowledge.

- (3) Concerning the question of consistency in answering, a clear-cut pattern emerged. Despite a long intervening period between administrations, all participants managed to answer at least half of the items they had correctly responded to in the first test. This suggests that some words in memory are much easier to access, and this facet of knowledge may not vary much over time. Further, the data from Q_Lex demonstrated that these reliably recognized items had significantly faster response times compared to other items which were recognized in only one or the other test. This also appears to depend on the individual. In other words, this result is not due to the facility of the items, but reflects a greater sensitivity to certain words among learners.

14 Conclusion

This chapter has provided a brief overview of a new testing tool that we think might have a useful role to play in L2 vocabulary research. The test we have described is a low-tech tool that assesses the ability of L2 learners to access words when they are presented out of context. Unusually, the test is standardized against the performance of native speakers on the same items. We think that this test method has a number of features to recommend it to vocabulary researchers, and we hope that it might be used in future to model how individual learners' ability to process basic vocabulary changes as their proficiency develops.

Appendix 1: Approximations to English

Letters randomly selected from the alphabet are known as zero-order approximation strings. Words placed in masks made from such a selection of letters are easy to recognize since the masking string does not resemble English, and the hidden word stands out against this background. To increase difficulty, first-order approximation strings can be used as masking strings. To construct these strings, a letter is chosen at random from a text, and then every *n*th subsequent letter is added to the string. The end result is a masking string that reflects the frequency of English letters. (The letter 'e' appears more often than 'z', for example.) First order approximations have a closer resemblance to English, so words hidden in this kind of masking string are better camouflaged. Second-order approximation strings reflect the distribution of 2-letter pairs in English words—the sequence 'ab' is much more likely to occur in these strings than the sequence 'jj', for example. As a result, these masking strings camouflage the hidden word more effectively still. The three examples below illustrate this effect. The zero-order masking string contains only one vowel, so it is unlike any word spelled in English. Conversely, the first and second order masking strings are increasingly English-like. (Note that in the second order string, the word 'vein' has appeared fortuitously in the masking string. This would need to be removed for content validity.)

Zero order approximation string: gwdfdqtablevwcu

First order approximation string: lusetablechtacvutno

Second order approximation string: einentablerveinem

Appendix 2: The Instructions Provided to Subjects in the Test

In this test, quickly search for hidden 5-letter English words. Look at the example below. Can you see 1 hidden word?

fodguestqborp

HOW TO USE THE TIMER

In this test, you use two fingers to control a timer



1) **START**
left click on the timer button to see the item

2) **SEARCH**
fod**g**uestqborp



3) **STOP**
push the space bar with your other finger and select your answer from a choice of 4 words

4) Repeat 1), 2) and 3) for each item



References

- Akamatsu, N. (2008). The effects of training on automatization of word recognition in English as a foreign language. *Applied Psycholinguistics*, 29, 175–193.
- Aizawa, K., Ishikawa, S., & Murata, M. (2005). *JACET 8000 eitango*. Tokyo: Kiriara Shoten.
- Cattell, J. M. (1886). On the time taken up by cerebral operations. *Mind*, 11, 377–392.
- Cattell, J. M. (1947). The time it takes to recognize and name letters, pictures, and colors. In: A. T. Poffenberger. (Ed.), *James McKeen Cattell: Man of science* (pp. 13–25). Lancaster, PA: The Science Press. (Translated from German; Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *1885 Philosophische Studien*, 2: 635–650.)
- Daller, H., Milton, J., & Treffers-Daller, J. (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press.
- Eyckmans, J., Van de Velde, H., van Hout, R., & Boers, F. (2007). Learners' response behaviour in Yes/No vocabulary tests. In H. Daller, J. Milton, J., & Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 59–76). Cambridge: Cambridge University Press.
- Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behaviour Research Methods, Instruments & Computers*, 35(1), 116–124.
- Grabe, W. (2009). *Reading in a second language. Moving from theory to practice*. Cambridge: Cambridge University Press.
- Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook 1 (2006)* (pp. 147–168). Amsterdam: John Benjamins.
- Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge: Cambridge University Press.
- Kamimoto, T. (2004). *The vocabulary levels test: A relationship between target items and distractors*. Paper Presented at the 2004 Eurosla conference, San Sebastian, Spain.
- Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press.
- Laufer, B., & Nation, P. (2001). *Passive vocabulary size and speed of meaning recognition: Are they related? EUROSLA Yearbook 1(2001)* (pp. 7–28). Amsterdam: John Benjamins.
- Meara, P. M. (1986). The Digame project. In V. J. Cook (Ed.), *Experimental approaches to second language learning* (pp. 101–110). Oxford: Pergamon Institute of English.
- Meara, P. M. (2005). Designing vocabulary tests for English, Spanish and other languages. In C. Butler, M. A. Gómez González, & S. Doval Suárez (Eds.), *The dynamics of language use: Functional and contrastive perspectives* (pp. 271–285). Amsterdam: John Benjamins.
- Meara, P. M., & Milton, J. L. (2003). *X_Lex v 2.05*. Swansea: Lognostics. Retrieved from <http://www.lognostics.co.uk/tools/>.
- Miller, G. A. (1963). *Language and communication*. New York: McGraw-Hill.
- Miralpeix, I., & Meara, P. M. (2014). Knowledge of the written word. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (pp. 30–44). Basingstoke: Palgrave Macmillan.
- Milton, J. & Fitzpatrick, T. (2014). *Dimensions of vocabulary knowledge*. New York: Palgrave Macmillan.
- Prinzmetal, W., & Silvers, B. (1994). The word without the tachistoscope. *Perception and Psychophysics*, 55(3), 296–312.
- Schneider, W., Essman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
- Segalowitz, S. J., Segalowitz, N. S., & Wood, A. G. (1998). Assessing the development of automaticity in second language word recognition. *Applied Psycholinguistics*, 19, 53–67.
- Stanovich, K. E. (1982). Individual differences in the cognitive processes of Reading: 1. Word Decoding. *Journal of Learning Disabilities*, 15, 485–493.
- Verspoor, M. H., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development: Methods and techniques*. Amsterdam: John Benjamins.

Part IV
Alternative Assessment

Alternative Assessment: Growth, Development and Future Directions

Vino Sarah Reardon

Abstract This chapter will attempt to focus on *alternative assessment* from a pedagogical perspective (i.e. in the context of learning environments that address students with and without *special needs*). Thus, the purpose of this chapter is threefold: (1) to investigate how alternative assessment has been used in the United States and elsewhere, (2) to define the general characteristics of alternative assessment (in particular, portfolios), and (3) to examine the direction that alternative assessment has taken in recent years within *SPED* (Special Education) programs, *ESL* (English as a Second Language) and/or *EFL* (English as a Foreign Language) classrooms. This chapter offers a significant contribution to the field of ESL/EFL by demonstrating that alternative assessment best serves the needs of English Language Learners (ELLs) with and without speech impairments.

Keywords Assessment · Alternative · Traditional · SPED · ESL · EFL

1 Introduction

Historically, students who have special needs (i.e. students who are physically or mentally challenged because of physical or cognitive impairments) have been placed in what is commonly referred to as SPED programs in the United States. No doubt, the goal has generally been to offer this *learning disabled* (LD) student population adequate academic support and preparation to *mainstream* these students into general education classrooms, if and when possible.

V.S. Reardon (✉)
Acaciaewood College Preparatory Academy, CA, USA
e-mail: vinoreardon@gmail.com

2 Background

In this context, alternative assessment (or *alternate assessment*) which is characterized by an untimed, free-response or open-ended format (Brown, 2004) has been found to be particularly useful in special education programs because it is better suited to individualized feedback. Since it is *formative* in nature (rather than *summative* as in the case of *traditional* forms of assessment), it is more oriented to *process* (as opposed to being *product-focused*), thus making it easier to assess and help students who may have special needs. Because of the extent to which these special needs students are at risk for academic failure, teachers are expected to assume “new” assessment roles, helping their LD students to *self-assess* themselves (for example, within the context of “multiple intelligences”) and hence engage in tasks that are more meaningful to them (Brown, 1999, 2002; Gardner, 1983).

It would be helpful to address the fact that an alternative or alternate assessment should not be perceived as either a traditional large-scale assessment or a diagnostic assessment that has been individualized. More importantly, it must be understood that alternate assessments can be administered to students who “differ” greatly in their ability (when responding to stimuli and/or providing responses) in the same way that the general student population does.

In the United States, the federal *Individuals with Disabilities Education Act* (IDEA, 2004) (Wright & Wright, 2007) has stipulated that students with disabilities participate in statewide assessments which include an appropriate form of alternative assessment (this type of alternate assessment subscribed to “modifications” earlier but is currently seen to favor “accommodations”), especially in the areas of reading and mathematics but also in English, as in the case of ELLs. Prior to the 2004 reauthorization of *IDEA 2004*, the term *modifications* referred to *changes* in terms of instructional level or delivery of content in relation to district-wide or statewide tests for students receiving special education services. As a result, modifications resulted in scaling down the standards by which these students were evaluated. Modifications indicated that curriculum and/or delivery of content had been altered to a large extent. It should be understood that when modifications were made, LD students were not expected to master the same academic content as others in a regular classroom setting. In such a situation, grades did not seem to reflect the true abilities of these LD students. For example, a student who was incapable of performing well in the content area of “complex fractions” in a math class might only have been capable of working in the area of “additions” in a very simplified format.

What then was the underlying implication here? It simply indicated that the student’s *instructional level* had changed considerably. Therefore, when one looked at the *grade* of the student, it became important to determine whether the student had received this grade in the standard curriculum for his grade level or in a modified curriculum. Consequently, the term “modifications” emphasized the *simplified* versions of these tests. This led to a rather complex situation.

Thus, beginning with *IDEA 2004*, the term “modification” has no longer been used in relation to district-wide and statewide testing because the federal *No Child Left Behind Act* (NCLB) mandates that students with learning disabilities be tested using the same standards as those without the same learning disabilities. According to some of the final regulations in the NCLB Act of 2007, students with disabilities are expected to have access to tests that allow them to demonstrate their competence on those tests (Elliot, 2010). Thus, to be included in an educational “accountability system” (i.e. by generating information about how a school or district or state is doing in terms of overall student performance), a student who is physically or mentally challenged would need to have access to a mechanism which provides *alternate* ways of assessing him/her. This can usually be achieved by administering the test in a different setting and changing the presentation of test items or response format.

Thus, in a research study conducted by Elliot (2010), the findings indicated that students with disabilities performed far better when fewer distractors were used or when more white space was added to the page. Based on these findings, the researchers were able to conclude that modified tests were helpful to students with disabilities as long as those who designed the modified tests were trained in creating tests that could better reflect learning outcomes. This study contributed to the understanding that students who qualified for AA-AAS (Alternate Assessment based on Alternate Achievement Standards) did *not* perform as well on multiple-choice tests while students who qualified for the AA-MAS (Alternative Assessment based on Modified Achievement Standards) demonstrated that they would need more time to make the required progress.

A study by Kearns, Towles-Reeves, Kleinert, Kleinert, and Thomas (2011) indicated that students who required augmentative and alternative communication devices needed additional time to acquire and maintain the skills; that is, when the subjects were severely impaired, only 50 % of that population could access an alternate assessment. Of equal and related significance, the objective of the study conducted by Kettler (2011) attempted to find a reliable screening method that would identify the small percentage of students that would qualify for AA-MAS. The Computer-Based Alternate Assessment Screening (C-BAAS) test was deemed appropriate to screen students who needed a modified assessment. It became increasingly clear that the results of this test could be used to help teachers and *Individualized Education Program* (IEP) teams offer the instructional support that some students (i.e. those who could *not* meet the required proficiency level within one year in the classroom) needed desperately (Kettler, 2011).

In the United States, an IEP team is expected to determine whether a student is eligible to participate in an alternate assessment (Musson, Thomas, Towles-Reeves, & Kearns, 2010). It is true that if identified inappropriately, students who qualify for the program may fall through the cracks; conversely, a student who qualifies may not become eligible to use the service when misdiagnosed.

However, when a student is identified appropriately, an IEP team becomes responsible for considering the SPED student’s unique needs in designing assessment *accommodations*. What are accommodations? Accommodations are best

envisioned as adjustments that are made to ensure that the learning disabled students have equal access to the curriculum so that they can have successful learning outcomes. Therefore, a student with LD can learn the same material but in a different way; in other words, a student who is a slow reader may have the ability to listen to the audio version of the same book (and be exposed to the same content) to allow them to better demonstrate the learning outcomes.

In a related situation, what is of importance is also how a disability intersects, for example, with second language acquisition as in the case of ELLs. It should not be assumed that accommodations designed for typical ELLs (without disabilities) would necessarily give adequate support to ELLs with disabilities (cf., Reardon & Nagaswami, 1997). The issue of disability and how it is affected by language proficiency (or the lack thereof) will be revisited later in this chapter.

In such a situation, the IEP has to be structured in a more specific way so that it can indicate accurately the knowledge and skills of the special needs ELL population. It must be clearly understood that the primary purpose of using alternative assessments is to evaluate students who are not able to participate in general state assessments even when they are provided *accommodations*, for example, by changing the seating arrangement of a student. It is generally the case that through the IEP, accommodations can be developed both formally and informally. In the United States, the 50 states vary a great deal in the specific accommodations that they allow; all 50 states have written guidelines that indicate allowable assessment accommodations for all students with disabilities, including ELLs. What should be remembered is that for any assessment to successfully measure a student's learning outcomes, the assessment needs to be both *valid* and *reliable*. The issues of *validity* and *reliability* will be revisited in a later section of this chapter.

No doubt, accommodations can offer the possibility for the LD student to use "assistive technology" (e.g. a tape recorder or a computer) to complete his/her project on "air pollution" successfully, for example, by not having to struggle with pencil and paper. Moreover, the student's IEP can outline these accommodations clearly, if necessary. For instance, what about those students who are identified as "deaf"? In a study conducted by Cawthon (2011), the researchers concluded that the communication *mode* seemed to affect assessment outcomes, so it was important to use the same mode that the students had been familiarized with in classroom instruction. In such a context, especially when attempting to establish which students needed to take the AA-MAS, a posttest questionnaire seemed to shed light on how items should be presented in a test (Roach, Beddow, Kurz, Kettler, & Elliott, 2010) to make alternative assessment more viable in terms of producing desirable learning outcomes. Thus, investigating if variables such as "visuals" or "bolded items" or "fewer distractors" made a difference in test outcomes became the focus of the study. The results of the test showed that some of these differences proved to be useful to those students who *perceived* them to be helpful.

Thus, this led to the premise that a student's individual learning style was critical in determining how instruction would need to be tailored. It should be underscored that students with disabilities vary greatly in terms of *how* they learn even when they are exposed to the same content. The inherent assumption here is that content

should be delivered differently to suit the needs of the individual learner in order to have appropriate learning outcomes. This leads us to the following question: what pedagogical implications does this have for classroom instruction?

3 Implications for Classroom Instruction

Evidence indicates that instructional planning would influence how delivery is sustained within the broader scope of implementing and evaluating the IEPs of students who have disabilities (Johnson, 2012). Since school districts are accountable for the learning outcomes of *all* students, those that have disabilities are not exempted from this group of students that get monitored for growth and progress on a yearly basis. What this implies is that test scores would need to be made more *valid* within the overall context of using modifications and/or accommodations. In a study conducted by Hager and Slocum (2011), the researchers examined various ways (by conforming to federal guidelines) through which alternative assessment could be used so that students could have successful learning outcomes.

Thus, at this juncture, it would also be helpful to emphasize that alternate assessment has been used with students whose learning disabilities are not too severe and who have responded more favorably to *general assessment tools* (i.e. the general assessment instruments that are used in regular classroom situations when these students are placed in “pull-out” K-12 programs). It should be mentioned that these students are *mainstreamed* as and when it is deemed appropriate.

In addition, it is important to focus on the fact that alternate assessment has been used in the United States even in “general education” classrooms (i.e. in regular classrooms as opposed to “special education” classrooms) as a way to extend day-to-day classroom activities and to promote problem-solving skills. It should be noted that the notion of *alternative assessment* has become more popular, especially in ESL and/or EFL classrooms, since the early 1990s as a result of teachers realizing that not all students and not all skills can be measured by traditional tests. (Brown, 2004). Alternate assessment is often seen as a continuous long-term assessment which fosters intrinsic motivation and uses criterion-referenced scores (the criterion is often established by the teacher, who uses portfolios, for example, as a form of alternative assessment). In contrast, traditional assessment is often perceived as one favoring a one-shot exam (e.g. a standardized exam) which is timed and promotes extrinsic motivation, often using norm-referenced scores (Brown, 2004). Thus, the term *alternative* was proposed (Huerta-Macias, 1995) based on what teachers felt were conceivably shortcomings displayed by traditional tests.

Proponents of alternative assessment believe that alternative assessment is a more equitable form of assessment than traditional assessment (Lynch, 2001) because it includes a variety of measures or instruments to assess students. For example, by using tools such as journals, performance-based assessment, student-teacher conferences, self and peer assessments, observations, and portfolios,

students can be measured in various ways. Therefore, Brown and Hudson (1998) felt that it would be more appropriate to refer to these measures as “alternatives” in assessment, which could be considered a subset of assessment (and not as something that would be excluded from test design or construction).

More importantly, in recent times, *alternatives* in assessment have been embraced by those educators in the mainstream who feel that traditional assessment alone cannot offer the panacea to account for students’ learning outcomes at the post-secondary/tertiary level. It appears that alternative assessment is more sensitive to heterogeneous populations from different backgrounds; that is, ESL and EFL students with and without disabilities (as well as those in SPED programs). As Hamayan (1995, p. 213) stated, “Alternative assessment refers to procedures and techniques which can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom.” More specifically, it would be useful to point out that it is particularly suited for those with limited English skills because as Huerta-Macias (1995, p. 9) claimed, “students are evaluated on what they integrate and produce rather than on what they are able to recall and reproduce.”

4 The Defining Characteristics of Alternatives in Assessment

It is important to underscore that alternatives in assessment need to use authentic or real-world contexts (Brown & Hudson, 1998). It would also be useful to point out that students should be assessed on what would be considered normal everyday classroom activities where both *process* and *product* are equally emphasized. Ideally, these alternatives should be able to provide pertinent information about the students (i.e., their strengths and weaknesses) within the overall context of developing their critical thinking skills. At this point, it may be helpful to examine why the following kinds of student performance should be observed, particularly in a language classroom: discourse-level skills, sentence-level oral production skills (e.g. grammatical features), responses to tasks, interaction with other students, and evidence of listening comprehension or nonverbal behavior, amongst other forms of behavior.

It is important to emphasize that in these types of learning environments, trained instructors (and not machines) should do the scoring, which implies that scoring criteria will need to be determined and raters will need to be trained so that the issue of *reliability* (i.e., consistency and accuracy) would not be compromised (Bachman, 1990). In this context, it seems more appropriate to use “alternatives in assessment” with our special needs students and EFL/ESL students (with or without disabilities) to better reflect the learning outcomes with the following forms of alternative assessment (Brown, 2004):

- Journals
- Conferences and interviews

- Observations
- Self- and peer-assessments
- Performance-based assessment
- Portfolios

Thus, at this juncture, it would be useful to examine each of these *alternatives in assessment* to gain a deeper understanding of how each type of alternative in assessment can be employed to better suit the needs of atypical student populations.

For example, *journals* (Brown, 2004; Staton, Shuy, Peyton, & Reed, 1987) could be employed as assessment tools in the following manner: as classroom-oriented dialogue journals (emphasizing interaction between teacher and student), language-learning logs (self-monitoring one's own achievement in one or more skills), diaries of feelings or attitudes, reflections, grammar journals (monitoring one's own errors), reading response journals, and so on.

As far as *conferences and interviews* (Brown, 2004) are concerned, they could be seen to have the following functions in the classroom: commenting on essays, responding to portfolios, focusing on aspects of oral presentations, monitoring research or project proposals, assessing general progress in a course, advising a student (in relation to a student's academic plan), setting personal goals for the course, and many such related functions. It should be noted that *systematic observations* (not just informal observations) can also be useful in terms of the information that can be gleaned about students (Spada & Frolich, 1995).

Additionally, it is helpful to see how Brown (2004) categorizes *self and peer assessments* into: direct assessment of performance (e.g. student monitors himself after an oral presentation); indirect assessment of performance (for instance, self- or peer-assessment targets larger amounts of time in the form of questionnaires); and metacognitive assessment (more strategic in nature) by setting goals/strategic planning. Students are expected to perform, produce or create something in real-world contexts/situations where classroom activities are an extension of day-to-day life. Students are also trained to tap into critical thinking skills. Thus, as mentioned, this kind of assessment is likely to provide information about students' weaknesses and strengths. Of course, such an assessment would require an open disclosure of standards and rating criteria so that teachers/students are in a better position to make informed decisions.

Clearly, *performance-based assessment* can include interviews, role plays, story retelling, oral speeches or presentations, oral summarizing of articles, oral reports, and so on. Pierce and O'Malley (1992) suggest that teachers could elicit stories from students by offering them a set of pictures (it would be best if teachers could give them visual cues). Role plays are also seen as a way to get students to produce language by transforming themselves into "characters" (Kelner, 1993).

Now, finally, this brings us to the most popular alternative in assessment, namely *portfolios*, which are used to maximize the positive *washback effect* (i.e., the positive effect of testing on instruction and learning).

5 The Use of Portfolios in the United States and Other Countries

For a number of years, portfolios have been used successfully in the United States with mentally challenged students in the field of special education. However, more recently, portfolios are being increasingly used with English language learners in ESL/EFL classrooms in different parts of the world. Portfolios can be envisioned as both *formative portfolios* (this emphasizes process) and *summative* (this emphasizes the learning outcomes) *portfolios* (Cooper & Love, 2001, cited in Ali, 2005). Genesee and Upshur (1996) believe that a portfolio is reflective of a student's growth and achievement in a certain skill or area. What needs to be remembered is that successful portfolio development is linked to well-stated objectives, clearly developed guidelines and criteria for assessment where a variety of entries included in the portfolio assume significance (Yawkey, Gonzalez, & Juan, 1994). The reflective nature of portfolios needs to be underscored in the context of *formative* (i.e., ongoing assessment during instruction) assessment, where the issues of positive *washback effect* and *validity* (i.e., whether a test measures what it is supposed to measure) assume significance.

The acronym CRADLE (i.e., Collecting, Reflecting, Assessing, Documenting, Linking, Evaluation) was suggested by Gottlieb (1995) to highlight the dynamic nature of portfolios. Portfolios are seen to have numerous advantages (O'Malley & Pierce, 1996; Weigle, 2002). Benefits include, for example, a student taking control of the learning process (the teacher is envisioned as a facilitator) because the portfolio becomes the evidence of this student's progress in a course. Furthermore, portfolios can accommodate assessment on a multidimensional basis because there are many different kinds of portfolios that are used for various purposes. From this perspective, portfolios are seen to be employed in much the same way in both the United States and Russia (Reardon & Kozhevnikova, 2008). As Brown (2004) underscores, a successful portfolio would need to subscribe to some of the following guidelines: stating objectives clearly, giving clear guidelines, communicating assessment criteria openly to students, designating time within the curriculum for portfolio development, establishing schedules for conferencing, and providing positive *washback* when generating final assessments.

It might be best if teachers could maintain *anecdotal records* (Tierney, Carter, & Desai, 1991) to keep track of their regularly scheduled conferences with their students so that students would be able to critique their own portfolios. As is commonly understood, portfolios could include numerous kinds of materials (Brown, 2004; Tierney, Carter, & Desai, 1991) including: tests and quizzes, essays, different kinds of writing assignments, reports, projects, poetry, creative prose, artwork, photographs, audio/video recordings of presentations, journals, book reports, grammar logs, checklists created by teachers and students, self- and peer-assessments, reflections, and so on.

Currently, there are five major types of portfolios, which are used in schools in different parts of the world: the *European Language Portfolio*, the *Collections*

Portfolio, the *Showcase Portfolio*, the *Assessment Portfolio*, and the *Electronic Portfolio*.

The *European Language Portfolio* (ELP) is perhaps the first one to be introduced to different systems of education. It appears to be the most popular one in Russia (Reardon & Kozhevnikova, 2008). Based on its widespread use, the ELP is seen to foster student autonomy, with self-assessment being a major aspect of it. Based on common perception, the ELP can be seen to be characterized by the following components:

- It expresses the student's "linguistic identity" (or "passport") by accounting for the second language that has been learned and the student's assessment of his or her current competence (the *passport* could contain certificates that the student has obtained to indicate competence) in the target language;
- It can also be viewed as a language "biography" that can account for one's goals and objectives (the *biography* could contain "self-assessment" materials) in the target language;
- It can be considered a "dossier" which contains a selection of work that monitors the student's growth and progress in the target language (Little, 2005).

The *self-assessment checklists*, which constitute an inherent part of the ELP, are derived from the "Common European Framework of Reference" (CEFR). Students can reflect on their experiences and assess their accomplishments against the checklist. They can also set themselves "goals" and assess their progress by taking "ownership" of the learning process. No matter how beneficial the use of the ELP is, some teachers and students find the ELP cumbersome (the storage of portfolios could prove to be problematic) even though the language portfolio provides clear-cut learning objectives and a way to record growth and progress.

The *Collections Portfolio*, an "aggregate of everything a student produces" (Moya & O'Malley, 1994), consists of a student's entire work, so it is popular at the elementary/primary school level because it represents a variety of daily assignments that a young learner deals with in the classroom and at home. At this age, students do not produce many entries in terms of daily assignments that can go into a portfolio, so the portfolio is quite manageable. For example, portfolios could include a young student's art work, photographs, and even handicrafts. They are almost always kept in the classroom, and are easily accessible to anyone who would like to see these portfolios (e.g. parents or school officials). Maintaining collection portfolios can be very motivating both for young students and teachers because the students can view their own achievements and also the accomplishments of their peers (there can be a big difference in the way these portfolios are personalized and designed). Teachers, on the other hand, can use portfolios to measure each student's progress on a daily or weekly basis.

The *Showcase Portfolio* is often used to exhibit a student's best work (O'Malley & Pierce, 1996). Entries could include (but are not limited to) short stories, essays, poems, audio and video samples. Samples in the portfolio are very carefully selected to illustrate a student's achievement in the classroom and outside the

classroom (i.e. samples that a student considers most representative of his best work). Sometimes, university professors also favor the idea of maintaining such portfolios.

The *Assessment Portfolio* is becoming more popular with teachers all over the world because it is so manageable in that it can use scoring guides and rubrics (i.e. predetermined criteria) to measure a student's work (O'Malley & Pierce, 1996). More importantly, it helps a student to become a responsible learner by taking control of his/her learning process. Using assessment portfolios also gives the teacher opportunities to monitor a student's growth over a period of time, which is an important factor in measuring educational success. Portfolios provide good evidence of a student's achievements for school officials at the secondary level.

The assessment portfolio can be particularly useful in EFL classrooms where a teacher can account for both students with and without *special needs*. Not too long ago, in my own EFL classes (Reardon, 2014) in the Sultanate of Oman, the majority of students claimed that they found *assessment portfolios* to be useful as reflective tools. By looking at their portfolios, they were surprised at how far they had come in acquiring writing skills, for example. Thus, when they looked at the third drafts of their essays, they were actually able to see the progress that they had made. No doubt, this can be especially significant in the case of LD students. For the teacher, it is also a time to reflect on how things could be improved in the future.

The *Electronic Portfolio* is also gaining in popularity these days. As Barrett (2000, cited in Ali, 2005) stated, it consists of the "use of electronic technologies that allow the portfolio developer to collect and organize artifacts in many formats (audio, video, graphics, and text). A standards-based electronic portfolio uses hypertext links to organize the material to connect artifacts to appropriate goals... not a haphazard collection of artifacts (i.e. a digital scrapbook) but a reflective tool". Thus, the electronic portfolio is something that most teachers find useful these days.

6 Portfolio Model and Implementation

What about a portfolio model that teachers can readily access? Moya and O'Malley (1994) offer a portfolio model that has the following characteristics: identifying the purpose (e.g. plan/focus), planning the contents (including how often to assess), designing an analysis of the portfolio (e.g. standards and criteria), preparing for instruction (e.g. giving feedback to students), and verification of procedures (establishing a system for validating decisions). This seems like a good starting point for a teacher who is looking for guidelines to make the portfolio more manageable.

What are some steps that can be followed in implementing portfolios that are manageable? Huang (2012) recommends the following seven steps: planning the assessment purpose, determining portfolio tasks, establishing criteria for assessment, outlining organizational steps, preparing the students, monitoring the portfolio, and assessing the portfolio.

Thus, by using a good model and following these steps, it is possible to make the portfolio a successful assessment instrument and allow the implementation of the portfolio with greater success in all kinds of learning environments.

7 Portfolio Objectives, Criteria, and Assessment

As teachers get students' consent (Ali, 2005) in developing objectives and criteria for portfolio assessment, they are in a better position to specify the time frame and portfolio organization because the latter become actively involved in their own learning and start *reflecting* (Ali, 2002) on their learning styles or strategies, their successful practices or their failures. Learners understand that they can take control of their learning. Self-assessment and peer assessment constitute an important part of portfolio assessment: they increase students' ability to monitor their own progress and set improvement goals. By getting students involved in peer review and including a feedback session midway through the course (Ali, 2005), the teacher can receive a more in-depth knowledge of the student as a learner in terms of his or her learning styles and strategies, and strong and weak points, which means that the teacher can individualize instruction for the student and become more involved in adjusting the curriculum objectives, if necessary.

As mentioned earlier in this chapter, the issues of *reliability* and *validity* are crucial for any assessment tool. Based on our common understanding of testing constructs, *reliability* refers to the consistency and accuracy of the assessment tool (Coombe, Folse, & Hubley, 2007). For a portfolio to be considered reliable, the teacher should develop clear-cut goals and objectives, detailed criteria, explicit and consistent ways of scoring a student's portfolio. The students should fully understand what they are supposed to do and how their work is going to be assessed. Furthermore, the student and the teacher can negotiate the goals and objectives of the portfolio, its content, organization and the expected learning outcomes. Thus, students can develop the feeling of ownership and responsibility when designing their portfolios.

The *validity* of an assessment tool is related to how well this tool does what it is intended to do; that is, whether the tool succeeds in informing the teacher about the student's progress toward some goal in a curriculum or course of study (Bachman, 1990). *Content or curriculum validity* should ensure correspondence between curriculum objectives and the contents of a portfolio. Therefore, the teacher should make sure that the goals of the portfolio are in line with curriculum objectives, the contents of the portfolio match its goal and organization, and the criteria developed are related to the objectives (Delett, Barnhardt, & Kevorkian, 2001). This way, the teacher will integrate instruction and assessment, with portfolios having high *content validity*. One other type of validity relevant to portfolio assessment is *face validity*. This refers to the way the students perceive the portfolio as an assessment tool. If learners do not have a clear understanding of the benefits of the portfolio for their learning outcomes, they will not regard it as a credible assessment tool, and

portfolios will have low *face validity* (Bernhardt, Kevorkian, & Delett, 1999). Thus, before introducing the portfolio as an assessment tool, teachers should carefully explain to the students its benefits and challenges. Then, as long as teachers can establish clear guidelines for how portfolios should be developed and evaluated, portfolios can become extremely *reliable*. No doubt, from the perspective of *washback effect*, portfolios can be considered *valid* tools of assessment.

When portfolios are appropriately designed, they are seen to have the following benefits: they can offer a multi-dimensional perspective of a student's progress over time, promote self-reflection or learner autonomy, and integrate learning, teaching and assessment. Thus, a portfolio can help the student to extend learning and the teacher to gain insight about the student's learning process (Banfi, 2003; Delett et al., 2001).

8 Rationale for Using Portfolios in EFL/ESL Classrooms

The rationale derives from the necessity for developing adaptable alternate assessment techniques that will best suit the needs of diverse bilingual student populations with or without learning disabilities. Typically, with limited English proficiency, EFL/ESL students score poorly on traditional tests (as in standardized tests). Thus, "single-measure approaches" should be avoided by using portfolios which could be used for ongoing "continuous assessment" (Haney & Madaus, 1989).

In this context, it is also crucial to point out that the problem of limited English proficiency is exacerbated when the ESL/EFL bilingual student has learning disabilities which affect learning outcomes adversely. Thus, in considering the following scenario where disability intersects with language proficiency, a *differential diagnosis* is often necessitated, for example, to distinguish between "normal disfluency" (associated with a bilingual student who is considered a "beginner" learning English) and "stuttering" (associated with a bilingual student who has a speech-language disability). Therefore, if an EFL/ESL student has such a speech-language disability (e.g. stuttering) it can make the issue of measuring learning outcomes rather complex. In this context, how can a teacher appropriately evaluate the problem and offer the proper intervention? Clearly, it is important to understand that in diagnosing nonnative speakers of English who stutter, clinicians use an established assessment instrument to measure speech-language proficiency (for example, the SPEAK test) that includes an assessment of *normal disfluency* while conducting a parallel assessment of *stuttering* (for example, the SSI test).

9 Normal Disfluency and Stuttering

What makes the issue of distinguishing “true” stuttering from normal disfluency rather complex is, as Van Riper (1982) notes, the fact that nonstutterers occasionally exhibit similar behaviors as stutterers although stuttering is almost always characterized by “struggle”. According to Peters and Guitar (1991, p. 73), some of the features that distinguish normal disfluency from stuttering are “...the *amount* of disfluency, the *number* of units in repetitions and interjections, and the *type* of disfluency...” (italics added).

Commonly, the speech of nonstuttering bilinguals is also characterized by disfluency in the form of false starts, interjections, and repetitions, often considered the hallmark of second language acquisition. Thus, this type of normal disfluency that is triggered by “interference” or “transfer” gives rise to a variety of possibilities. Conceivably, normal disfluency or stuttering might occur in a similar way in the two languages (Yudaken, 1975, cited in Jankelowitz & Bortz, 1996; Woods & Wright, 1998); or, stuttering may occur in only one of the two languages. Then again, it might occur in both languages but not to the same degree (i.e., depending on the “level” of proficiency), thus underscoring the need to investigate the relationship between level of proficiency and degree of stuttering.

10 Level of Linguistic Proficiency and Degree of Stuttering

When Jankelowitz and Bortz (1996) assessed the language proficiency of an Afrikaans-English bilingual stutterer who used the two languages interchangeably, they reported that linguistic competence influenced frequency, distribution, and nature of stuttering. Findings demonstrated that the subject, a 63 year-old English-Afrikaans speaker, who was a *compound bilingual* (i.e., a bilingual who learns two languages at the same time), stuttered less in Afrikaans, which was his dominant language. Apparently, although both languages were spoken at home, Afrikaans predominated. Clearly, as the authors indicated, the findings demonstrated that linguistic competence influenced stuttering.

In addition, a pilot study conducted by Reardon (1998), involving a Russian-English *coordinate bilingual* (i.e. a bilingual who learns the second language later) stutterer, served to clarify these previous findings by demonstrating that the subject, a Russian-English bilingual, stuttered more in his second or less dominant (less proficient) language, English. Thus, both case studies contributed to the hypothesis that language proficiency seemed to play a role in promoting stuttering; that is, the less proficient a stutterer is in a particular language, the more likely it is that he would stutter.

Thus, Reardon’s (2000) investigation sought to clarify and extend previous research that used very small samples ($N = 1$) in investigating the interaction

between stuttering and language competence in the context of bilingualism. As with past studies (Reardon, 1998), the findings of this investigation (Reardon, 2000) served to support an inverse relationship between language proficiency and stuttering ($r = -0.56$). That is, the greater the proficiency, the less the stuttering. Conversely, the less the language proficiency, the greater the stuttering.

Because the findings of this study indicated that the greater the stuttering the less the proficiency, it led to the following conclusion: as a stutterer gains proficiency in a language, it is less likely that he/she will stutter to the same degree either qualitatively or quantitatively in that language. Interestingly, this finding seems logical in the context of stutterers who engage in substitutions to avoid stuttering. It makes sense that stutterers who are more proficient linguistically would be more able to substitute words that they anticipate as problematic.

Clearly, this finding has pedagogical implications for intervention in EFL/ESL classrooms. It is strongly recommended that a student's speech therapist collaborate with a language specialist (i.e., an EFL/ESL specialist in the case where English is spoken as a foreign or second language) in order to bring about effective intervention. In such a scenario, it is quite possible that as the stutterer becomes more proficient in the language that he stutters in, his stuttering would decrease in that language.

11 Conclusion

Thus, from a pedagogical perspective, it seems crucial that in designing an IEP for a student who stutters and is being offered alternative assessment, that speech-language pathologists understand the differences among EFL/ESL students who are culturally and linguistically different so that they can "accurately discriminate an actual language disorder from the natural progression of second language acquisition and acculturation," (Morsink, Thomas, & Correa, 1991, p. 282). Therefore, it is critically important that the EFL/ESL instructor and the speech-language pathologist work together as an interactive IEP team to enhance learning outcomes in the classroom by using the appropriate alternative assessment tools, for example, portfolios.

In summation, does alternative assessment work in EFL/ESL classrooms? In general, the benefits of alternative assessment are well supported in the literature, especially in relation to portfolios (Allen, 2004; Balhous, 2008; Caner, 2010; Lo, 2010). Alternative assessment provides the teacher with extensive information about students as autonomous learners. As research indicates, most students have positive attitudes toward alternative assessment, in particular toward portfolios.

Clearly, it can be demonstrated that alternative assessment can be successfully used with all students (with and without special needs) in EFL/ESL classrooms at all levels of education. Thus, I do believe that this chapter has contributed to the understanding that alternative assessment can best serve the needs of ESL/EFL populations with and without speech-language disabilities.

References

- Ali, S. Y. (2002). Portfolio assessment: Motivating ESP students to write better through continuous assessment. In S. Troudi, S. Riley & C. Coombe, C. (Eds.), *EFL challenges in the new millennium*. TESOL Arabia 2001 Conference Proceedings (Vol. 6).
- Ali, S. Y. (2005). An introduction to electronic portfolios in the language classroom. *The Internet TESL Journal*, 11(8). Retrieved from <http://iteslj.org>.
- Allen, L. Q. (2004). Implementing a culture portfolio project within a constructivist paradigm. *Foreign Language Annals*, 37(2), 232–39.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Balhous, R. (2008). The self-assessed portfolio: A case study. *Assessment and Evaluation in Higher Education*, 33(4), 381–393.
- Banfi, C. S. (2003). Portfolios: Integrating advanced language, academic, and professional skills. *ELT Journal*, 57(1), 34–42.
- Bernhardt, S., Kevorkian, J., & Delett, J. (1999). *Portfolio assessment in the foreign language classroom*. Washington, DC: National Capital Language Resource Center.
- Brown, H. D. (1999). *New vistas: An interactive course in English*. White Plains, NY: Pearson Education.
- Brown, H. D. (2002). *Strategies for success: A practical guide to learning English*. White Plains, NY: Pearson Education.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.
- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32, 653–675.
- Caner, A. (2010). Students' views on using portfolio assessment in EFL writing courses. *Anadolu University Journal of Social Sciences*, 10(1), 223–236.
- Cawthon, S. (2011). Making decisions about assessment practices for students who are deaf or hard of hearing. *Remedial and Special Education*, 32(1), 4–21.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor: The University of Michigan Press.
- Delett, J. S., Barnhardt, S., & Kevorkian, J. A. (2001). A framework for portfolio assessment in the foreign language classroom. *Foreign Language Annals*, 34(6), 559–68.
- Elliot, S. (2010). Effects of using modified items to test students with persistent academic difficulties. *Exceptional Children*, 76(4), 475–495.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Gottlieb, M. (1995). Nurturing student learning through portfolios. *TESOL Journal*, 5, 12–14.
- Hager, K., & Slocum, T. (2011). Using alternate assessment to improve educational outcomes. *Rural Special Education Quarterly*, 30(1), 24–29.
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212–226.
- Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan*, 70, 683–687.
- Huang, J. (2012). The implementation of portfolio assessment in integrated English course. *English Language and Literature Studies*, 2(4), 15–21.
- Huerta-Marcia, (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal*, 5, 8–11.
- Jankelowitz, D. L., & Bortz, M. A. (1996). The interaction of bilingualism and stuttering in an adult. *Journal of Communication Disorders*, 29, 223–234.

- Johnson, P. (2012). The benefits and administration of alternate assessments. *New Horizons of Learning*, 10(1). Retrieved from <http://education.jhu.edu/PD/newhorizons/Journals/Winter2012/Johnson>.
- Kearns, J., Towles-Reeves, E., Kleinert, H., Kleinert, J., & Thomas, M. (2011). Characteristics of and implications for students participating in alternate assessments based on alternate academic achievement standards. *Journal of Special Education*, 45(1), 3–14.
- Kelner, L. B. (1993). *The creative classroom: A guide for using creative drama in the classroom, preK-6*. Portsmouth: Heinemann.
- Kettler, R. (2011). Computer-based screening for the new modified alternate assessment. *Journal of Psychoeducational Assessment*, 29(1), 3–13.
- Little, D. (2005). The Common European Framework and the European Language Portfolio: Involving learners and their judgments in the assessment process. *Language Testing*, 22, 321–336.
- Lo, Y. (2010). Implementing reflective portfolios for promoting autonomous learning among EFL college students in Taiwan. *Language Teaching Research*, 14(1), 77–95.
- Lynch, B. K. (2001). The ethical potential of alternative language assessment. In C. Elder (Ed.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 228–239). Cambridge: Cambridge University Press.
- Morsink, C. V., Thomas, C. C., & Correa, V. I. (1991). *Interactive teaming: Consultation and collaboration in special programs*. New York, NY: Macmillan.
- Moya, S. S., & O'Malley, J. M. (1994). A portfolio assessment model for ESL. *The Journal of Educational Issues of Language Minority Students*, 13, 13–36.
- Musson, J., Thomas, M., Towles-Reeves, E., & Kearns, J. (2010). An analysis of state alternate assessment participation guidelines. *Journal of Special Education*, 44(2), 67–78.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. White Plains, NY: Addison-Wesley.
- Peters, T. J., & Guitar, B. (1991). *Stuttering: An integrated approach to its nature and treatment*. Baltimore, MD: Williams and Wilkins.
- Pierce, L. V., & O'Malley, J. M. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.
- Reardon, V. S. (1998). *Disfluency: A cross-linguistic comparison*. Unpublished manuscript, University of Kansas.
- Reardon, V. S. (2000). *Stuttering in the context of bilingualism: The relationship between language proficiency and degree of stuttering*. Unpublished doctoral dissertation, University of Kansas, USA.
- Reardon, V. S. (2014). *Special needs learners, diagnosis and intervention: Alternative assessment*. Paper Presented at the Oman 14th International ELT Conference, Muscat, Oman.
- Reardon, V. S., & Kozhevnikova, L. (2008). Alternative assessment and the use of portfolios in the United States and Russia. In *International Academic-Applied Conference, Moscow, Russia*. Moscow: Cambridge University Press.
- Reardon, V. S., & Nagaswami, G. (1997). *Attenuating speech disability with e-mail*. Paper Presented at the Thirty-first Annual Convention & Exposition of the Teachers of English to Speakers of Other Languages, Inc., Orlando, Florida.
- Roach, A., Beddow, P., Kurz, A., Kettler, R., & Elliott, S. (2010). Incorporating student input in developing alternate assessments based on modified academic achievement standards. *Exceptional Children*, 77(1), 61.
- Spada, N., & Frolich, M. (1995). *Communicative orientation of language teaching observation scheme*. Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Staton, J., Shuy, R. W., Peyton, J. K., & Reed, L. (1987). *Dialogue journal communication: Classroom, linguistic, social and cognitive views*. Norwood, NJ: Ablex.
- Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood: Christopher Gordon.
- Van Riper, C. (1982). *The nature of stuttering*. Englewood Cliffs, NJ: Prentice-Hall.

- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Woods, D. W., & Wright, L. W., Jr. (1998). Dismantling simplified regulated and experimental breathing: A case of a bilingual stutterer. *Journal of Behavior Therapy Psychiatry, 29*, 179–186.
- Wright, P., & Wright, P. (2007). *Wrightslaw: Special education law* (2nd ed.). Virginia: Harbor House Law Press Inc.
- Yawkey, T. D., Gonzalez, V., & Juan, Y. (1994). Literacy and biliteracy strategies and approaches for young culturally and linguistically diverse children: Academic excellence P.I.A.G.E.T. comes alive. *Journal of Reading Improvement, 3*(3), 130–141.

Alternative Assessment: Student Designed Test Evidence in an Iranian EFL Context

Jafar Dorri Kafrani and Mohammad Reza Afshari

Abstract In the past decade, educators have realized that alternative assessments are an important means of gaining a dynamic picture of students' academic and linguistic development (Tannenbaum, 1996). Alternative assessment consists of portfolio assessment, self-assessment, projects, observations, presentations, journal keeping, interviews, and student designed tests (Coombe, Folse, & Hubley, 2007). The purpose of this study is to find out how the process of having students design tests would help them and their classroom teachers in assessing students' progress and learning. To this end, 120 junior high school male third graders studying at Imam Khomeini School located in Tehran, Iran were selected, grouped, and asked to design tests based on their English course book. Each group was given two chapters of their course book to design tests. Students were free to write as many test items in any format for any skills or sub-skills they would like. The data collection procedure was done in the classroom so they could not use sample test items nor could they use their workbooks in which students can find some sample tests. The collected questions were analyzed in terms of test format as well as skills and sub-skills. Furthermore, in order to have a better understanding of the probable reasons behind designing such tests by the students, both students and their teacher were interviewed. The results of the study revealed that students benefited from their tests, and the study helped them to review the book content in detail. However, the results revealed that students did not pay enough attention to certain key parts of each chapter, and their test-item formats suffered from a lack of variety.

Keywords Alternative assessment • Student designed test

J.D. Kafrani (✉)
University of Tehran, Tehran, Iran
e-mail: jafar_dorri@yahoo.com

M.R. Afshari (✉)
Islamic Azad University, Science and Research Branch, Tehran, Iran
e-mail: muhammadreza.afshar@gmail.com

1 Introduction

The importance of testing and assessment is not lost to scholars, teachers, and students in the EFL/ESL literature. Testing and teaching are so closely interrelated that it seems it is impossible to work in one field without being concerned with the other. Among all testing methods, alternative assessment has become a vital tool for educators wanting to gain an understanding of learners' academic and linguistic developments. Hancock (1994) defines alternative assessment as an ongoing process involving the student and the teacher in making judgments about the student's progress in non-conventional language strategies.

Student-designed tests are one of the many types of alternative assessment; however, there is a dearth of research on this subject in the literature. Educators believe that student-made tests enable the teachers to see where the gaps in their students' understanding are and at the same time it provides the students with opportunities to review for an upcoming test and to relate the course to their own interests. Furthermore, student-designed tests are good practice and review activities that encourage students to take responsibility for their own learning (Coombe, Folse, & Hubley, 2007).

Among all the alternative assessment methods, the focus of this chapter is on student-designed tests. We investigated the effectiveness of students' involvement in testing and assessment to see whether it could help them to better learn the course content and have a better understanding of their own progress. In addition, we set out to know if this type of task could help the teacher diagnose their students' weak points so as to enable them to find reminders for their area of weakness.

2 Review of the Literature

There is a mutual relationship between language tests and language learning and teaching. Tests assist teachers in a number of different ways. They help teachers diagnose their students' strengths and weaknesses. They often give teachers insights into the process of teaching and feedback on learning (Bachman, 1990; Brown, H.D. 2004). Thus, testing and assessment have gained great importance in the literature.

Lambert and Lines (2000) define assessment as "the process of gathering, interpreting, recording and using information about pupils' responses to educational tasks". Likewise, Coombe, Folse, & Hubley (2007) state that assessment refers to varied ways of gathering information about learners' language abilities as well as their achievements. Assessment can be either formative or summative. Formative assessment examines the way students develop, but summative assessment looks at what students have achieved during a course of study (Lambert & Lines, 2000).

Regarding assessing students' development during the course, Hamayan (1995) suggests alternative assessment and defines it as the procedures and techniques

which can be used in the context of instruction in *everyday school or classroom activities*. Alternative assessment might also happen out of a classroom and the subjects being tested might be asked to demonstrate their knowledge in different ways (Smith, 1999; Tannenbaum, 1996).

Tsagari (2004) lists the following advantages of alternative assessment: (a) aids in the evaluation of the process and product of learning as well as other important learning behaviors, (b) enables the evaluation and monitoring of instruction, (c) supplies meaningful results to a variety of stakeholders (d) relates learning to cognitive psychology and related fields, (e) represents a collaborative approach to assessment, (f) supports students' psychologically, (g) promotes autonomous and self-directed learning, and (h) provides new roles for teachers. Furthermore, she has listed the most common methods of alternative assessment in her article as: conferences, debates, demonstrations, diaries/journals, dramatizations, exhibitions, games, observations, peer-assessment, portfolios, projects, self-assessment, think aloud, story retelling, and student-designed tests.

An approach within alternative assessment is to have students write tests on course material. This process results in greater learner awareness of course content, test formats, and test strategies. As mentioned previously student-designed tests are good practice and review activities that encourage students to take responsibility for their own learning (Coombe et al., 2007; Brown, J.D. 1999), and enable the teachers to see where the gaps in their understanding are and at the same time provide the students with opportunities to review for an upcoming test and make the course more relevant to their own interests. Baron (2004) also argues that if students are to take more responsibility for their own learning, they should be involved in the assessment process, and that students should not merely be passive recipients of results, but have a voice in the designing of some aspects. She believes this procedure (a) involves the students in the assessment process (b) helps students to realize what materials and what types of issues are the most important, and (c) provides a fresh pool of questions and promotes partnership between teacher and the students.

3 Method

The purposes of this study are multi-faceted. The first aim is to determine what students chose to include in terms of (a) test content such as listening, speaking, reading, writing, vocabulary, grammar, spelling, and pronunciation in terms of the number of test items and (b) test format (multiple choice, true/false, short answer, matching, etc.).

Moreover, the researchers sought to find out whether students focus on all aspects of a unit to design their tests or whether certain parts of a unit are considered important for students in designing their tests.

To these ends, the following research questions were considered:

1. Are there any specific areas students focus on regarding test content and test format in student-designed tests?
2. Will students be creative in their tests in general or will they follow a certain test format?
3. Will students consider all sections of a unit worthy of inclusion when writing their tests?

In the following section, the researchers describe the process of data collection and give information about the participants and the instrument for data collection.

4 Participants

One hundred and twenty 14-year old male junior high school third grade students participated in the trial. All of them were studying English as a foreign language and were chosen, in part, because their English teacher agreed to allocate a session for this study.

The school is located in district 15 in the South East of the Tehran province. Students took part in English classes over three sessions a week, and each session lasted 90 min. It was the last month of the educational year in Iran when the researchers had students design questions from their English books. Their teacher, 33, held a BA in TEFL and had 12 years of teaching experience.

5 Instruments

The main instrument in this study is student-designed tests. In addition, we conducted a semi-structured interview with the teacher of the class at the end of the study to gauge his opinion on the probable reasons behind certain aspects of the student-designed tests. The investigators also wanted to know the teacher's opinion about the merits and demerits of the present study and whether or not motivating students to design tests would foster their learning. In addition, students participated in a focus-group interview to see if they liked the experience.

6 Procedure

Based on the research aims, the students in each class were put into five groups consisting of *four* students each. All groups were heterogeneous—containing a top, two mediocre, and one weak student. The categorization of the groups was based on each student's mean score from eight previous formative tests they had taken. Finally, after collecting and analyzing the data, the results were shown to the

teacher. The classroom teacher was asked to share his opinions of the student-designed tests and the merits of the research project in a semi-structured interview with the researchers, which took 30 min (see Appendix 1 for the interview questions). The interview was recorded, with the interviewee's consent, then transcribed and analyzed.

Following that, based on the second and third research questions, students were put into groups of 15 to take part in a focus group interview. The interviews which took nearly 40 min were recorded, transcribed, and analyzed. In the interview with students, the researchers were able to find students' rationale behind the tests they designed (see Appendix 2 for the interview questions).

7 Data Analysis

The collected data from the student-designed tests were carefully studied by the researchers. The data was labeled, categorized, and then counted. The interviews were recorded, transcribed, and analyzed. Below is a report on the analysis of the collected data.

8 Student-Designed Tests Results

The test items developed by the students were carefully labeled, categorized, and then counted. The outcome was organised into two main categories: (a) skills and sub-skills, and (b) test format, which will be elaborated upon in turn.

9 Skills and Sub-skills

The first category analyzed is skills and sub-skills. It was important for the researchers as well as the classroom teacher to see which skills and sub-skills were more popular in SDTs. The analyzed data revealed that grammar tests topped the list. As Table 1 illustrates, 62 % of the student-designed test items were grammar related. Next behind grammar were vocabulary items, at 21 %, which is a substantial drop. Pronunciation, at 10 %, was the third most common test item. The last two test item categories were reading and writing with 5 and 2 % respectively. As the table depicts, none of the designed tests contained items related to listening and speaking.

Table 1 Skills and sub-skills covered by items in student-designed tests

	%
Grammar	62
Vocabulary	21
Pronunciation	10
Reading	5
Writing	2
Listening	0
Speaking	0

Table 2 Test format in student-designed test

Test item format	%
MCQ	50
Short answer	16
Unscramble	13
Transformation	9
Spelling	6
True/False	4
Matching	2

10 Test Format

The second categorization of the data is test item format. The format of test questions included in the SDTs is important for the researchers to know since they wanted to determine if students see the language as discrete items or integrated.

As shown in Table 2, the most common question format in the tests was multiple choice. They top the list at 50 %. The next most prevalent test item format is considerably below MCQ and stands in second place with only 16 %. After short answers, unscrambled sentences were the third most popular item format with 13 %. The last four item formats are transformation, spelling, T/F, and matching with 9, 6, 4, and 2 % respectively.

11 Teacher Interview After Data Analysis

The teacher of the class was interviewed to garner his opinion regarding the study. Following this, the recording of the interview was transcribed for further analysis. According to the teacher of the class, the procedure had its advantages and disadvantages; however, the positive points outnumbered the negative ones. He believed that student cooperation in such tasks would result in peer learning and that the resulting student discussions over the right answer will no doubt help students in general and weak students in particular to master the lesson. He stated:

“Students learn from each other. They ask questions and discuss the answers and this ends in learning. They also review what I had taught them.”

Also, he believed that the experience of designing their own tests would give students self-confidence in their exams. In addition, there was competition among different groups to write better questions and the interaction among each group member was impressive. Another benefit noted was that this adds variety to in-class activities, which stops the class from being monotonous and boring. Finally, according to the teacher, the most important outcome of the study is that students study the contents of the course book.

However, we recorded a number of perceived disadvantages of student-designed tests. First of all, they were considered time consuming: “It is really time-consuming to spend a session on this task, particularly if this is going to be done every now and then”. Second, a problem that is thought to exist in this activity is task deviation. That is, those groups who finished the task sooner than the rest may distract others in the class because they had nothing else to do. His recommendation for this problem was to keep the students busy with other tasks.

The teacher also stated that based on what students had designed he could find out which part of each lesson was not of interest to students or might not have been learned. This can be determined since there is likely to be no test item from those parts or the items would be very poor. He defines poor items as the ones for which it is easy to find the answers or the test format is not selected appropriately.

12 Student Interview After Data Analysis

Students took part in a focus group interview and the list below summarizes what they think of as advantages of student-designed tests (Table 3).

What students mostly commented on in their responses to interview questions was the revision over the content of the course book. Students believe that the experience helped them to read the course book in depth and find answers to their questions. They also liked the group work and the cooperation involved in writing tests. The discussions they had when coming up with distractors and the right answer helped them to learn better and recognize their weak points. What also interested them was the increased familiarity with different test types, which helps prepare them for future exams.

Table 3 Advantages of student-designed tests from students’ perspectives

Advantages	%
Review of the course book content	55
Recognition of our weak points	20
Familiarity with question samples	15
Preparation for future exams	8
Cooperation between students	2

Table 4 Disadvantages of student-designed tests from students' perspectives

Disadvantages	%
No disadvantages	70
Uncooperative students	15
Lack of time	11
Copying from test books	2
Inability to write good tests	1

Students were also asked to name some disadvantages associated with the task. Table 4 summarizes their views:

As Table 4 depicts, the majority of students (70 %) believe that there are no disadvantages to this activity. However, 15 % of the students stated that some group members were uncooperative, and other group members had to shoulder their responsibilities. Ten percent of the students felt they needed more time (more than an hour) to design better quality test items. In addition, 2 % of the students believed that some group members simply copied test items from test books and used them. Finally, 1 % of the students stated that they were unable to design good tests and needed some sort of training on how to develop a good test.

13 Findings and Discussion

This study has some very interesting positives which stunned the classroom teacher and the researchers. The collected data from the SDTs are a reflection of the Iranian educational testing system context these students are in. In the Iranian context, students have to take tests which are mostly multiple choice questions replete with grammar, especially in special high school entrance exams or university entrance exams. This is in-line with one of the most important findings of this study regarding item types (Grammar: 62 %; Multiple-choice questions: 50 %).

While SDTs contained many vocabulary and grammar questions their formation was found to help students to review course content, particularly if it is carried out in a group. Individuals might consider some specific sections of a unit to be more important than other parts which may lead to an imbalance in course material coverage, whereas in groups, everybody is involved in making decisions on what to include in the design. This group task can therefore provide students the opportunity to read, review, and learn; and consequently have a new look at what they considered unimportant or might not have learned in the first place. SDTs guide the students toward what to study. Therefore, we consider SDT an effective tool in students' learning.

In addition, SDTs can be a tool for the teacher to provide an engaging opportunity for students to review the content of the course book. Teachers usually design tasks and activities such as role-plays, games, information gap activities, etc., in order to review what they have covered during a month or two. SDT could be used as a fresh alternative for students to review the content.

An interesting occurrence in this study was the reduction of test anxiety mentioned by students. Since students are involved in the process of designing tests, and the final product is shared amongst the groups, students became familiar with a wide range of test items and formats. Since the teacher will use student-made tests as the classroom quizzes, this gives students more self confidence in answering exam questions because students have already designed or studied similar test items in advance. Moreover, they have studied more of the content and have discussed the right answers in groups which particularly helps weak students to learn the content. Thus, students will have less stress in exams since they have had ample preparation and are familiar with exam format and content.

As the collected data supports, there is not much focus on the skills of listening, speaking, and writing in the Ministry of Education books that are being taught in schools in Iran. On the contrary, the books focus merely on reading, vocabulary, grammar, and pronunciation. It seems that the reason why the student-designed tests had a greater focus on grammar, vocabulary, and reading, as Table 1 shows, is the contents of their books.

Considering the format of the questions; students mostly designed tests using MCQs. We believe that the high proportion of MCQs is due to the fact that students have to take an English MCQ test once a month. The purpose of this test is for the school and school teachers to check students' progress. These tests have influenced students' point of view regarding test format and test content.

14 Some Guidelines

Based on the results of the investigation, as well as the researchers and the classroom teacher's experience in this study the following guidelines are humbly suggested for those who are interested in utilising SDTs in their own specific context.

1. Allocate some time in your course.

It is crucial to allocate some specific time in our course plans in general and lesson plans in particular. When teachers are thinking and writing a plan for a course of study, they should allow some time for SDTs. Following are some aspects of the plan for a SDT.

2. Brief your students on the process.

When students know why they are designing tests and are told about the probable benefits, they will be more willing to design tests. Also, teachers should talk about the process. This awareness reduces students' stress and puts them at ease when generating tests.

3. Group your students based on their abilities (homogeneous/mixed-ability).

Teachers have two options in grouping students: (a) homogeneous groups, (b) heterogeneous groups. Each way of grouping might have its own advantages

and disadvantages. However, what worked well in this study was groups with mixed-ability students in which the good students helped the weaker students. In homogenous groups, the weak-student groups might be left on their own and consequently, less learning might occur and the final product (SDT) might not be very effective.

4. SDTs should be done in the classroom.

There are many test books, workbooks, and on-line sample tests which are at students' reach outside the classroom. Moreover, some students might miss the homework and come to class without doing their homework. Due to one of the goals of SDT (group activity, learning, reviewing the content, etc.), it is highly recommended that the task be done in the classroom. This also allows time for feedback during the process. While students are doing the task and writing their tests, teachers can monitor group dynamics and provide feedback where necessary.

5. Use students' final products.

A very important suggestion here is to **use** students' tests in one way or another. This will give students a sense of achievement and ownership as well as willingness for further group activities and more classroom tasks. Students' final product can be posted on the school website, classroom boards, or even used as classroom quizzes.

15 Conclusion

Student-designed tests can assist students to review the course material. Moreover, they foster learning and motivate students to be actively involved in the classroom dynamics. In addition, student-generated tests can broaden the teacher's view on what students consider important or trivial. Likewise, the student-made questions can mirror the sections that students consider unimportant. By engaging students in designing tests, the teacher can reduce students' test anxiety and make students like exams and tests. When the classroom quizzes are selected from students' productions, they can make sure they will get a good mark which helps them to be more confident at the exam session.

Appendix 1: Interview Questions with the Classroom Teacher

1. How did you find the experience?
2. Do you think students are capable of designing tests? Why? Why not?
3. How do you think this experience will help you in your teaching?
4. How will this experience help students in their learning process?

Appendix 2: Interview Questions with the Students

1. Did you like designing tests? Why? Why not?
2. How did the group work help you in designing tests?
3. What did you learn from this experience?
4. What are the advantages and disadvantages of student-designed tests?

References

- Bachman, L. F. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University Press.
- Baron, L. (2004). *Student's creating their own tests is a way to promote active learning*. Retrieved from <http://www.literacyplus.ca/Baron/baron-student-creat-tests.html>.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Pearson Education.
- Brown, J. D. (1999). *Testing in language programs*. New Jersey: Prentice Hall.
- Coombe, C., Folse, K., & Hubley, N. (2007). *A practical guide to assessing English language learners*. Ann Arbor, MI: Michigan University Press.
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212–226.
- Hancock, C. R. (1994). Glossary of selected terms. In C. R. Hancock (Ed.), *Teaching, testing and assessment: Making the connection* (pp. 235–240). Lincolnwood, Ill: National Textbook Company.
- Lambert, D., & Lines, D. (2000). *Understanding assessment. Purposes, perceptions, practice*. London: Routledge.
- Smith, K. (1999). Language testing: Alternative methods. In B. Spolsky (Ed.), *Concise encyclopaedia of educational linguistics* (pp. 703–706). Amsterdam: Elsevier.
- Tannenbaum, J. (1996). Practical ideas on alternative assessment for ESL students. *ERIC Digest* (ED395500).
- Tsagari, D. (2004). Is there life language testing? An introduction to alternative language assessment. *CRILE Working Papers*, 58, 2–23.

From a Culture of Testing to a Culture of Assessment: Implementing Writing Portfolios in a Micro Context

Elizabeth Noel

Abstract Changes in education theory have pushed the English language teaching community toward a learner centered approach to education. Learner centered classroom methodologies have brought about the need for learner centered evaluations of students. Alternative assessments have been widely adopted over the last 25 years and in particular there has been a focus on the use of portfolios in assessing college level writing. Reasons for this include the belief that portfolios can mirror good classroom practice and target language use, integrate curriculum goals, help students and teachers to self-reflect, and highlight the communicative purpose of writing. However, there are challenges to overcome in the use of assessment portfolios as their creative nature allows for wide variations in construction and interpretation, which brings reliability into question. Some of these challenges and benefits are being felt in the English for Academic Purposes Department at the University of Technology and Business (UTB) where a modified curriculum created the need for different assessment procedures. This chapter briefly outlines the change in ELT from a culture of testing to a culture of assessment and then presents some of the changes occurring in the micro context of UTB, where portfolio assessment is being implemented to help solve curricula and other issues.

Keywords Assessment • Portfolios • Micro context

1 Introduction

Education theory since the 1980s has undergone major change. The change has been characterized by a shift in the teachers' role from knowledge provider to facilitator and the view of the student from passive receiver to an individual acting and reacting within a social environment. The new emphasis has pushed the English language teaching community toward a learner centered approach to education

E. Noel (✉)

Higher Colleges of Technology, Abu Dhabi, UAE
e-mail: Lizzybee68@hotmail.co.uk

(Kohonen, 1999, p. 280). Learner centered classroom methodologies have in turn brought about the need for learner centered ways of evaluating students and according to Short (1993, p. 630), these ways have dominated the educational reform dialogue. Alternative assessments have been widely adopted over the last 25 years and in particular “interest in and commitment to portfolios for assessing college writing has swelled enormously” (Hamp-Lyons & Condon, 2000, p. 176). Reasons for this include the belief that when implemented effectively portfolios can mirror good classroom practice and target language use, integrate curriculum goals, provide further evaluation of teachers and the curriculum, and even improve student awareness and self-esteem (Callahan, 1995, p. 118). However, there are major challenges to overcome in the use portfolios for assessment. Their creative nature allows for wide variations in construction, and therefore interpretation, which brings reliability into question.

Some of these challenges and benefits are being felt in the English for Academic Purposes Department at the University of Technology and Business (UTB) where a new curriculum has created the need for assessment procedures that better reflect philosophies of student centeredness. This is especially crucial in the Middle Eastern context where “the development of individual work skills tends not to be perceived by the student as part of the learning process...the students will more readily apply themselves to the more academic tasks of passing tests” (Richardson, 2006, p. 111). An ideal scenario therefore, is for tests to foster language and wider learning skills.

This chapter briefly outlines the change in ELT from a culture of testing to a culture of assessment and then presents some of the changes occurring in the micro context of UTB. This chapter outlines the reasons for choosing portfolio assessment, such as validity, authenticity and washback effects, and then goes on to explain how issues of reliability and authenticity of authorship will be met. The study focuses mainly on the starter and exit levels of English at UTB.

2 A New Culture of Assessment

Testing, teaching and learning are without doubt heavily intertwined. Anticipation for a test can influence curricula and classroom decisions as well as guide students in their own revision and preparation processes, especially when those tests are of a high-stakes nature. Traditional standardized tests refer to pencil and paper exams, often one shot, under timed conditions that allow for one right answer. During the test students are forbidden to interact and feedback is often minimal. “Interestingly, much of the research supporting the power of testing to influence schooling is based on traditional standardized tests and concludes that such tests have a negative impact on program quality” (Herman, 1992, p. 74). This is easy to understand if a language class is organized around a test that has no social aspect. The inclusion of

social aspect is crucial because when the need to interact with someone through writing or speaking is removed it can be argued that “the learners are dealing with strings of words and not with language at all” (Rinvolutri, 1999, p. 195).

Twenty-five years ago Fredriksen and Collins (1989, as cited in Shohamy 2001, p. 142) challenged the assumption that the key goals of testing should be to contribute to improved learning and be highly connected to classroom practice. The alternative assessments that have since arisen to try and meet these demands are varied. Through them the learner is evaluated “using activities and tasks that integrate classroom goals, curricula and instruction and real life performance. It emphasizes the communicative meaningfulness of evaluation and the commitment to measure that which we value in education” (Kohonen, 1999, p. 284). In short the new culture of assessment entails a shift from testing for testing’s sake to assessment for learning’s sake. Assessment is no longer something teachers do to students, but a process that requires the student’s direct involvement. This helps to foster both cognitive and affective development of the individual.

3 Challenges and Changes at the University of Technology and Business

The first steps toward assessment for learning’s sake are taking place in the English for Academics department (EAP) at UTB. UTB is in its fourth year of operation and has undergone major curricula changes since its founding in 2006. EAP has by no means been exempt from those changes. The original EAP curriculum has been adapted almost to the point of no recognition in an attempt to better meet the needs of its students within the Middle Eastern context of the United Arab Emirates (UAE). There are two challenges affecting UTB and the wider Gulf region. Firstly, “at an early stage in the learning process students develop a passive, teacher centered learning style which inhibits progress in the language classroom” (Richardson, 2007, p. 247). Secondly, students are often “sure they can pass the course with the minimum effort and consider the English class session a fun period” (Rizk, 2006, p. 96). This self-assurance is often accurate as the Ministry of Education in the UAE only requires a band 5 IELTS/TOEFL 500 for studying on academic programs. In most western located universities, a band 6 or above is mandatory. This has led to a huge challenge for EAP because students’ single orientation to pass IELTS at a band 5, and pressure from the wider university to equip students with the skills of band 6 have collided. One of the biggest concerns has been writing because in TOEFL the writing does not contribute to the total point score considered by UTB and in IELTS an amalgamated grade is accepted. Therefore, a student may receive a score as low as 3.5 in one skill and still enter their academic program if other skill scores make the average 5.

To try and address these issues, two steps have been taken. Firstly, a more learner centered curriculum has been employed with a key aim of providing students with both language and learning skills. The new curriculum is organized around an integrated skills approach and has five levels each lasting seven weeks. Secondly, before enrolling in compulsory credit language courses (CLC), which are heavily focused on academic reading and writing and make up part of the final degree, students must pass an exit test from EAP 5. This means that those students who gain IELTS 5 or TOEFL 500, but fail EAP 5 *will* be granted access to their programs but *will not* be granted access to CLC, and therefore can potentially be held back from graduating. Instead those students will have to take remedial English classes until such a time that they can pass the exit/entrance test. These two initiatives have called for assessment procedures within EAP to be reevaluated to better meet two needs;

- align assessment with the more student centered syllabus
- focus exam preparation not only on requirements of IELTS and TOEFL but also on the academic requirements of credit English courses referred to as CLC.

One step considered appropriate to help meet these needs is to introduce portfolio assessment.

4 Portfolio Assessment at UTB

Portfolios have long been used in professions such as architecture, journalism and photography for keeping a log of personal accomplishments. A second language portfolio is defined by Kohonen (1999, p. 286) as “a purposeful selective collection of learner work and reflective self-assessment that is to document progress and achievement over time with regard to specific criteria”. Though portfolios have been cited as tools for assessing general language ability (Kohonen, 1999), it is the writing portfolio that has been most widely adopted (Genesse & Upshur, 1996, p. 101). Writing portfolio assessment will be implemented across the five levels of EAP with the primary purpose of evaluating individual student progress in writing ability at the end of each seven-week course. In order to be a portfolio more than one piece of writing must necessarily be collected. Though the literature implies the more writing the better, Weigle (2002, p. 215) cautions that “Concern for adequately sampling the domain must be balanced by concerns of practicality”. As each level is only seven weeks long, in practical terms three to five pieces, in addition to other assignments, is adequate and realistic. A minimum of five pieces of edited writing will therefore be made mandatory from level one to three, four in level four and three in level five when students make the transition from paragraph to essay.

A secondary purpose of the writing portfolio is to provide a tool for teachers to assess student needs at various stages in the course. The range of pieces will allow for a broader interpretation of student abilities and therefore a more accurate picture of what students can do. This is particularly true of academic situations (Weigle, 2002, p. 202) because the range of genres allows teachers to identify strengths and weaknesses with particular genres of writing. Furthermore, by reviewing multiple drafts of writing, teachers can recognize where extra guidance with the writing process is required. For example, students may be able to edit at the sentence level but not at the paragraph or whole essay level. An added benefit of the collection is that it measures growth along specific parameters over a given length of time, in this case five blocks of 7 weeks, so teachers can see if those identified weaknesses have successfully been addressed. Due to these advantages, portfolios allow for better construct validity than traditional methods. “They give a better picture because they give a broader picture of what the writer can do, over a long period of time, under valid circumstances, in response to a number of writing opportunities” (Callahan, 1995, p. 127).

Each level will produce specific types of writing depending on curriculum goals (see Appendix 1). For example, in level one the five pieces consist of an informal description of a process, person and place, a review, and a short narrative. In level five the three pieces are an analysis of tabulated or graphic data, a critical response and summary of a reading, and a five paragraph essay. Hamp-Lyons and Condon (2000, as cited in Weigle, 2002, p. 212) suggest that student control of content is essential if student efficacy is the primary goal. However, they warn that students do not always make the best choices and may be swayed, for example by a preferred topic. Therefore, students will have limitations to their choices and be offered guidance from teachers and eventually peers. Choices will become increasingly wide as the levels progress and may include selections from multiple attempts at single genres, how to document understanding of the writing process, as well as portfolio layout and design of cover and content pages. It is only in level five that students will be expected to choose an essay that represents their best effort in writing. In order to make these choices students will need to reflect on their achievements.

5 Reflection

Reflection is a natural part of portfolio practice because it makes “students agents of reflection and decision making and thus gives them control over their own learning” (Genesee & Upshur, 1996, p. 105). As previously mentioned, portfolios can be vehicles for teachers to assess student abilities, but better still they allow students to do this for themselves, a skill greatly valued in modern pedagogy. Over the five levels students will be expected to become increasingly reflective and able to

self-assess. This gradual increase is because the concept of reflection will be new to most students at UTB, and as Ellili and Chaffin (2007, p. 307) point out, students in the Middle East need “to be guided along a continuum from total dependence on the teacher toward greater autonomy”. Simply asking students to be reflective is unlikely to be successful. The guidance can further be supported by Gottlieb’s (1995) CRADLE approach to portfolios. CRADLE is an acronym for collection, reflection, assessment, documenting, linking and evaluating. She suggests that a movement along the CRADLE continuum will ensure better portfolio practice. Accordingly, in level one, students are primarily only expected to collect and as they progress through the levels they will reflect, assess, link and evaluate.

Reflective behavior will be encouraged and developed in the following ways. First in level one, students will be given the opportunity to improve the grades of the individual pieces by responding to teacher feedback and editing work appropriately before resubmission. The focus in level one is primarily about familiarizing students with portfolio collection and organization. By allowing students to improve grades it is hoped they will be motivated to carefully consider teacher feedback and also recognize that writing is a cyclical process and not a one shot event. Students will be asked to include the additional drafts in the portfolio for submission. In level two students will begin to develop process writing skills of multiple drafts with self-editing using evaluation frames. Level two students will also have to provide short answers to a self-report form (Appendix 2) which further encourages students to reflect on what they have achieved. In level three students will begin to work collaboratively using evaluation frames to check each other’s work as well as their own. Peer assessment provides an authentic audience, beyond the teacher, to communicate with. It provides a basis for comparison with their own work, which in turn can aid self-assessment practices; as Donato and McCormick (1994, p. 461) mention, they have something solid to make comparisons with.

Peer assessment “fosters the idea that writing is a process of communicating to an authentic audience” (Mangelsdorf, 1992, p. 274). Based on research into second language learning, peer assessment promotes L2 development (Saito, 2008). Writing portfolios lend themselves to peer assessment through the sharing of written work. Saito (2008) advises training students to peer assess as in her study the trained group gave better peer advice than the untrained group. At UTB, students will be trained gradually in class through the whole class assessing writing before it is done in pairs. Peer assessment also encourages the development of interpersonal intelligence which “is characterized by the ability to understand and respond to others” (Chamot, 1995, as cited in Arnold & Brown, 1999, p. 4). By including peer and self-assessments a concern for educating the whole person and not just “from the neck up” (Rogers, 1975, as cited in Arnold & Brown, 1999, p. 5) is being fostered, thus realigning assessment with the more student centered curriculum. It is hoped classroom instruction involves language development through meaningful learning experiences. Portfolios aid a better understanding of the processes behind learning to write and therefore a better understanding of themselves and also each other.

Students will continue with self and peer assessments in level four and five. Here they will also take an active role in designing the evaluation frames by studying writing examples and considering what makes one better than another. Furthermore, in levels three to five, instead of using the self-report, students will be expected to write a reflective paragraph (level 3, 4) and essay (level 5) because as Weigle (2002, p. 215) states “reflection is one of the key elements of a portfolio and an essay is clearly the most direct way for evaluators to gain insights into student’s reflective processes”.

6 Authenticity

If the washback effects of portfolio assessment at UTB are improved, the validity of portfolios becomes even greater. Indirect tests of writing may have led to strange practices within the writing classroom, but portfolios require students to not only write more but write in more varied ways (Herman, 1992, p. 74). These effects contribute to portfolio assessment being considered not only valid but authentic. In an academic context authenticity is close to absolute in portfolio assessment because as Weigle (2002) explains:

In many writing programs, where many or all of the essays written in class are included in the final portfolio, the test tasks (portfolio contents) and the target language use tasks (classroom writing tasks) are virtually identical, which of course is the ultimate in authenticity (p. 203).

In the context of EAP, timed writing will be considered authentic as this is a requirement of undergraduate courses. Timed pieces can therefore also be included in the portfolio and in turn be reflected upon by the student. Furthermore, timed in class pieces can act as a guide to determine the authenticity of the author. Plagiarism is an issue that should be taken seriously by all educators. Because portfolio work may be completed at home, it is important that teachers are aware of in class standards. If differences between in and out of class work are suspicious, the teacher can conference with the student to establish if plagiarism has occurred and why. Martin suggests that “most cases should be dealt with as matters of etiquette rather than theft” (Martin, 1994, as cited in Owen, 2007, p. 216).

7 Reliability and Scoring Procedures

Like authenticity of authorship, reliability in portfolio assessment requires careful consideration. One of the greatest strengths of the portfolio is its uniqueness for each writer. However, this strength is also a weakness. “Intentional uniqueness is

thought to contribute to the high validity and washback potential of the procedure. However, the diversity of the products to be evaluated can create problems in terms of reliability” (Bailey, 1998, p. 218). Reliability with indirect tests of writing is high as the right or wrong answers allow for consistency in scoring. According to Callahan (1995, p. 128), holistic scoring techniques are currently the method of choice for scoring portfolios. However, the wider the range of contents the more complex the process becomes, and analytic scales may in these cases produce more reliable results (Weigle, 2002, p. 121). It may be easy to assess whether writing has a clear thesis statement. However, assessing humor or engaging quality is axiomatically more complex.

Therefore, planning for portfolio assessment needs to be systematic and carefully executed if it is to be successful and fair. It will be ensured that all those assessed and assessing are aware of what the goal posts are through regular conferencing and meetings. Teachers will be encouraged to regularly meet with each other to improve consistency. To aid this, negotiations are being made to provide a training day in the procedures and scoring of alternative assessments. Portfolios this autumn will be double marked and in the event of wide discrepancies a third opinion will be taken. This could present logistical issues of time. Class numbers in EAP range from six to twenty so teachers will have between twelve and forty portfolios to mark as each teacher has two classes. Therefore, although this approach will be adopted this autumn, it may need reconsidering come spring. Different analytic scoring frames will be used for each level to try and increase reliability.

In level one the analytic rubric will be scored out of twenty for each individual written piece making a total of 100 and a further 15 marks will be given for characteristics of the portfolio as a whole. Level one, as already discussed, is primarily focused on collecting work and familiarizing students with this process which is why the majority of marks in level one are assigned to the individual written pieces. The analytic criteria are set out below (Table 1).

Throughout the levels teachers will be able to increasingly focus interest on characteristics outside of the individual texts to assess the students’ ability to process write and apply themselves to different purposes and audiences. This will be reflected in the rubrics where increasing weight will be given to portfolio and writer characteristics. In level five multiple rubrics will be used to assess the portfolio as Gottlieb (1995, p. 13) states that “the heart of alternative assessment is anchored in rubrics that are aligned with specified tasks...[that]... serve as the yardstick for measurement...and standards of achievement”. The marking criterion for level five has three parts, individual texts, writer characteristics and portfolio contents. The latter two are set out (Table 2).

The reflective essay should provide background information to the reader about the choices made by the level five students and allow for a more knowledgeable reading; thereby they allow informed scoring to take place. As this is just the beginning of portfolio assessment at UTB, characteristics of the portfolio have been

Table 1 Level one grading criteria

Criteria for individual texts	Marks
<i>Assigned topic:</i> Write on the assigned topic using thoughtful details and appropriate vocabulary	5
<i>Sentences</i> Write 5–8 sentences	5
<i>Grammar</i> Accurate use of the assigned grammar (min. 4 times)	4
<i>Error correction</i> Errors from the previous week must be corrected and new sentences written	4
<i>Academic quality</i> Typed or written neatly Submitted on time	2
<i>Total for submission</i>	20
Criteria for portfolio	Marks
<i>Contents</i> The portfolio contains all the required texts including: contents page, rubrics and 5 texts	5
<i>Layout</i> The contents page is accurate to the portfolio layout and papers are organized	5
<i>Error correction</i> Each text (5) has been rewritten with errors corrected	5

limited to content and organization in order to provide more time for the marking of reflective practice as this will be new for teachers. In the same way students should not be introduced to everything at once, nor should staff. It was felt that the organization of the portfolio would also demonstrate reflective behavior as if the organization was logical it should complement the reflective essay, and again allow for a more informed reading of the portfolio itself. The three individual texts all have their own rubrics and focus on key aspects of the task. These rubrics are not considered final and will be edited and improved as EAP learns more about portfolio assessment.

8 What Happened Next?

The use of portfolio assessment continued at UTB until the Fall semester of the academic year 2013/2014. In general, the project received positive feedback from both students and teachers. It was especially popular with those students who stayed with the program for multiple levels as they had a very clear and satisfying record of their progress. The grading of portfolios required a large amount of collaboration

Table 2 Level five grading rubric

Characteristics of writer				
<i>Reflective essay</i>				
	1	2	3	4
Introduction of writing tasks for reader through reflective essay	Provides little information about assignments and writing process. Weak sense of audience and does not successfully orientate reader	Provides some information about assignments and the writing process. Shows some awareness of audience though this is mostly mechanical	Provides a useful discussion about assignments and writing process. On the whole the reader is orientated	Provides ample, rich discussion of assignments and writing process; clear sense of audience and purpose. Aware of writing as a way to do something beyond fulfilling assignments
<i>Evidence of the writing process being understood</i>				
	1	2	3	4
	Evidence is unorganized and shows little or no understanding of the writing process	Evidence is often unorganized and shows only basic understanding of the writing process	Evidence is on the whole organized and shows good understanding of the writing process	Evidence shows strong understanding of the writing process
Content of Portfolio				
<i>Organization and Completeness:</i>				
	1	2	3	4
A portfolio should include; contents page, reflective essay, 3 complete texts, including reading responded to, and additional work to show how the final version was achieved	3 or more parts are missing. Contents page has many irregularities. Content organized illogically	Two to three parts are missing. Contents page has some irregularities. Content organized illogically at times	Contains most of the required documentation but one part may be missing. Contents page is still accurate. Content organized in a mostly logical way	Contains all required documentation. Contents page is accurate. Content organized in a logical way

between teachers, but this was also a chance for teachers to share experiences and look for ways to improve the project. One of the key developments was the option for students to submit a digital portfolio. Pieces of writing were submitted either as word documents or as iPages.

Despite positive feedback from students and staff, alternative projects in both reading and writing and their associated assessments were suspended in the Fall semester of the academic year 2013/2014. Reasons include management changes, a belief that traditional test taking skills should be prioritized and the belief that alternative assessments are too time consuming for part time teachers. These are real concerns for students, teachers and management, especially when the tests act as a gateway or barrier to tertiary education. Financial costs must also be considered. However, with careful consideration the two projects, with their associated assessments, could be modified to meet time and cost restraints. To date, their numerous benefits have been sadly outweighed by these restraints. The projects represent a forward looking view of assessments in line with education theory and are certainly worth investing in.

9 Conclusion

UTB is attempting to meet the needs of its exam orientated students by implementing assessment practices that encourage students to acquire both language and learning skills. It is hoped that the learning skills will allow students to continue their language development more effectively after they have entered their academic programs with the relatively low level of IELTS and TOEFL required in the UAE. UTB is trying, through its curriculum, to gradually train students to be more self-reliant. According to Genessee and Upshur (1996, p. 99) “classrooms that use portfolio assessment are more student centered, collaborative and holistic”. However, to achieve these very desirable goals students and staff will need continued guidance in new practices and procedures, and both those assessed and assessing need to be fully aware of where the goal posts in standards are. Portfolio assessment is the first of several alternative assessments that will be introduced at UTB since the fuller adoption of a learner centered curriculum and the recognition that traditional tests alone were having a negative influence on its success. Only time will tell if the bridges UTB is attempting to build between language learning, instruction, evaluation and life learning skills will be successful.

Appendix 1: Assignment Continuum

<p>Writing task Continuum for Portfolio Assessment. For each level the writing genres to be included are listed, under these are bullet pointed suggestions of how to incorporate these genres into the syllabus. They do not have to be taught in the order listed, however some may need to come after other parts of the curriculum and this is indicated below *</p>				
Level 1	Level 2	Level 3	Level 4	Level 5
<p>In level one students are expected to write accurate sentences and not formal paragraphs so the emphasis should be on this. Here they are familiarizing themselves with writing in English. Word total is up to the student however 6 sentences should be the minimum.</p>	<p>In level two students will begin to write formal paragraphs however two of the genres remain informal. They will only use process writing of editing and evaluating (with evaluation frames) on the formal writing. Each paragraph/ email should contain a minimum of ten sentences.</p>	<p>In level three all writing is formal and should be planned written and edited, with the use of evaluation frames. Students will also peer edit some pieces. Each paragraph/email should have a minimum of ten sentences.</p>	<p>In level four students are expected to submit four formal pieces. They should edit at the sentence and paragraph level. Peer marking should take place with most pieces.</p>	<p>In level five students are expected to submit three formal pieces. One of these is an essay which should be 3 to 5 paragraphs long (30 to 50 sentences) . They should be planned, and edited at the sentence paragraph and essay level. The final essay submission should be chosen by the student with teacher support.</p>
<p>1.Description of a person:</p> <ul style="list-style-type: none"> • self, • classmate • family member • someone famous 	<p>1.Informal emails:</p> <ul style="list-style-type: none"> • What I've done recently • A thank you email 	<p>and 2.Formal emails:</p> <ul style="list-style-type: none"> • to ask for an extension • to explain absence from class • to present plan for paragraph 	<p>1.Formal emails:</p> <ul style="list-style-type: none"> • to apply for employment with a attached curriculum vitae (CV) 	

<p>2.Description of a process:</p> <ul style="list-style-type: none"> the water cycle how to make a dish how to pass an exam 	<p>2.Formal description of an object:</p> <ul style="list-style-type: none"> favourite item piece of technology equipment from the past 	<p>3.Formal description of a process:</p> <ul style="list-style-type: none"> how to learn a language how to change a tire how to make new friends at university 	<p>2.Description of tables and graphs:</p> <ul style="list-style-type: none"> dying languages provides tabulated and graphic data as does globalization 	<p>1.Description and analysis of table/graph:</p> <ul style="list-style-type: none"> clarity on line package provides a clear review students could find their own tabulated/graphical data to analyze
<p>3.Description of a place:</p> <ul style="list-style-type: none"> A tourist attraction in Dubai The UAE home country, The university 	<p>3.Writing about my opinion:</p> <ul style="list-style-type: none"> Women/men are better drivers Uniform should/should not exist at university Students should/should not fail for missing more than 15% of classes 	<p>4.Compare and contrast paragraph:</p> <ul style="list-style-type: none"> small / large car UAE / home country. Living at home / living in dorms 	<p>3.Opinion paragraph:</p> <ul style="list-style-type: none"> Globalization Gender issues Dying languages. 	<p>2.Essay:</p> <ul style="list-style-type: none"> Informative opinion
<p>4.Simple narrative:</p> <ul style="list-style-type: none"> What I did at the weekend A childhood memory My last holiday *after past tense lesson 	<p>4.Formal comparative paragraph:</p> <ul style="list-style-type: none"> Then and now (Dubai, travel, food, leisure) 		<p>4.Cause and effect paragraph:</p> <ul style="list-style-type: none"> Women in the workforce Alcohol in society 	
<p>5.A review:</p> <ul style="list-style-type: none"> of a restaurant of a video game of a film of a book from the extensive reading program 	<p>Describing graphs:</p> <ul style="list-style-type: none"> unit 7 Science of core text provides a good introduction to this skill 	<p>5.Summary of word text:</p> <ul style="list-style-type: none"> from text books from extensive reading program from a newspaper/ internet site 		<p>3.Critical response to reading:</p> <ul style="list-style-type: none"> from teachers' collections in resource cupboard any suitable newspaper/magazine article (please add to collection)

Appendix 2: Self-report Form for Level 2

Name.....

Answer these questions as honestly as you can about your portfolio.

1. Creating a portfolio has helped me understand how to write

A lot a little not at all

Can you explain your choice:.....

.....

.....

2. My favorite piece of writing in my portfolio is

my email my description of an object writing about my opinion

my comparison paragraph my description of a graph

Can you explain your choice:.....

.....

.....

3. My least favorite piece of writing in my portfolio is

my email my description of an object writing about my opinion

my comparison paragraph my description of a graph

Can you explain your choice:.....

.....

.....

References

- Arnold, J., & Brown, D. (1999). A map of the terrain. In J. Arnold (Ed.), *Affect in language learning* (pp. 1–24). Cambridge: Cambridge University Press.
- Bailey, K. (1998). *Learning about language assessment*. New York: Heinle and Heinle Publishers.
- Callahan, S. (1995). Portfolio expectations: Possibilities and limits. *Assessing Writing*, 2(2), 117–151.
- Chamot, A. U. (1995). Learning strategies and listening comprehension. In D. Mendelsohn & J. Rubin (Eds.), *A guide for the teaching of second language listening* (pp. 13–30). San Diego, California: Dominic Press.
- Donato, R. & McCormick, D. (1994). A sociocultural perspective on language learning strategies: The role of mediation. *The Modern Language Journal*, 78(iv), 453–464.
- Ellili, M., & Chaffin, E. (2007). Emirati students' readiness for autonomous language learning. In A. Jendli, S. Troudi, & C. Coombe (Eds.), *The power of language: Perspectives from Arabia* (pp. 247–260). Dubai: TESOL Arabia.
- Fredriksen, J., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18, 27–32.
- Genesse, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Gottlieb, M. (1995). Nurturing student learning through portfolios. *TESOL Journal*, 5(1), 12–14.
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice theory and research*. Cresskill, NJ: Hampton Press.
- Herman, J. (1992). What research tells us about good assessment. *Educational Leadership*, 74–78.
- Kohonen, V. (1999). Authentic assessment in affective foreign language education. In J. Arnold (Ed.), *Affect in language learning* (pp. 279–294). Cambridge: Cambridge University Press.
- Manglesdorf, K. (1992). Peer reviews in the ESL composition classroom: What do students think? *ELT Journal*, 46(3), 274–284.
- Martin, B. (1994). Plagiarism: A misplaced emphasis. *Journal of Information Ethics*, 3(2), 36–47.
- Owen, H. (2007). ESL students: Fostering skills to avoid plagiarism. In A. Jendli, S. Troudi, & C. Coombe (Eds.), *The power of language: Perspectives from Arabia* (pp. 247–260). Dubai: TESOL Arabia.
- Richardson, A. (2007). Appropriate methodology for project based learning in ESP. In A. Jendli, S. Troudi, & C. Coombe (Eds.), *The power of language: Perspectives from Arabia* (pp. 247–260). Dubai: TESOL Arabia.
- Richardson, A. (2006). Learning how to question. In M. S. Lahlou & A. Richardson (Eds.), *English for specific purposes in the Arab world* (pp. 111–122). Dubai: TESOL: Arabia.
- Rinvoluceri, M. (1999). The humanistic exercise. In J. Arnold (Ed.), *Affect in language learning* (pp. 194–210). Cambridge: Cambridge University Press.
- Rizk, S. (2006). ESP for business students: A new approach. In M. S. Lahlou & A. Richardson (Eds.), *English for specific purposes in the Arab world* (pp. 95–110). Dubai: TESOL Arabia.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25, 553–582.
- Shohamy, E. (2001). *The power of tests*. Essex: Longman.
- Short, D. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 27(4), 627–656.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.

An Essential Tool for Continuous Assessment: The Learning Portfolio

Esra Gun Alayafi and Pinar Gunduz

Abstract This chapter discusses the rationale for the adoption of the learning portfolio (LP) as a tool for improving instruction, providing a means of continuous assessment, providing structured and systematic feedback to learners, and keeping track of their progress throughout a course. The chapter then describes the implementation stages of the LP at Sabancı University School of Languages (SL) in Turkey and presents an assessment of the current practices and procedures. Finally, future goals are proposed in light of the collected feedback.

Keywords Learning portfolios · Continuous assessment · Turkey · Learner development · Process writing

1 Introduction

According to Paulson, Paulson, and Meyer (1991), a portfolio is a selection of learner work that demonstrates the learners' effort, progress and achievement in a range of areas. In broad terms, portfolios may include writing and speaking tasks, mini projects and learner development tasks. They are ideally compiled over a period of time in order to better represent learners' development. As Trevitt, Stocks, and Quinlan (2011) put forward, portfolios differ from any other 'products' of learning in that they document 'process' rather than just 'product'. They also demonstrate students' effort and progress over the duration of a course and therefore better represent learning outcomes. Because portfolios provide us with an overall picture of students' work, they "rescue us from the contradiction in many of the paradoxes or binary oppositions that lie at the heart of good learning and teaching" (Elbow, 1994, p. 40).

E.G. Alayafi (✉) · P. Gunduz (✉)
Sabancı University, Istanbul, Turkey
e-mail: esragun@sabanciuniv.edu

P. Gunduz
e-mail: pgunduz@sabanciuniv.edu

In addition, as Paulson et al. (1991) state, the portfolio acts as a bridge between teaching and assessment because it provides ample opportunities for teachers to receive feedback on the development of learners and to do remedial work. If portfolio tasks are done regularly and matched to course content and objectives, teachers can easily see their students' strengths and areas for further development. Thus, teachers can cater for their students' needs in a better and more individualized way. According to Huot (2002) and Klenowski (2002), portfolio assessment has become popular because it can serve different purposes at the classroom level. One of these purposes is that it aligns teaching and assessment so as to facilitate productive learning. Portfolios demonstrate students' cognitive and linguistic abilities in depth and show the progress that they make over a period of time with regard to course objectives. Therefore, with the portfolio approach, assessment becomes more closely linked to teaching practices with the curriculum as the driving force of teaching and learning processes.

The portfolio approach to assessment makes it possible to "break out of the 'assessment mindset' that has so long whispered in our ear 'You can only measure what is easily measurable'" (Elbow, 1994, p. 42). However, learning is a complex process and assessment should reflect this. In that sense, portfolios are a better representation of the complexity and individuality of the learning process because they better reflect students' actual abilities through a wide range of tasks, multiple drafts, and other aspects focusing on self-assessment and learner development. Therefore, portfolio evaluation helps us address the *real* assessment issues: "What do we really want in successful students?", "What are we trying to produce?" (Elbow, 1994). What's more, portfolios are different from traditional assessment methods in that they "encourage a focus on the importance of discovery, experimentation" (Huot, 1994, p. 325).

Taking all these benefits into consideration and the possibility that it would have positive backwash on everyday teaching and learning, the Learning Portfolio (LP) was implemented in the SL, and has now become an essential part of our program.

2 The Teaching Context and Rationale

The medium of instruction at Sabancı University is English. All undergraduates are required to take the Sabancı University English proficiency exam or bring an equivalent internationally recognized exam score in order to begin studying their major. Otherwise, they enroll in the School of Languages (SL). The SL has an intake of about 700 students a year at different proficiency levels, from zero-beginners to upper-intermediate students. The SL aims to provide students with the necessary foundation skills and knowledge to excel in their interdisciplinary academic studies. In addition to helping students develop their language awareness, knowledge and skills in English, it also helps them develop critical and

creative thinking through the provision of high quality instruction and the promotion of independent study.

The SL instructors, task groups and the director are in contact with professors teaching the faculty courses in Sabancı University with the aim of conducting research in order to understand their changing needs. As well as interviews with the professors, we also analyze their course books, attend some lectures, and analyze their assessment methods. Based on the expectations of various faculty courses, we try to align our curriculum objectives and assessment types to better prepare our learners for their future studies. To that end, recent needs analysis research revealed that more work is needed towards improving learners' writing abilities in the faculties, and thus it required prioritization over other academic competencies.

Considering the above principles and needs, our initial aim was to include a continuous assessment method for evaluating writing in particular. Previously, our assessment practices focused mainly on more traditional, summative means of assessment. However, the feedback we received from colleagues and learners over the years revealed that there was a need in our program for a more continuous type of assessment that focused on process rather than one final product, especially for writing.

From an assessment point of view, the portfolio approach provides teachers and institutions with a tool to evaluate student performance in a more authentic way. Rarely are we required to undertake a writing task under strict time limits. Writing is most often completed in our own time, using various resources if necessary, and is edited a few times until the writer is finally happy with the product. A timed essay, on the other hand, is the most typical form of assessing writing.

Doubtless, the timed essay has benefits over alternative forms of assessment. It is standard, easy to prepare and administer, and ensures every learner takes the exam under exactly the same conditions. Moreover, marking is relatively less subjective with the use of well-written criteria, blind grading, and multiple grading. However, a portfolio better represents students' actual abilities as it reflects students' performance over a longer period of time under a variety of conditions. Traditional product-oriented assessment methods focus on two aspects of test usefulness, namely reliability and practicality. On the other hand, the portfolio assessment covers other areas of test usefulness, which are "construct validity, authenticity, interactiveness, and impact" (Weigle, 2002, p. 175). Taking all these into consideration, it was clear that besides the summative writing exams that we already had in our program, there was a need for a more process-based assessment method to cater for our learners' needs in a better way.

Another priority was to encourage learners to approach their study systematically and to put more effort into their self-development as language learners. We aimed to help our learners become less dependent on their teachers, and to equip them with study habits that would prepare them for their future studies. One obvious benefit is that keeping a portfolio increases students' learning responsibilities (Barootchi & Keshavarz, 2002). For us, this was a good starting point to break inefficient and ineffective study habits such as memorization or procrastination that were predominant among our learners due to their previous study habits,

a lack of understanding of effective study strategies and methods, and cultural reasons. Our attempts to foster good learning habits such as talks, presentations and workshops helped only in a limited way in that they raised learners' awareness of such issues and strategies, but they were not instrumental in inspiring learners to change their habits. As a result, a portfolio task group was formed to look into possible ways of incorporating the portfolio into our current program.

3 The Implementation of the LP in the School of Languages

The Learning Portfolio (LP) Project at Sabancı University School of Languages emerged as part of the curriculum and assessment renewal based on a thorough needs analysis process—in particular on teacher and learner feedback. The feedback and the needs analysis process clearly showed our learners' need for a more systematic and more process-based approach to the teaching and testing of writing. Moreover, needs analysis research revealed that more work needed to be done towards learner development, and that the LP could address these concerns. The next section describes the implementation stages of the LP at Sabancı University School of Languages (SL).

4 Decisions Related to Content

4.1 Matching LP Content with the Curriculum

In a portfolio, tasks can be linked to specific curriculum objectives and learning outcomes. When this is the case, tasks can be “geared towards a relatively narrow target language use domain” (Weigle, 2002, p. 179). This enables the teacher to see more clearly the extent to which objectives have been achieved, and which objectives require remedial teaching, thus having a positive backwash effect on instruction.

Therefore, while making decisions regarding the content of the LP, we started by taking a detailed look at the curriculum, course objectives and course materials in order to create LP tasks that well suited the needs of our learners. After curriculum objectives and the desired learning outcomes were identified, these were matched with the tasks and materials already available in our course books. These materials were compiled in the form of a booklet to make the portfolio more organized and to enable it to be implemented and utilized regularly as part of the program. Students were given their portfolios at the beginning of the course. For each task in the LP, learning objectives and outcomes were outlined in detail for both teachers and students. A task checklist that reminded students of the specific requirements was

also added for each task in the LP. These requirements and course objectives were also clearly indicated in every LP task to make the task more meaningful (see Appendix 1). This was done with a view to allowing teachers to easily identify areas in which their students needed further development and to aid the design of lessons and learning materials accordingly.

4.2 Variety in Task Types and Conditions

For a more comprehensive representation of learners' abilities, our LP includes a variety of task types. We tried to incorporate tasks that had a good balance in terms of variety, style and requirements because as Weigle (2002) states "the writing ability is not a simple construct but involves numerous processes, and (that) a single writing sample written for a specific audience and purpose is extremely limited in its ability to represent the writer's ability to write for other situations, audiences, and purposes" (p. 186).

Initially, we focused on writing. We included tasks that were personalized in nature as well as academic text-based tasks. For example, students were asked to write learning diary entries (see Appendix 2). In other tasks, students were asked to formulate short answer responses based on an academic text that they had studied in class (see Appendix 3). This ensured that students had writing practice in different styles, genres, text types and length. This also gave the chance to students who prefer freer, more personalized writing tasks over academic writing to demonstrate their abilities.

As well as task types, we also gave importance to varying task conditions. The fact that we set the tasks under a variety of conditions made the assessment more authentic. This is because the majority of writing we do in the real world is not done under strict time constraints, and thus, the assessment of writing should also not "rely solely on in-class writing as evidence of writing ability" (Weigle, 2002, p. 185). This also enabled us to give students with different learning styles and preferences the opportunity to demonstrate their learning outcomes in a more suitable way, as traditional assessment types most generally put good test-takers at an advantage. The larger variety of student samples also allow both teachers and students to see to what extent learning objectives have been achieved, and which specific objectives require more attention.

4.3 Fostering Learner Development

The portfolio approach does not merely act as an assessment tool, but also helps students to become less dependent on their teachers. Since learners' reflection on their learning process is acknowledged as an essential component of education (Wolf & Reardon, 1996), we believed if tasks focused on various aspects of learner

development such as self-reflection and goal-setting as well as developing a specific skill such as writing, they could be used as effective tools to foster better study skills and eventually begin to improve autonomy in learners. In the portfolio, we envisaged that some tasks could ask learners to reflect on, assess and evaluate their own learning processes and thus require them to make conscious choices about their own learning. For example, Appendix 4 illustrates a personalized task that focuses on learning styles and strategies and self-reflection.

There are several different tasks that are designed for the purpose of self-reflection in the LP. One example of these are the 'Task Checklists' which consists of linguistic, stylistic, organizational and content-related requirements for successful completion of a task (see Appendix 5). Teachers encourage learners to analyze these checklists for guidance before completing a task and upon completion of a task to reflect on their own work. These checklists also serve another unique purpose which is for teachers to refer to while giving feedback to learners and evaluating the success of a task.

Another self-evaluation method that is used in the LP aims to encourage students to self reflect on their progress in a given period of time. In our case, these tasks are done twice a semester; midway and at the end of the course. Using these tasks, students have a critical look at their own work and identify their strengths and areas for development. Based on 'Can-Do' statements adapted from The Common European Framework of Reference (CEFR), students are asked to set their own goals to further develop their learning (please see Appendix 6). Following this, teachers evaluate the extent to which the learners have been able to prioritize their goals and learner action plans. Students' self-evaluation is also discussed in one-to-one tutorials. We believe in the importance of self-evaluation because introducing mutual responsibility between teachers and learners to carry out self-assessment results in improvements in instruction and learning, both through raising awareness regarding the quality of students' written or spoken work and also fostering in students a more goal orientation outlook in their studies (Fulcher, 2010).

Once a predetermined number of tasks are completed, students review their work up to that point, identifying areas that they need further work on and setting learning goals for themselves. They draw up an action plan specifying areas that require most attention and ways of achieving these goals (see Appendix 7). Students are then invited to attend tutorials with their teacher where they go over the goals and the action plans. To encourage students to engage in more in depth evaluation of their work and progress as well as coming up with tangible future targets, teachers invite students to discuss their choices during one-to-one tutorials. Teachers guide their students with questions to train them to think more critically of their work.

It is also possible to design portfolios in a way that caters for flexibility and choice on the learners' part, which are important aspects of learner autonomy. To this end, in our context, learners are asked to select from tasks they have completed to be evaluated on. They are required to choose and explain their performance on tasks they have chosen, based on given criteria. For example, they are asked to

choose the work that best represents their abilities or the one(s) they have shown most progress on. They document their selection and justifications in their portfolio in the designated section and also discuss it with their teachers in prescheduled tutorials. The process of task selection engages self-evaluation and enhances awareness of meta-cognitive processes, thus forcing students to make conscious choices about the quality of their work (see Appendix 8).

For such practices to be successful in creating less dependent behavior, learners need extensive scaffolded training and guidance from their teachers. It can only then bring about favorable learning habits including how to take control and responsibility of one's own learning.

5 Decisions Related to Assessment

Because a LP is a collection of work and thus contains many samples, it may not be realistic to double mark portfolios as we would normally do with standardized writing exams. What's more, blind marking is favorable in such exams to make the assessment of writing more reliable. However, such a practice would contradict the nature of the portfolio since the focus is not only on evaluating one single product but on the process and development. As a consequence of this, subjectivity may increase. In addition, in order to create a positive backwash on students' learning and development as learners in the SL, we wanted to also assess students' meta-cognitive processes such as goal-setting and self-reflection within the LP. However, such processes are not tangible and hard to assess fairly as they are based on personal judgments.

For these reasons, we have taken some measures to increase the reliability and consistency of scoring. Most important of these measures are writing clear specifications and guidelines, supplying task previews, training the graders and conducting standardization sessions, and having a clear set of criteria.

5.1 *Specifications and Guidelines*

Marking should be carried out with high standards, and marking procedures need to be consistent to yield reliable scores. Clear specifications and guidelines not only help maintain marking across different levels and sections in the same institution, but also make it possible to be more consistent in developing tasks and maintaining institutional standards (Weigle, 2002). Although certain elements of the portfolio need to be flexible, some aspects need to be standardized to achieve consistency. For example, in our guidelines, we saw the need to specify the number of tasks that need to be set as in-class or outside class work, individualized versus collaborative tasks, or the number of free or academic, text-based tasks.

5.2 *Task Previews*

To make the assessment of the LP as reliable and consistent as possible, under the supervision of a level assessor, teachers preview tasks in order to clarify expectations for task fulfilment before a task is set. We also standardize expectations by going over task checklists (see Appendix 5). Teachers brainstorm possible ideas students can include in their responses as well as different ways they could organize their work so that students are evaluated fairly afterwards. This also ensures that different teachers provide similar types of guidance to students when they set the task.

5.3 *Standardization*

Rater-training and standardization sessions are also an indispensable part of LP evaluation to ensure reliable grading. These sessions take place before each LP evaluation is carried out. It is the level assessor's responsibility to choose some samples that reflect various ability levels. The teachers read and discuss their views on the samples with regard to the criteria. As well as discussing the main strengths and areas for improvement, teachers also discuss and agree on the grade a specific sample would get. The level assessor supervises the process and assists with emerging issues.

5.4 *Criteria*

Another way we address the issue of grader subjectivity is through an easy-to-use set of criteria that balances less tangible components of the LP with concrete and evidence-based aspects. For instance, while the number of tasks a student completes or a student's attendance in portfolio tutorials could be considered concrete and easy to measure aspects of portfolio evaluation, development of learning in response to feedback or identification of strengths and weaknesses are more subjective aspects of the criteria we use in the SL.

A clear set of criteria is indispensable for reliable and consistent grading. The teachers are required to familiarize themselves with the criteria and attend standardization sessions where we grade sample portfolios using the criteria. This is crucial to ensure consistency and inter-rater reliability. Maintaining high standards in grading procedures and ensuring reliability through the use of clear criteria is "especially important in language programs that have several proficiency levels, as it reduces the likelihood that students will be promoted or held back in error" (Weigle, 2002, p. 183). Teachers also make use of the criteria while giving written or oral feedback to their students.

In the SL portfolio, the criteria are provided to the students in the portfolio booklet, and teachers use this page in the portfolio when they are grading the portfolio. This way the scoring criteria “becomes a teaching tool as well as a testing tool” (Weigle, 2002, p. 182). Social aspects of writing can also be encouraged and can be incorporated in the criteria. For example, teachers can evaluate to what extent a student has incorporated feedback on their work (Weigle, 2002). In our criteria, this is evaluated as a separate band, and students receive a score for the degree of progress they have shown in response to their tutor’s feedback. Therefore, they are held accountable for incorporating feedback.

5.5 Avoiding Plagiarism

One risk with portfolios is related to task conditions. Since many tasks are not carried out under test conditions, there is an increased risk that some students may be tempted to get assistance from others while completing their tasks. If the portfolio tasks are recycled in time, there is also the risk that students may get the portfolios from students who were previously enrolled in the program and thus plagiarize. In our case, most of the portfolio tasks are either newly created or extensively revised to cater for the needs of the new student group. However, there are also tasks that are recycled, and the number of students is too large to spot cases of plagiarism if precautions are not taken. To avoid issues of assistance and plagiarism, we follow several guidelines.

First of all, students are provided with written guidelines in their portfolios explaining expectations. There is a statement of academic integrity which students need to sign to show they have read and understood the statement and that the work in their portfolio is their own and completed without any assistance. This helps us to make our expectations clear from the very beginning of the course. The second measure is the use of plagiarism detection software for longer pieces of work. Students upload their work onto our online learning platform, which then compares the written work against other students’, the Internet, and the original text if it is a text-based task. It then detects if and to what extent a student’s work was plagiarized. Finally, we have documented a set of guidelines which clearly indicate what teachers need to do if they suspect a student has received assistance or plagiarized.

6 Provision of Information to Students

Before the portfolio was implemented general guidelines were produced for both teachers and students. Teacher guidelines include information related to task submission. In this part, the details about task setting procedures and plagiarism detection procedures are documented. Feedback and evaluation principles and guidelines are another important part in the guidelines document. These give

information about what to pay attention to while giving written oral feedback for individual tasks or after portfolio evaluation; and how to grade portfolios. There is also a separate section on tutorials, as they are an integral part of the portfolio system. Types and frequency of tutorials are also specified.

The student guidelines include information about the rationale for the portfolio, important reminders, assessment related information, and the academic integrity information. These aim to make the expectations and requirements clear and transparent from the first day. There is also a page which shows the calendar of events such as when to submit a certain task, when they will have tutorials, and the evaluation dates.

7 An Assessment of Current Practices and Procedures

The feedback cycle on the LP started even before it was implemented. A specific project group consisting of curriculum and assessment group members had done the initial planning, and designed portfolios for every level. This group then shared their work with the teachers and asked for their opinions, comments and suggestions on the design, content and grading. After revisions were made, assessors focused on their own level, and collected feedback on a regular basis.

Teachers were asked to contribute their ideas, comments and suggestions regularly during forums and meetings such as task previews before a task was set, during standardization sessions after students submitted their tasks, and before LP evaluations on specific tasks. Teachers were also sent detailed surveys in the middle and at the end of the course (see Appendix 9).

While surveys focused on the general views, perceptions and attitudes towards different components and design of the LP, the forums and meetings concentrated specifically on either tasks or grading procedures. This provided us with a macro perspective in that it helped us to gather feedback both on how the LP fits with the rest of the program and how it is viewed in general. Additionally, it enabled us to view the LP from a micro perspective and get feedback on specific details in a systematic way.

In the SL, learner feedback is as important as teacher feedback and is taken into serious consideration by graders. This was also the case with the LP since if the students did not appreciate and understand the value of the portfolio, then it would not have the desired effect. Therefore, in addition to teacher surveys, students were also given questionnaires twice a course. Also, every level held learner forum meetings with representatives from each class. At these forums, students' perceptions of the portfolio were discussed in addition to other items related to course content.

The gathered feedback showed that the students appreciated the value of the LP and thought of it as one of the most useful learning tools in the system. They also believed that the weighting of the portfolio towards their overall course grade could be increased because they believed it is a good representation of their actual performance (see Fig. 1).

Teacher 1: "I think the re-writing focus is very important. This is really where the students can show their understanding of the feedback and their growth in writing ability. Tutorials for portfolio feedback were the most productive element of the whole course for my students, I felt."

Teacher 2: "Everything about the portfolio went great. Of course there is room for improvement as with anything else, but overall, the students took it seriously, and did their best to follow the deadlines and produce high quality work."

Teacher 3: "I like the booklet idea because it gives the students a focus, and ensures that teachers mark and give feedback regularly. And despite the modest weighting, washback so far has generally been positive in that students on the whole continue to take it seriously".

Teacher 4: "Learning portfolio in terms of continuous assessment was good, the practices of continuous assessment could be even more."

Teacher 5: "I found criteria easy to use - once or twice there was a strange case but i think these were easily worked out - I like that it is quite quick to use- and students are writing much more systematically... I like that it isn't really a feeling of mini assessments but something in between... they are learning a lot from it".

Teacher 6: " I think it has been very beneficial - it systematizes our approach to writing without being an over-standardized straight-jacket - it's enough of a carrot to get students doing more writing and to ensure all teachers are doing it:)"

Student 1: "I usually don't take the initiative to write something, but the LP inspires me to write."

Student 2: "We needed a lot of practice to improve our writing and the assignments in the LP helped especially improve our writing skills. So we were very happy with the portfolio."

Student 3: "It is the best element of the SL program. The weighting should definitely be increased."

Fig. 1 Extracts from teacher and student feedback

Especially in the first year that we started to implement the portfolio, it was very important for us to collect feedback from both teachers and students to understand if the portfolio served its purpose and in what ways we could further improve it. Both teacher and student feedback from all levels indicated a positive perception of the portfolio. Most of the feedback was quite encouraging and in line with our primary goal of incorporating a continuous assessment method into our program that would bridge instruction and assessment while fostering learner development and enhancing learners' language competencies (see Fig. 1).

One of the questions in the student survey focused on writing: *“To what extent do you agree with the following statement: “The Portfolio has helped me in improving my writing skills”*. Of the 92 students who responded to the survey, 94.15 % agreed with the above statement. This proved that learners believed that keeping a portfolio was an essential tool in their learning process.

8 Future Goals

The feedback we received from teachers and students also provided us with ideas and suggestions on how to further improve the LP and what amendments can possibly be made to the current design and content.

9 Continuous Speaking Assessment

One of the most commonly raised suggestions from both teachers and learners was the need to make the assessment of speaking continuous and more process-oriented when compared to the traditional oral exam. Students suggested that such an approach would provide much needed relief for the exam-anxiety they have been experiencing and encourage them to place more importance on speaking on a daily basis rather than studying towards it before the exam. As for the teachers, it was suggested that making speaking part of the portfolio would give them the chance to evaluate their learners' speaking competencies through a variety of tasks, enabling them to focus on a range of speaking sub-skills.

10 Allowing for Peer Assessment

In our current system we get the students to self-assess, and set their own learning goals as well as draw an action plan. However, teachers also suggested incorporating peer assessment. Such practices are considered highly effective and

informative among formative assessment methods. As Brown states (2004, p. 276), “self-peer-assessment are among the best possible formative types of assessment and possibly the most rewarding”. Possible tasks that we are considering for peer assessment are filling out peer checklists and questionnaires, rating someone’s oral presentation holistically, peer editing and peer proof reading. Peer assessment is also important for acquiring meta-cognitive awareness in learning. This is because being able to judge to what extent given criteria have been achieved is the initial step towards becoming able to produce high quality output.

Based on the feedback we received, we are also considering expanding our task variety to include more styles and genres—specifically creative writing. For instance, we are thinking of getting students to write response papers, short stories at lower levels, and answers to document-based questions. In addition, we have started working on creating mini-projects for our learners in order to incorporate elements of project-based learning in our portfolio. This, we believe, will enhance task variety as well as increase the amount of collaborative work. Such additions and changes to the Learning Portfolio will truly supply evidence of students’ learning progress rather than only specific learning outcomes.

11 Conclusion

Having implemented the Learning Portfolio for over a year, we were able to observe several desired outcomes. For instance, there has been an improvement in learners’ commitment to and enthusiasm for developing their skills. We have also witnessed that learner responsibility and awareness towards their language learning in terms of the attendance rates for tutorials, response to feedback, their ability to self-reflect and set goals for their own learning have improved significantly. As the research results also indicate, students appreciated the value and positive effect of keeping a learning portfolio on their learning in general.

Although at times it was challenging for teachers to keep up with the demands of the portfolio in terms of arranging time to give written and oral feedback, the teachers also embraced the Learning Portfolio as an indispensable component of our assessment scheme. As also stated in the research results and the future goals section of our chapter, they would like to make the learning portfolio an even more inclusive tool that could best represent students’ performance.

Appendix 1: Matching Tasks with the Curriculum

TOPICS	OBJECTIVES
Displaying familiarity with: The stages of cloning	WA.4. Describing the process of cloning

After studying Unit 9 Output 1 “How is it done” from Beyond the Boundaries Level 1 Book Two

Write a paragraph describing the process of cloning. (70-90 words)

Appendix 3: A Sample Short Answer Response

After studying Unit 9 Input 1 “Organ Transplantation” from Beyond the Boundaries Level 1 Book Two

There are long waiting lists for donor organs. What are two possible solutions to this problem?

(70–90 words)

Appendix 4: A Sample Learner Development Task

After studying Unit 1 Input 3 “Student Types” from Beyond the Boundaries Level 1 Book One

What type of a student are you (logical, intuitive or independent)?
What are the strengths and weaknesses of the type you chose? Write a description of yourself. (150-160 words)

Appendix 5: A Sample Task Checklist

After studying Unit 9 Output 1 “How is it done” from Beyond the Boundaries Level 1 Book Two

Write a paragraph describing the process of cloning. (70-90 words)

Now check your answer by answering the questions in the first table only.

Task checklist - TO BE FILLED IN BY STUDENTS	
Use of Language – Have I used the target language correctly? (present simple passive)	<input type="checkbox"/>
Task Fulfilment – Is my answer complete and accurate according to the text in the book?	<input type="checkbox"/>
Rhetorical Pattern – Have I used ‘sequencing phrases (e.g. to begin with, next, lastly)’?	<input type="checkbox"/>

Task evaluation – TO BE FILLED IN BY TUTORS			
	Yes	Partial	No
Signature			
Comments:			

Appendix 6: A Sample Self-reflection Task

These “can do” statements are to encourage you to reflect on your own language ability and assess your progress throughout the course. If you have a greater awareness of your own language learning, it will help you to focus more clearly on areas of your English to develop.

When I complete *Route 2*, my level of language will be approximately A2+ on the *Common European Framework of Reference for Languages*. This means:

I will have enough basic language to deal with everyday classroom situations.

I can give short descriptions and tell other people information on topics about my studies.

√√ I can do this well √ I need more practice of this X I can't do this

I can understand and follow the process of answering a writing prompt.	
I can understand the question and respond appropriately.	
I can make the topic of a paragraph clear.	
I can use some linkers to connect my opinions.	
I can use an appropriate text pattern(s) to answer the question.	
I can define and explain simple terms and concepts.	
I can make comparisons and contrasts between objects and concepts.	
I can describe the causes and effects of ideas and concepts.	
I can give a description of objects, people, places and situations.	
I can write in an academic style.	
My language is mostly grammatically correct.	
My spelling and punctuation is mostly accurate.	
I can use a variety of vocabulary.	
I can write in an objective and impersonal style.	
I can rewrite parts of a text using my own words.	
I can improve my writing if I...	
1. _____	
2. _____	
3. _____	

Tutor's comments:

Appendix 7: A Sample Self-reflection and Goal-setting Task

Student self-reflection & goal setting

	Example tasks
My strengths in writing are • •	_____ & _____
In the last month, I developed most in	_____ & _____
When I look at my earlier work I see	_____ & _____

I would like to learn more about

-
-
-

I can do this if I

-
-
-

Tutor's comments:

Appendix 8: A Sample Task Selection Task

SELECTION OF TASKS FOR EVALUATION #1

Please pay attention to your tutor's comments in each task before you make your selection. In this task you can mention all or some of the points below:

- *Task fulfilment*
- *Use of language*
- *Development / explanation of ideas*
- *Organisation and linking of ideas*

Please choose **two** tasks that best reflect your development in the first part of the course.

The first task I chose for my second LP evaluation is _____
because _____

The second task I chose for my second LP evaluation is _____
because _____

Tutor's comments:

Appendix 9: Teacher Survey Questions

	Very well	Quite well	Not very well	Not at all well
1. How well was the writing material in Beyond the Boundaries in the tasks in the Learning Portfolio?				
2. How satisfied were you with the following aspects of the criteria for the evaluation of the Learning Portfolio?	Very well	Quite well	Not very well	Not at all well
a. Ease of use				
b. Clarity				
c. Provision of feedback to students				
3. How satisfied were you with the following aspects of the implementation of the Learning Portfolio?	Very well	Quite well	Not very well	Not at all well
a. Previewing of tasks in route meetings				
b. Timing of tasks on the calendar				
c. The amount of guidance provided to teachers				
4. What were the strengths of the portfolio?				
5. How did the portfolio tutorials go?				
6. What are the areas of improvement for the LP? You can consider tasks, format or the grading documents or procedures.				
7. Have you got any other suggestions or comments about the Learning Portfolio?				

References

- Bag, M., Dobie, G., Einer, J., Ferris, D., Garinger, D., Mustafa, A., & Yanova, J. M. (2003a). *Beyond the boundaries—English in an academic environment—Level One book one*. Sabanci University Press.
- Bag, M., Dobie, G., Einer, J., Ferris, D., Garinger, D., Mustafa, A., & Yanova, J. M. (2003b). *Beyond the boundaries—English in an academic environment—Level One book two*. Sabanci University Press.
- Barootchi, N., & Keshavarz, M. H. (2002). Assessment of achievement through portfolios and teacher-made tests. *Educational Research, 44*(39), 279–288.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Elbow, P. (1994). Will the virtues of portfolios blind us to their potential dangers? In L. Black, D. A. Daiker, J. Sommers, & G. Stygall (Eds.), *New directions in portfolio assessment: Reflective practice, critical theory, and largescale scoring* (pp. 40–55). Portsmouth, NH: Boynton/Cook.
- Fulcher, G. (2010). *Practical language testing*. London: Hodder Education.
- Huot, Brian. (1994). Beyond the classroom: Using portfolios to assess writing. In L. Black, D. A. Daiker, J. Sommers, & G. Stygall (Eds.), *New directions in portfolio assessment: Reflective practice, critical theory; and large-scale scoring* (pp. 325–334). Portsmouth, NH: Boynton/Cook.
- Huot, B. (2002). *(Re)-articulating writing assessment for teaching and learning*. Logan, UT: Utah State University Press.
- Klenowski, V. (2002). *Developing portfolios for learning and assessment: Processes and principles*. London: Routledge Falmer.
- Paulson, F. L., Paulson, P. R., & Meyer, C. (1991). What makes a portfolio a portfolio? *Educational Leadership, 48*(5), 60–63.
- Trevitt, C., Stocks, C., & Quinlan, K. M. (2011). Advancing assessment practice in continuing professional learning: Toward a richer understanding of teaching portfolios for learning and assessment. *International Journal for Academic Development, 17*(2), 163–175.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wolf, D. P., & Reardon, S. F. (1996). Access to excellence through new forms of student assessment. In D. P. Wolf & J. B. Baron (Eds.), *Performance-based student assessment: Challenges and possibilities. Ninety-fifth yearbook of the national society for the study of education, part I* (pp. 1–31). Chicago: University of Chicago Press.

Believing in the Power of the Child: Reggio Recognizing the Affective

Nayer Iqbal Ali Chandella

Abstract In today's globalized world, educators face complex challenges. These challenges demand educators to be more outward-looking. We must keep in mind growing intolerance among communities. When educators rely on predetermined sets of 'competencies' to be acquired through rote memorization, they deny children opportunities to develop critical thinking and creativity. Such practices deprive children opportunities to learn in a collaborative, reflective manner. In examination systems that contribute to a strong desire for grades, real objectives of education are seldom achieved. Assessment for learning encompasses all factors influencing a student's learning. Assessment is more than test results quantified. Good assessment practices must collect student's data and make sense of it and use it to change teaching and learning in a positive direction. There is a need for innovations in curriculum, enabling all children to become successful learners and confident individuals. My research explores whether schools can become centres of curricular innovation. My paper presents a small scale research project based upon the Reggio philosophy. My narrative describes the Reggio approach to early childhood education and explores aspects of this approach adapted to the Pakistani context. I particularly focus on portfolios and documentation as means of assessment.

Keywords Social constructivism · Reggio · Recognizing the affective · Curriculum of curiosity · Documentation and portfolios as assessment tools

1 Introduction

When a child's capacity for participation is underestimated, their agency in their own lives is denied and the value of involving them goes unrealized. We need a community where there is collaboration between parents, teachers, and students, a school setting where a child's interests are recognized and learning stimulates

N.I.A. Chandella (✉)

Lahore College for Women University, Lahore, Pakistan

e-mail: Chandella5@yahoo.com; nic201@alumni.exeter.ac.uk

© Springer International Publishing Switzerland 2017

R. Al-Mahrooqi et al. (eds.), *Revisiting EFL Assessment*,

Second Language Learning and Teaching, DOI 10.1007/978-3-319-32601-6_16

curiosity, where teacher and students are co-explorers. This community is possible and does exist. The philosophy known as Reggio Emilia was conceived shortly after World War II, when Loris Malaguzzi decided to rebuild the war-ravaged school system. Literacy was the area of focus; however, writing and reading were embedded in social and cultural contexts.

Carl Rogers' classic text 'Freedom to Learn' (1983) helped me to refine and understand a different approach to learning. In 'Freedom to Learn' Carl offers a lucid account of this approach to learning and the wonderful world it makes possible:

I want to talk about learning. But not the sterile, quickly forgotten stuff that is crammed into the mind of the helpless individual tied into his seat by bonds of conformity! I want to talk about learning - the curiosity that drives the child to absorb everything, the student who says, "I am discovering, drawing 'in' from the 'outside', and making that a real part of me." The learning in which s/he exclaims: "No, no, that's not what I want; "Wait! I am interested in this; Ah! Here it is what I want to know!" (pp. 18-19).

I saw and experienced the world painted by Rogers (1983); I experienced this at a junior branch of a private school system in Pakistan (popularly known as TNS). It is a private school that has chosen to implement the Emilio Romagna approach to learning.

The Reggio philosophy, named after the region in Italy where it started after World War II, aims at pre-school and primary students. It is a self-directed, holistic experience for youngsters utilizing the support and guidance of community. It is based on a belief that children learn best when they have some control over their learning. I was drawn to the philosophy because it focuses on the child as a partner in the learning process. Children can learn not only from the teacher but also from the environment and from each other. While reflecting on an amiable, collaborative, and active system of educating young people, I wondered: do we truly value the child as resourceful and competent? And can we promote learning as a reciprocal process resulting in growth?

My interest in the Reggio philosophy developed during a visit to an art exhibit which displayed the works of toddlers in the Junior Section of the school. Amazed at the fine work of children, I wanted to investigate whether assessment for learning that encompasses all factors influencing a student's learning is possible? I wanted to know—why did TNS choose to adopt the Reggio approach to instruction/assessment, and what were the challenges they faced?

There is a widespread and mistaken view that the Reggio approach is incompatible with assessment practices. My paper discusses how researchers can adopt new methods to assess children for true learning.

Thus my narrative describes the Reggio approach to early childhood education and explores aspects of this approach adapted to the Pakistani context. I particularly focus on assessment through portfolio and documentation. My chapter is a narrative of the application of a constructivists' approach to learning. I focus on Learning not as an event but as a process: a continual growth. The emphasis is on the affective domain of learning that deals with our attitudes, values, and emotions. My

conclusions are based upon the analysis of data collected through both observation, and teacher's journals to elaborate on the challenges faced by the teachers.

2 Accepting the Challenge of Change

In a globalized world, educators are faced with increasingly complex challenges that need to be considered. We are concerned about growing intolerance among communities. What is depressing is the sad reality that the spread of education has made no serious difference. Our schools rely on predetermined sets of 'competencies' to be acquired through rote memorization, thereby denying children creative and critical thinking skills. Creating an encouraging environment for children necessitates profound change. Our classrooms deprive children of the opportunities to learn in a collaborative, reflective manner. It seems we have lost sight of the real objectives of education.

A curriculum establishes a vision of the kind of society we want. There is a need for innovations in teaching practices to enable all children to become successful learners and confident individuals. Reforms are needed in changing times. Educational policy documents in nations such as Britain, Pakistan, Bangladesh and Singapore share similar aims. Curriculum and teaching innovations in Pakistan are part of that global trend. Its aim is to make children uniquely enabled to thrive as participative citizens.

In most countries the capacities of children for participation are underestimated, their agency in their own lives is denied. We need a community where there is collaboration between parents, teachers, and students, a school setting where a child's interests are recognized and learning stimulates curiosity, where teacher and students are co-explorers. This community is possible and does exist. As stated previously, the philosophy known as Reggio Emilia was conceived shortly after World War II, when Loris Malaguzzi decided to rebuild the war-ravaged school system. The community responded enthusiastically. The city established an educational system where teachers and children enacted a highly integrated curriculum. The model follows socio-romantic ideology, and is based on a humanistic, child-centred curriculum. Although literacy was an area of focus, writing, reading and assessment were embedded in meaningful cultural contexts.

3 Theoretical Foundations

According to constructivists, students are more than passive storehouses of fed information; instead, they come to school with values, histories, and perceptions. Therefore, meaning is interpreted and constructed by learners in the process of exchange with the source (Stevenson, 1997). Thus, an officially sanctioned curriculum can have multiple readings in negotiated ways (Luke, 2002).

A constructivist framework challenges teachers to create environments where students are encouraged to think and explore (Brooks, 1993). A constructivist learning environment is a place where learners support each other by using various tools and resources in pursuit of learning (Wilson, 1996).

Reggian's is a socio-constructivist model, influenced by Vygotsky (1962) who states that children (and adults) co-construct their knowledge through the relationships that they build with other people and their environment. Ciari, Dewey and Piaget have contributed to various embodiments of the Reggio philosophy. It promotes an image of the child as a capable participant. A reciprocal relationship exists between teacher and child, and much attention is given to detailed observation and documentation. Collaboration is important because learning occurs in many places, not just at home or in school.

The foundational philosophy of the Reggio approach is based on the value of inquiry-based learning, and the existence of a child-centered curriculum. Education is a holistic endeavor of developing every aspect of a child: mind, body, emotion and social competence. Collectively and individually, every person and situation can be a potential learning opportunity. Thus two defining principles of this approach are:

- (i) Continued research and learning
- (ii) Emphasis upon the social and physical environment

Vygotsky (1962) stresses that learning occurs through interaction between adults and children. Adults, as more skillful partners, provide modelling, fostering development of both cognitive and social processes. The Reggio philosophy asserts that children also have the potential to contribute to the learning process by provoking questions and interpretations. Certain elements of Piaget's (1969) perspective on a child's cognitive development, namely constructive outlook on cognitive conflicts, are reflective of the philosophy. Similarly, the Reggiorians view errors as opportunities for learning.

Dewey's philosophy of 'progressive education' sees learning as mutual where both teachers and students interact and cooperate in the educational process (1997). Dewey emphasized the role of social interaction. Thus, Dewey's emphasis on a learning process characterized by transformation of understandings is valued by the Reggiorians.

Loris Malaguzzi's thinking reflects a social constructivism that draws from Dewey, Piaget, Vygotsky, and others. Focusing on preschool years, Malaguzzi (1993) drew a powerful image of the child, social from birth, intelligent and curious. His vision of an education is based on relationships and seeks to support reciprocal relationships among children, family, teachers, society, and the environment. Reggiorians consider cultural and physical environment as important, therefore each Reggio inspired school is different from others in many aspects (Fig. 1).

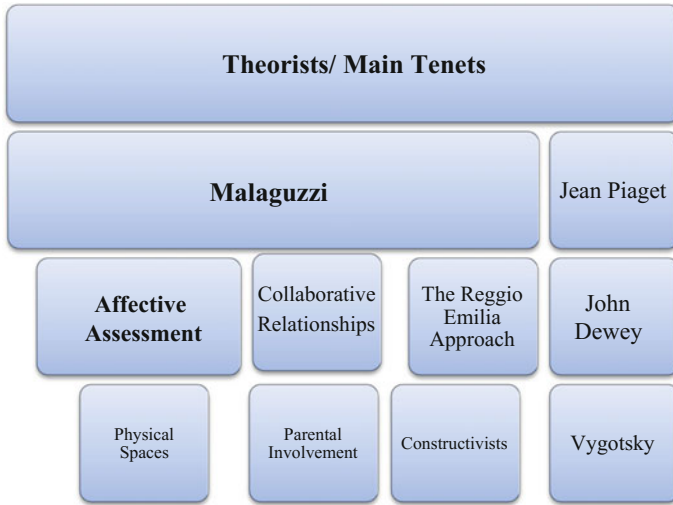


Fig. 1 Theorists influencing the Reggio approach to learning

4 Social Constructivism, Reggio: Recognizing the Affective

Malaguzzi (1993) pays importance to individual as well as group interests and thus supports a form of socio-constructivism. It is through a process of re-reading, reflection and revisiting that children are able to organize what they learn from their experiences. These processes are socially constructed and the child is an active constructor of knowledge. The teacher is a protagonist who engages in the same processes along with learners. Social constructivism is shown in respecting the child’s need to generate questions and revisit choices. Education is based on communication in teacher-child-parent relationships. This creates a very rich learning environment for all. In the Reggio Approach the teacher is a researcher and a co-creator of learning. This echoes the work by Vygotsky (1962), who thought it was necessary to look at the social environment with which individuals interact.

Human beings are not limited to their biological inheritance, as other species are, but are born into an environment that is shaped by generations. In this environment, they are surrounded by artefacts that carry the past into the present (Cole, 1996). In other words, to the biological inheritance is added cultural inheritance. This is linked to human development by Vygotsky (1978), who states that development is not simply a matter of biological inheritance; it is enriched through the individual’s cultural inheritance as well. Thus, a classroom should be a collaborative community that works towards shared goals and collaboration. Curriculum is a means not an end. If the aim is to engage students in productive activities that are personally and socially significant, ‘covering’ the curriculum should not be the ultimate goal of

education. Thus in a Reggio school a child is seen as a protagonist, a collaborator and a communicator. Teachers are involved in designing initial learning materials for exploration; teachers do not structure a curriculum that must be adhered to, rather, the curriculum is developed as the school session unfolds. Students' interests in particular topics help shape the activities and projects that make up their learning curriculum. This emergent curriculum provides multiple learning opportunities where students not only learn facts and skills, but are also encouraged to be inquisitive and explorative. Reggio educators believe that children have the capacity for representing ideas in a wide variety of symbolic and graphic modes—what Malaguzzi (1993) called the 'hundred languages of children'.

5 Romanticism: Undoing the book

In language education, a number of curricular developments have their roots in Romanticism. These developments celebrate imagination and the self. Peckham (1976) sums up Romanticism as: organic rather than mechanical, becoming rather than being, and relationships rather than entities. Wordsworth, in 'The Prelude', speaks of "*attending to a voice more profound than written language, a voice which fosters thought that was not the prisoner of books*. Books can uphold, and feed, and leave in quiet, or in the hands of the pedant, they can *work to derange, to subvert, to lay waste*" (Wordsworth, in Palgrave's Golden Treasury, 2002). In "Ode: Intimations of Immortality," Wordsworth celebrates the Platonic glory of the child-philosopher:

Thou best Philosopher, who yet dost keep
 Thy heritage, thou Eye among the blind,
 That, deaf and silent, read'st the eternal deep,
 ———— Mighty Prophet! Seer blest!

On whom those truths doth rest. ("Intimations of Immortality")

Romanticism emphasizes individualism, importance of imagination and glorification of human qualities. For Reggiorian educators it means having the 'courage to teach' rather than fulfilling the requirements of state curricula. However, finding the resources to do this is a struggle; nevertheless, a significant number of those who 'dare' is a beacon to follow.

6 The Reggio Approach: Living and Learning Communities

Reggio Emilia respects children as creative knowledge builders and leaders of their learning. It gives them the space, time and individual support to explore and learn from the world around them. The approach fosters creative enquiry and empowers children to become co-constructors of their learning. Participating adults are companions on the learning journey, not the leaders. The role of the adult is to document this journey, putting emphasis on the process of the explorations not the outcomes.

Teachers work with pupils to help them draw on their experiences, to discover things that excite them and make them curious. Pupils are supported to identify an idea that they would like to research further. The outcomes are communication and presentation of enquiries that take various forms, including video or written documents. Students are given a voice, a right to make decisions about things that affect them. The teachers use their professional expertise with a repertoire of skills that suit immediate requirements of the child. Educators in Reggio see themselves as a reference point, a part of engaging dialogue, fostering a strong and rich vision of the child (Edwards, Gandini, & Forman, 1998; Katz, 1998).

It is a concept in which each child's intellectual, emotional, and social potentials are cultivated. The principle involves youngsters in long-term engrossing projects, carried out in a love-filled setting. Reggio inspired teachers are seekers, observers, choreographers and mentors. They are guided by the urge to help children use their natural gifts and excitement to uncover the mysteries that lead to genuine 'learning'.

However, one of the most important drivers of learning is how that learning is assessed and it has been observed that students' attitudes towards their studies are strongly affected by the nature and timing of assessment. Assessment procedures not only measure learning outcomes but these also classify students' achievement. Well-timed and well-designed assessment can have a powerful impact on how students approach their learning.

There is a widespread and mistaken view that the Reggio approach is incompatible with assessments. It is limited in its interpretation to a standardized unit of measurement. Therefore, the questions that my research seeks to answer were, how can the Reggio approach, featuring children's construction of learning through expressive language be combined with assessments or demonstrations of children's learning according to defined standards? Some of these answers that build on a distinguishing feature of Reggio early education can be found in the approach called assessment for learning.

7 Assessment of Learning: Assessment for Learning

Assessment and evaluation are fundamental components of teaching and learning. Assessment is the process of collecting and documenting information on individual's learning, while evaluation is the process of analyzing, reflecting, summarizing and making decisions based on this information. The purpose of assessment is to inform teaching and improve learning. Hence, assessment is an integral part of the teaching and learning process.

Recent developments and demands of society have affected education. In particular, theories such as constructivism and multiple-intelligence demand change in traditional approaches to learning, teaching and assessment. For earlier academics, the learning of basic knowledge was very important. Knowledge was abstract, and conceived as the accumulation of stimulus-response association. Assessment was to be based on testing basic knowledge.

This traditional assessment approach mostly promotes students to memorize rules rather than conceptual understanding, and focus on small, discrete components of the learning domain (Dochy, 2001). These tests seldom provide information about genuine learning and are not enough to assess higher order cognitive skills such as problem solving, critical thinking and reasoning (Romberg, 1993). These do not measure a students' ability to organize relevant information (Shephard, 2009), and only assess easy to test-memorization of rote skills and procedures (Mumme, 1991).

Whereas the constructivists proclaim that students are able to acquire and socially construct their knowledge and understanding. Therefore, students' prior learning, their problem solving skills and their collaborative learning skills are important (Baki & Birgin, 2004). However, in this approach students' learning cannot be assessed within a short time using multiple choices tests (Mumme, 1991). Therefore, a broader range of assessment tools are needed to assess the students' skills such as open-ended problem solving, critical thinking and reasoning. For this reason, alternative assessment approaches are needed in assessing both learning process and learning outcomes.

Feminist pedagogy (Fiksdal, 2001; Maher & Tetreault, 1994; Ropers-Huilman, 1998) focuses on how grading or assessment points to the use of power in the classroom and how prior standards used in assessment are arbitrary or unfair. Assessment for learning is a recent trend in education which encompasses the broad nature of all factors influencing a student's learning.

Student's success in demonstrating what they know or are able to do varies. Their level of success depends on factors such as the time of day, situation, types of questions asked, familiarity with the content and willingness to perform at any one time. Children require time to demonstrate their achievements through learning opportunities that are appropriate and within the range of that which they can do independently. The rate and depth with which individuals will engage vary from beginning to end.

Learning is an active process. Therefore, assessing the process is critical and it should occur while the learning is happening rather than assessing the final product. Ongoing assessment informs the approach needed to design and deliver developmentally appropriate instructional activities. The best opportunities to assess learning occur in instructional encounters while students work individually and in groups. This implies that assessment techniques should focus on what students know as well as what they do not know. This can be achieved through alternative assessments, measuring students' performance and developments in the learning process.

Motivation and desire are foundations of learning. If students don't want to learn or are unable to learn, there will be no learning. According to Stiggins (2005), "Desire and motivation are not academic achievement characteristics. They are affective characteristics" (p. 199). Popham (2009) concurs that "Affective variables are often more significant than cognitive variables" (p. 230). That is why my chapter advocates an affective domain of learning.

8 The Affective Domain

The affective domain involves attitudes, interests, preferences, values, and emotions. Emotional stability, motivation and personality are examples of affective characteristics. Teachers must know who can work unsupervised and who cannot, who needs to be encouraged to speak in class, who is interested in science but not in social studies. Most teachers can describe their students' affective characteristics based on their informal observations and interactions with the students. But affective behaviours are rarely assessed formally in schools. Affective assessment measures learners' interests, values, inclinations, attitudes towards the relevance and importance of the content they are to learn and beliefs concerning their own ability to learn. According to Stiggins (2005):

We assess dispositions in the hope of finding positive attitudes, values, sense of academic self so we can take advantage of these and build on them, to promote greater achievement. And if our assessments reveal negative feelings, then we must strive for educational experiences that will result in the positive dispositions we hope for (p. 204).

Affective assessments target two primary contexts: teacher-led conversations and child directed activities. For example, the assessment of teacher-led conversations with children (whole or small group) consider questions such as: how can teachers facilitate the conversation? Do children and adults listen to each other? Is the purpose of the conversation to share what children already know or build new knowledge? Can children help each other by critiquing or explaining ideas to each other? The assessment of child-directed exploration considers the quality of the exploration: whether the scenario is sophisticated and complex so that the children open up to multiple solutions or whether it is limited? It further deliberates upon the role of the teacher, the teacher's response to children's ideas and questions and how he or she deals with conflict and issues of sharing and equity (Fig. 2).

Fig. 2 Affective assessment

9 Assessment Tools

Assessment strategies should encourage children to show what they know and what they can do, rather than focusing on what they do not know or cannot do. Focusing on children's thinking rather than a particular answer or solution provides valuable information about a child's learning. Sometimes their thinking is revealed through dialogue or demonstrated through their behaviours. The teacher's greatest assessment tool is a continual process of observation and documentation of learning because children show their understanding by doing, showing and telling. Therefore, teachers need to use the assessment strategies of observing, listening and asking probing questions to assess children's achievement. In addition to documented observations, other assessment tools include anecdotal records, photographs, videotapes or audio recordings, work samples and portfolios.

The idea of the school child as a passive recipient of knowledge is being rejected in favour of a socially active and participative child, democratically bringing existing knowledge and ideas into the classroom. If the schools permit students to make decisions about the content, processes and outcomes of their learning, it will mean providing a more flexible, less prescriptive environment that encourages worthwhile knowledge and promotes creativity.

This type of arrangement sees a school as a democratic and communal site, where the curriculum is co-created by teachers and students working together with an understanding of each other's roles and responsibilities. Too many of the things that students experience in school do not matter in their lives. Children who do not find anything of their worlds represented in the school curriculum are less likely to find it motivating, and think it is 'not for me'. I wanted to explore whether schools can respond to changing policies and participate in emerging programmes. TNS is a learning community where teachers have managed to balance child choice with curricular needs, following the Reggio approach centred on constructivist practices.

The quest for new methods of assessment, conducted thoughtfully, is undeniably worthwhile. The resulting data provide evidence needed to persuade educators that Reggio-inspired schools are superior both in quality and support of children's learning.

10 Focused on the Future: The New School

When I read about the Reggio approach, I thought that it would be hard to incorporate into practices of teaching, especially with a gruelling curriculum. However, after my visit to TNS (The New School), I saw it being implemented successfully.

After years of engaging with some of today's foremost thinkers such as Roger Schank and Noam Chomsky, Beaconhouse, a private educational establishment in Pakistan, decided to establish TNS; a school that aspires to be a beacon of hope in a world of testing-obsessed schools, lighting the way for the future of schools in Pakistan. TNS formally commenced classes on 10th September 2007 to cater to children aged 3–11 in Nursery to Year 5. It is essential for all children to be able to think creatively, express themselves and be able to adapt to a changing world. Beaconhouse recognized and responded to this change and TNS was a first step towards it. The program is based upon the constructivist principles promoting continuous assessment and progress. Affective assessment, which is the focus of my research, is fostered and supported by three important aspects of the program: the environment, project-based learning, and documentation.

11 Believing in the Power of the Child

Nothing excites a child more than having the opportunity to learn something that no one else knows. At TNS, children are contributors to and participants in the learning process. Because children are allowed to be leaders, the adults around them spend a considerable period of time learning from them (in contrast to the traditional model, where the teacher is the authority and provider of information). According to Malaguzzi (1993), this image is of a child with extraordinary strengths and capabilities, engaged in co-construction of knowledge in relation with others. Reggio educators believe a child has rights not just needs (Fig. 3).



Fig. 3 Not just any place (The New School, TNS)

At TNS there is emphasis on building rich meanings rather than completing tasks. Students are intellectually engaged, and feel a sense of collaborative partnership. Classrooms like these are extremely rewarding places to teach and learn.

12 Progettazione: The Curriculum

The curriculum Reggio Emilia creates is called *progettazione*. It acknowledges and reinforces the power of children and the power of social interactions. It is an emergent curriculum not mandated from administrative or standards setting bodies. This means it is not scripted; rather, it is drawn and changes with the needs and observations of the children, families, teachers, and community. It employs real life situations and long-term projects to enhance interaction and learning opportunities. Mundane tasks such as setting a lunch table or cleaning up afterward are included in the learning experience. The long-term projects are of interest to the children because the children themselves want to explore the topics. Educators may introduce challenging problems for further exploration.

These general directions are informal and subject to change. These aid in creating a developmental practice, not prescribed from a desire to achieve a particular competency or skill. The teachers are constantly listening and observing children's actions and interactions to develop '*progettazione*'. By selecting topics suggested by children, a message is reinforced that children's thoughts are valued. It is a fact that an evolving curriculum is more difficult to manage but it's recommended as it is incredibly exciting for all concerned (Edwards et al., 1998). Mistakes do occur within the concept of proximal development (Malaguzzi, 1993). But failure should not be feared. It is from failure that growth comes; provided that one can recognize and learn from it.

I am drawn to the idea of children orchestrating their own learning while thoughtful and purposeful adults work with them in long-term projects. Reggio educators talk about thoughtful actions taken by adults to extend children's thinking. This concept recognizes that children should make their own discoveries instead of being told what to do. I think children need time, time to ponder, time to explore, time to repeat things over, to shape their own ideas and to realize their potential as image-makers and knowledge-builders. While watching the children at a months-long project on space, one visitor remarked, "This topic is too complex for preschool children; children should do projects on things that are in their backyard! Children can't visit outer space!" Adults, especially educators, do like to decide what children can and cannot do and what they should and should not learn (Fig. 4).

The question is who gets to decide what can interest the individual? What we can and cannot understand? Who stands between us and what we want to know and tells us where we're not allowed to wander? Yes, space is far; it's also in our backyard. Through project work and documentation of the learning process,



Fig. 4 Project based learning: knowledge-builders at TNS

children were given the facility to co-construct and negotiate knowledge, which ultimately belonged to them.

13 Setting Sail: Project-Based Learning

In the Reggio approach, the teacher uses the child's work as a guide to gauge what the child knows. Children do not have the verbal or writing skills to tell others what they know. Children use the languages of drawing, painting, clay, or wire to express what they understand. Studying their artwork allows teachers to understand children and their learning. Teachers use drawings as a basis for discussion and educational growth. Art is not a 'subject', but a language for children to express understanding of their environment. Art is not used as a decoration; the art studio is a place of research.

Projects or flexible plans are co-constructed by children and teachers which generate documentation and are regenerated by documentation. They may start from an idea or problem posed by children or from an experience initiated by teachers. Projects can last as long as the children remain interested in the topic and continue exploring through negotiated learning (Gandini & Goldhaber, 2001). Timing is at children's pace, not that of a wall clock. Project work facilitates learning of subject matter as well as related skills. Group work provides opportunities for shared problem-solving. Through project work, children produce art that reflects their creative, communicative and intellectual potential. As mentioned, this is what Malaguzzi (1993) describes as The Hundred Languages of Children (Gandini, 1998a, 1998b). These symbolic representations express the capacity of children to produce superior quality work which teachers document in their discussions with children.

In many ways project work transcends social, racial and hierarchal patterns and dissolves the rigid patterns of society. Project work involves asking pertinent questions, working as groups to find the answers, and finding creative ways to report those answers. Parents become partners, and peers. Learning is meaningful as all relationships are strengthened. Project learning is holistic reading, writing, researching, constructing, measuring, experimenting, and reporting. One goal of projects is to dig deeper. Make learning complex, more layered.

Every effort we make to slow down means more opportunity for higher-level thinking (Abramson, Seda, & Johnson, 1990; Enright & McCloskey, 1985). Repeated experiences help children develop language skills. I was interested in how children developed vocabulary and the teacher's role in helping children expand their vocabulary. Thus, I observed children in as many situations and activities as possible (for 3 days). I noticed that a core vocabulary developed among the children due to routines such as identifying the weather (it's raining) and social interactions (I am sorry). Vocabulary was shared by all children, including children who spoke no English (English is a second language to Pakistanis).

The recurring behaviours were conferring, sharing and responding. Conferring was interactions with teacher; sharing was among children talking about what they have done. There were of course differences in the individual children's motivation to learn language. All things were not equal for the children, despite their similar experiences in the classroom. However, their language learning differences were opportunities that showed how different children behaved as they acquired a new language. Children developed skills for writing and reading, but they developed them in balance with more childlike modes of expression.

There were no instructional tasks, focusing on specific skills, no assigned worksheets, and no alphabets to be taught each week, and no designated time to practice in writing one's name. Rather, writing and reading skills were incorporated into aspects of daily experiences. Children appeared to have fun, and specific academic standards were met through activities (Genishi, Yung-Chan, & Stires, 2000).

Projects that inspired me most were about 'space' and persona dolls. Persona dolls stimulated an exchange of ideas around issues of unfairness and injustice (Details of projects and pictures in Appendix 1).

14 Documentation: An Epistemology of Caring and Concern

We know that assessment is much more than test results arbitrarily designed and quantified. The intent of good assessment practice is to collect data and make sense of it in order to change teaching and learning in a proactive direction. The most valuable contribution of the Reggio approach is the use of documentation of a child's experience. Documentation, in the forms of observation of children and extensive recordkeeping, has been encouraged and practiced in many early childhood programs. However, compared to these practices in other traditions, documentation in Reggio focuses on children's experience, memories, thoughts and ideas in the course of their work. This method of teaching focuses on purposeful listening (Malaguzzi, 1993).

Documentation is important as it provides information about children, their learning and their role in the learning experience. It encourages discussion among

educators and parents, reinforcing collaboration in the learning process, where the adult educates the child and the child educates the adult. The documents reveal how the children planned, carried out, and completed the displayed work. In-depth documentation reveals the learning paths that children take and the processes they use in their search for meaning. A commonly noted feature of children in Reggio schools is their meta-cognitive understanding of their own learning processes. The Reggio Emilia experience fosters children's intellectual development through a systematic focus on symbolic representation. Young children are encouraged to explore their environment and express themselves through multiple paths and all their 'languages', including the communicative, symbolic, cognitive, imaginative, and relational. The Reggio approach respects every child's potential for developing competencies. Documentation does not mean measurement. Assessment is based on regular documentation of children's work that is stored in the portfolios, binders, and journals

Documentation of children's work and ideas contributes to the quality of the learning in the following ways:

- (i) Documentation contributes to the extensiveness and depth of children's learning. Children become curious, confident and interested in what they have achieved (Malaguzzi, 1993).
- (ii) The process of preparing and displaying documents provides debriefing or re-visiting of experience during which understandings are clarified and strengthened. It assures children that their ideas and work are taken seriously. Taking children's work seriously encourages responsibility and commitment, and develops both delight and satisfaction in the processes.
- (iii) Documentation is based on the evaluation of work as it progresses. Planning decisions are made on what children found interesting or challenging.
- (iv) Documentation makes children's learning visible. Of particular relevance to Pakistani educators, documentation provides information about learning that cannot be determined by formal standardized tests. The variety of media used provides evidence of growth that can't be gauged through examination.

15 Elements of the Documentation Process

At Reggio, documentation is seen as the process of co-constructing knowledge during concrete experience (Kvale, 1992). Pedagogical documentation is a process of visualization (Dahlberg, Moss, & Pence, 1999), a social construction where choices are made about what to document. This means that the teacher who is documenting is a co-constructor, a participant, attempting to understand children and how they learn. Thus teachers collect, examine, analyze, recollect, re-examine their data many times over, to gauge the extent of growth in children.

The content concerns what the children are saying and doing. This content is obtained in different ways including notes, audio-video recordings and



Fig. 5 Children contributing to documentation

photographs. Meaning is produced through interpretation and reflection. This reflection is done in a rigorous, systematic manner involving teachers, parents and children. Involving children in the process of documentation is a powerful and valuable learning tool, for both the children and their teachers (Fig. 5).

The documents give teachers a common frame of reference for examining the developmental domains categorised as physical/cognitive, language/literacy, social/emotional, and creative/aesthetic. For each worksheet teachers selected a specific behaviour and then wrote anecdotal notes about a target child. They also took photos as a record of what they saw and developed their notes and photos into a more complete observation consisting of three parts (notes, photographs, and interpretation). Their lens for documenting was ‘micro,’ with a ‘who’ of an individual child, a ‘what’ of a behavioural domain, a ‘where’ of the classroom, a ‘when’ of day/week of class, and a ‘how’ of the process. The teachers labelled and analyzed child behaviour according to the learning categories. The teachers compiled their observation pages into a Child Notebook—a piece of finished documentation that was used in the classroom. Meanwhile, the worksheet pages were also put into the classroom. A blank book for each child soon became embellished by the work of the students. They added observational information throughout the semester by recording speech and behaviour on sticky notes attached to the appropriate expectation. By the end of the semester, all of the worksheet pages and informal notes for each child were put together into a Child Development Assessment Report, so that parents and teachers could benefit from the accumulated evidence of the child’s growing competencies. After the implementation of a planned activity, teachers created an activity piece that communicated the goal and objectives for the activity as well the children’s response to it. This multiple purpose documentation is hard work. It requires time, but it is rewarding. I was able to reflect and see the ‘process’ wherever the child wanted it to go, rather than being forced in a particular direction (Fig. 6).

It helped me in many of my own practices in the classroom. Listening through active participation, documentation, and the formation of small groups are a few of the methods that inspired me most. However, unless educators become conscious, change cannot occur. Project based learning and documentation offer possibilities for implementation and deserves consideration for wider application in our education system.



Fig. 6 Documentation: reflecting on the learning process

16 The Portfolio as an Assessment Tool

One of the alternative methods in education used in the assessment of students' performance is the portfolio. The necessity of using portfolios is acknowledged by many studies (Birgin, 2003; De Fina, 1992; Gussie & Wright, 1999; Mumme, 1991). According to these researchers, the portfolio gives more reliable and dynamic data about students for teachers, parents and also students themselves. The use of the portfolio as an assessment tool is a process with multiple steps. The process takes time, and all of the component parts must be in place before the assessment can be utilized effectively. General assessment alone is not a goal for a portfolio. Portfolios are most useful for addressing the student's ability to apply what has been learned. According to Arter and Spandel (1991), a portfolio is a purposeful collection of student work that exhibits to the student, or to others, her efforts in one or more areas. Paulson, Paulson and Mayer (1991, p. 60) define the portfolio as a purposeful collection of student's work that exhibits progress and achievement in one or more areas. The collection must show involvement of students in selecting content, the criteria for selection and evidence of student self-reflection. Collins (1992a, 1992b, p. 452) identifies the portfolio as "evidence with a purpose. Evidence is documentations that can be used by one person or group of persons to infer another person's knowledge, skill, and/or disposition". According to Barton and Collins (1997) portfolios are authentic forms of dynamic assessment. Therefore, portfolios should be on going so that they show students' efforts, progress, and achievement over a period of time. Thus, in this study the portfolio is promoted as a systematic and purposeful collection of the evidence which reflect the success, performance, and efforts of the students in one or more areas over a period of time.

The *documentation portfolio at TNS* was a collection of work over time showing growth and improvement in students' learning. Portfolio assessment provided more authentic and valid assessment of students' achievement; comprehensive views of students' performances in contexts; encouragement for students to be independent and self-directed learners; and improved communication among teacher, student and parents. Additionally, they provided opportunities for the learner to demonstrate their weakness and strengths; take responsibility for their own learning; and enhance communication with their teacher. Portfolio assessment revealed a

potential to demonstrate students' learning process and learning product over time. As a result, portfolios gave detailed information about students' development in the learning process to teacher, parents and students themselves.

Portfolios of children's work were displayed and sent home at key intervals and transitions. This was done to present a view of the learning process, and they showed both attempts and accomplishments. These were purposeful collections of examples of student's work that show the questions, interests, creative and analytic processes. Teachers prepared memory books, to trace the experience of children. Process research (formative evaluation) was central to program improvement.

Reggio educators consider pedagogical documentation to be an instrument for reflection and not assessment (Dahlberg et al., 1999; Oken-Wright, 2001; Rinaldi, 2001). Unlike most of the high-stakes tests that students undergo, students are not evaluated on a single occasion, the assessments are multifaceted. Teachers use graphic organizers such as Correlation Chart, Stacked Line Graph, DIGA (Describe, Interpret, Generalize, and Apply), KWHL (Know? Want to know? How will I find out? What did I Learn?), and the Learning Matrix.

The goal of portfolio grading is to strike a balance between product and process. In other words, a student's learning process is just as important as their ability to produce a well-shaped product which meets standard rubric-like criteria, so portfolio grading focuses on both the learning process as well as the quality of products. Consequently, portfolios are graded as a whole, and each item in a portfolio can be used to showcase a student's best work or provide evidence for a student's self-assessment of their learning process and growth. Portfolios present a wide perspective of learning process for students and enable continuous feedback for them. They also help students gain important abilities such as self-assessment, critical thinking and monitoring one's own learning (Asturias, 1994; Micklo, 1997). The use of portfolios can additionally answer these questions: what kind of troubles do students have? Which activities are effective or ineffective? How efficient is the teaching process? I support De Fina's views that portfolio assessment has many advantages over standardized testing. This is shown in Table 1 (De Fina, 1992, p. 39).

As shown in Table 1, portfolio assessment enables measuring high-level skills with meaningful activities for students instead of measuring low-level skills in a limited time; it also allows the use of multi assessment methods instead of using only one measurement method, making assessment a continuous process of determining the student's weaknesses and strengths. Many theoretical and empirical studies have reported the superiority of portfolio assessment to traditional assessment tools in education (Asturias, 1994; Baki & Birgin, 2004; Barton & Collins, 1997; Birgin, 2003; De Fina, 1992; Ersoy, 2006; Klenowski, 2000; Kuhs, 1994; Mullin, 1998; Norman, 1998; Sewell, Marczak, & Horn, 2002).

Although using portfolios has advantages, it has some disadvantages as well. Portfolios are considered time consuming and challenging to evaluate. However, if we use portfolios from a representative sample of students rather than having all students participate, this approach may save considerable time, effort, and expense.

Table 1 Portfolio assessment versus standardized testing (De Fina, 1992, p. 39)

Portfolio assessment	Standardized testing
Occurs in the child's natural environment	Is an unnatural event
Provides opportunity for student to demonstrate his/her strengths as well as weaknesses	Provides a summary of child's performance on certain tasks Provides little diagnostic information
Gives hands-on information to the teacher on the spot	Provides ranking information
Allows the child, parent, teacher, to evaluate the child's strengths and weakness	Is an one-time "snapshot" of a student's abilities on a particular task
Is ongoing, multiple opportunities for observation	Assesses artificial task, which may not be meaningful to the child
Assesses realistic and meaningful daily literacy tasks	Asks child to provide a singular desired response
Invites the child to be reflective about his/her work and knowledge	Provide parents with essentially meaningless and often frightening numerical data
Invites the parents to be reflective of child's work and knowledge	Forces teacher-administration conferencing
Encourages teacher-student conferencing informs instruction and curriculum; places child at centre of the educational process	Reinforces idea that the curriculum is the centre of the educational process

As for evaluation—the use of rubrics can help facilitate greater consistency between raters.

Moreover, portfolio assessment places new demands on teachers in terms of professional development; time to learn how to use portfolio; and preparation time to create new materials and lessons. Teachers also need additional time for reviewing and commenting on students' work. Such requirements force teachers to develop themselves in their fields. However, research shows that teachers see portfolios as a worthwhile burden with tangible results in instruction (Koretz, Stecher, Klein, & McCaffrey, 1994; Stecher, 1998).

Children learn about themselves and their environment if they are given the chance to take risks, form opinions, discover new ideas, and share findings about the hidden beauties of the world. In our country, many demands are placed upon students and teachers to increase performance on standardized tests. Of course we want our students to be proficient in areas such as reading, writing and mathematics, but children also need freedom to explore. Many children learn best through guided discoveries, experiences and collaboration with others.

The experiences that children at TNS are given will advance them in more ways than we could ever imagine. I saw children open up, express themselves, and bring out hidden talents which increased their confidence.

17 Tree House and Wheel Barrows: Exciting Discoveries at TNS

The philosophical underpinnings of Reggio Emilia take inspiration from Piaget (1969), Dewey (1938), Vygotsky (1978), and Gardner (1983). The practical application of which will require time, energy, resourcefulness and commitment. The example of TNS proves the benefits of adapting Reggio Emilia principles to develop language (English) abilities in culturally diverse elementary-age students.

Fundamental to a feeling of ‘belonging’ is the ability to express ones’ self. Sometimes this can’t be done verbally (Pakistan is a multiethnic society). In settings like this, with no common verbal language, English holds great importance. It is therefore extremely important to find other forms of expression to build on language abilities—something fundamental to Reggio principles.

The Reggio approach aims for a paradigm shift when children, teachers and parents actively work together to produce new knowledge. One parent I met at the School remarked “the school has helped me discover my children”. As a teacher of young adults, I yearned for the day that a parent would say that to me. But instead they used to demand that I prove that my classroom learning was developmentally sound. What they didn’t realize is that children’s development is being suppressed by assessment after assessment. When I observed young children at TNS, I remembered Pamela Houk’s poem ‘if’ and I agree, “*what a child can really do if he or she is allowed to explore the world, discover new ideas, and express feelings*” (cited in Edwards et al., 1998, p. 293).

As I conclude my narration, I feel a deep sense of pride in the ongoing research that has made this concept a reality. I wanted to highlight the achievements and the difficulties faced during the implementation of the Reggio approach in a Pakistani context. Much work is still to be done. We still have a lot to learn about how to develop programmes that promote a positive sense of belonging in the early years. However, we know how to make a start:

Learning and teaching should not stand on opposite banks and just watch the river flow by; instead, they should embark together on a journey down the water. Through an active, reciprocal exchange, teaching can strengthen learning how to learn (Malaguzzi, cited in Edward, 1993, p. 56).

18 My Reflections and Observations

My investigation of the Reggio Emilia approach has taught me a great deal about learning and especially assessment. It has also moved me to appreciate the young child as a capable being. In the process of my research, I have re-affirmed my faith in child-centered learning and affective assessment.

Prior to my visit to TNS, I wondered how the idealistic vision of a Reggio school could be implemented in a country where there is so much emphasis on testing; I wondered how the time to listen to students and their ideas could be found. But, in my observations at TNS School, I saw teachers striving to implement some of the Reggio ideals and make the school a happier place. The visit provided me with in-depth knowledge of project-based learning and documentation. Many aspects of this approach appealed to me. I liked the strong emphasis on creativity and problem solving with peers (with adult guidance) that gives many students opportunities to improve social and communication skills. I especially liked the opportunities for students to improve communication skills through discussions of visual representations of their work.

However, there are some potential snags and cultural challenges involved too. These challenges might not be overwhelming for an elitist school such as Beaconhouse where only children of affluent families are enrolled but could be elsewhere.

The Challenges faced in implementing the Reggio project are:

- (i) Reggio Emilia is a city located in affluent wine country in Northern Italy. There is very little unemployment. Families have ample financial resources. In contrast, many communities in Pakistan have a myriad of problems including financial difficulties.
- (ii) In Reggio Emilia, the relationships among administration, teachers, and families are reciprocal. In Pakistan, they are not. In some cases, but not all, Pakistanis tend to view parents as amateurs and teachers as professionals. Parents view schools as institutions with rules and standards. In Reggio Emilia, the teachers, parents, and community work together, to build a school with a sense of history, tradition, and culture. They constantly reassess what they are doing in order to improve themselves and their schools.
- (iii) In Pakistan, we rely on setting and meeting standards to provide quality schools. Understanding the differences in cultures has helped me understand what we are up against when trying to make changes in assessment.

I was struck by the powerlessness that many teachers feel over the present process of educational change in the country. The greatest contrast found, in my opinion, was the absence of willing participation of families and a lack of understanding of affective assessment.

I found a lot of things I could relate to as a teacher, but because of the expectations forced upon me as a teacher, I worry that I may never be able to realize them. This is a huge concern for me, because I feel we have to attend to the necessary evils of accountability and standardized measures of achievement.

Our education system is hierarchical. Its structure is rigid and that rigidity reinforces the status quo. We admit standardized tests are biased and yet we use more standardized testing and measuring. The question remains, how does a focus on children's exploration and production of symbols fit into the current educational context, obsessed and anxious over high standards?

Allowing children to learn this way requires trust—trust that delving into a long-term project gives children what they need. It requires the administrators to trust the teachers—that they will make sure the students are meeting the learning standards and that they will address any areas that aren't satisfied during the project work. Usually, schools are not comfortable with this. The most frequent criticism is that the system neglects the academic basics that children need to master at an early age. The focus on testing, and the rivalry for places in educational institutions is particularly intense. This may explain parental resistance and why different approaches to assessment are not taken seriously. For these reasons, TNS is the only school that is presently using the approach. “Children have a lot to say”, said Amina as she stood in the school's atelier, a narrow room off the library lined with pictures that the children drew. “It is our job to make sure that they are able to work. We can't stand in their way”.

We need to find ways to balance our curriculum to meet the needs of society as well as the needs of each child. We need to give our children the chance to discover their voices and give them freedom to use them. What children have to say can be valuable. It could make a world of difference in young lives for someone to hear their treasured thoughts and ideas. I acknowledge that the transition to standard elementary school might be difficult for the young learners. After experiencing an unrestrained environment where they were recognized as valid components of the school community and quite capable of making curricular decisions, to then move to a highly structured elementary school with standardized tests, could be problematic, yet I prefer to see the glass half-full.

19 A Note of Thanks

I am thankful for the time given to me by the teachers, administrators and of course children and parents. I am especially grateful for the material; the brochures, pictures and the samples of how documentation is carried out, and the consent to use this in research. I have fond memories of TNS, which for me gave a rare glimpse of possibilities...




Appendix 1

See Tables 2 and 3.

Table 2 Language and literacy

Using writing and drawing as a method of representing ideas	 A photograph showing a child's drawing of a person and a piece of paper with handwritten text in Chinese. The text reads: "I was making squiggly lines. I got the idea to make squiggly lines from up in my head. The lines looked like the legs of an ant." data-bbox="531 206 838 295"/>
Using books and pictures to tell stories	 A photograph of a child in an orange shirt painting a picture on a white surface. There are paint containers and brushes nearby.
Being able to match sensory experiences to descriptive terms, sour/sweet, cold/warm,	 A photograph showing several children sitting on the floor, engaged in a sensory activity. They appear to be touching or playing with various materials.
Retelling sequence of events in a story	 A photograph of a group of children in a classroom setting. They are standing and appear to be participating in an activity or discussion.
Reading from print outs/computers	 A photograph of a young girl sitting at a desk, looking at a computer screen. She is pointing at the screen, which displays text and images.

Table 3 Projects based learning: children busy at work

The project	The child	The teacher
I am an artist	Wow <i>beautiful paint class</i>	
	Oh I love drawing And paints too	We will explore the materials without direction, moved from idea to idea to idea (painting paper, to painting leaves, to printing...),
	I think I will make a cartoon film —I love to draw figures	“Baber I notice that you are interested in figures. Would you like to study figure drawing? Would you like to add some writing to your figures too? Do you want to get some books when we go to the library tomorrow?”
	Can I use red all over! Hmm, yaa at sunset its red!	Do you think the sky is red?

References

- Abramson, S., Seda, I., & Johnson, C. (1990). *Literacy development in a multi lingual kindergarten classroom. Childhood Education, 67*(2), 68–72.
- Arter, J. A., & Spandel, V. (1991). *Using portfolios of students’ work in instructions and assessment. Education Measurement: Issues and Practices, 11*(1), 34–44.
- Asturias, H. (1994). Using student’s portfolios to assessment, mathematical understanding. *The Mathematics Teachers, 87*(9), 698–701.
- Baki, A., & Birgin, O. (2004). Reflections of using computer-based portfolios as an alternative assessment tools: A case study. *The Turkish Online Journal of Educational Technology (TOJET), 3*(3), 79–99.
- Barton, C., & Collins, A. (1997). *Portfolio assessment: A handbook for educators*. New York: Dale Seymour Publications.
- Birgin, O. (2003). *Investigation of the application level of a computer-based portfolio*. Unpublished master’s thesis, Karadeniz Technical University, Trabzon, Turkey.

- Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum.
- Carl, R. (1983). *Dialogues: Conversations with Martin Buber, Paul Tillich, B.F. Skinner, Gregory Bateson, Michael Polanyi, Rollo May, and other*. Ann Arbor: University of Michigan Press.
- Ciari, B. (1950). *In dialogue with Reggio Emilia: Listening, researching and learning*. London: Routledge.
- Collins, H. (1992a). *Manifesto for learning*. New York: Continuum International Publishing Group.
- Cole, M. (1996). *Handbook of early childhood literacy*. New York: Sage.
- Collins, A. (1992b). Portfolios in science education: Issue in purpose, structure and authenticity. *Science Education*, 76(4), 451–463.
- Dahlberg, G., Moss, P., & Pence, A. R. (1999). *Beyond quality in early childhood education and care: Postmodern perspectives*. London: Routledge.
- De Fina, A. A. (1992). *Portfolio assessment: Getting started*. New York: Scholastic.
- Dewey, J. (1938). *Experience and education*. New York: MacMillan.
- Dewey, J. (1997). *Democracy and education: An introduction to the philosophy of education*. New York: Simon & Schuster.
- Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Research Dialogue in Learning and Instruction*, 10(1), 11–20.
- Edwards, C. (1993). Partner, nurturer, and guide: The roles of the Reggio teacher in action. In C. Edwards, L. Gandini, & G. Forman (Eds.), *The hundred languages of children: The Reggio Emilia approach to early childhood education* (pp. 151–169). Norwood, NJ: Ablex.
- Edwards, C., Gandini, L., & Forman, G. (1998). *The hundred languages of children: The Reggio Emilia approach—advanced reflections*. Portsmouth: Greenwood Publishing Group.
- Enright, D. S., & McCloskey, M. (1985). *Integrating English*. Reading, MA: Addison-Wesley.
- Ersoy, F. (2006). Opinions of teacher candidates as to the portfolio assessment. *Elementary Education Online*, 5(1), 85–89. Retrieved from <http://ilkogretim-online.org.tr>.
- Fiksdal, S. (2001). *A guide to teaching effective seminars: Conversation, identity, and power*. London: Routledge.
- Gandini, L. (1998a). Educational and caring spaces. In C. Edwards, L. Gandini, & G. Forman (Eds.), *The hundred languages of children: The Reggio Emilia approach to early childhood education* (pp. 161–178). Norwood, NJ: Ablex.
- Gandini, L., & Goldhaber, J. (2001). Two reflections about documentation. In Lella Gandini & Carolyn Edwards (Eds.), *Bambini: The Italian approach to infant-toddler care* (pp. 124–145). New York: Teachers College Press.
- Gandini, L. (1998b). Educational and caring spaces. In C. Edwards, L. Gandini, & G. Forman (Eds.), *The hundred languages of children: The Reggio Emilia approach to early childhood education, Advanced Reflections* (pp. 161–178). Norwood, NJ: Ablex.
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. London: Heinemann.
- Genishi, C., Yung-Chan, D., & Stires, S. (2000). Talking their way into print: English language learners in a pre-kindergarten classroom. In D. S. Strickland & L. M. Morrow (Eds.), *Beginning reading and writing* (pp. 66–80). New York: Teachers College Press, Routledge.
- Gussie, W. F., & Wright, R. (1999). *Assessment of the implementation of portfolio assessment in K-8 school districts in New Jersey*. Paper Presented at the Annual Meeting of the Eastern Educational Research Association, Hilton Head, SC. (ERIC Document Reproduction Service No. ED429996).
- Katz, L. (1998). What can we learn from Reggio Emilia? In C. Edwards, L. Gandini, & G. Forman (Eds.), *The hundred languages of children: The Reggio Emilia approach—Advanced reflections* (2nd ed., pp. 27–45). Westport, CT: Ablex Publishing.
- Klenowski, V. (2000). Portfolios: Promoting teaching assessment in education. *Police & Practice*, 7(2), 215–236.
- Koretz, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Education Measurement: Issues and Practice*, 13(5), 5–16.

- Kuhs, T. (1994). Portfolio assessment: Making it work for the first time. *The Mathematics Teacher*, 87(5), 332–335.
- Kvale, S. (Ed.). (1992). *Psychology and postmodernism*. London: Sage.
- Luke, A. (2002). Curriculum, ethics, metanarrative: Teaching and learning beyond the nation. *Curriculum Perspectives*, 22, 29–55.
- Maher, F. A., & Tetreault, M. K. (1994). *The feminist classroom*. New York: Basic Books.
- Malaguzzi, L. (1993). For an education based on relationships. *Young Children*, 49(1), 9–11.
- Micklo, S. J. (1997). Math portfolio in the primary grades. *Childhood Education*, 97, 194–199.
- Mullin, J. A. (1998). Portfolios: Purposeful collections of student work. *New Directions for Teaching and Learning*, 74, 74–87.
- Mumme, J. (1991). *Portfolio assessment in mathematics. California mathematics project*. Santa Barbara: University of California.
- Norman, K. M. (1998). *Investigation of the portfolios as an alternative assessment procedure*. Unpublished doctoral dissertation, The University of Memphis, USA.
- Oken-Wright, Pam. (2001). Documentation: Both mirror and light. *Innovations in Early Education: The International Reggio Exchange*, 8(4), 5–15.
- Palgrave, F. (2002). *Golden Treasury*, Oxford Paperbacks; 6 New Ed edition.
- Paulson, F. L., Paulson, P. R., & Meyer, C. A. (1991). What makes a portfolio a portfolio? *Educational Leadership*, 48(5), 60–63.
- Peckham, M. (1976). *Romanticism and behaviour: Collected essays II*. Columbia: University of South Carolina Press.
- Piaget, J. (1969) *The psychology of the child* (H. Weaver, Trans.). New York: Basic Books.
- Popham, W. J. (2009). *Unlearned lessons: Six stumbling blocks to our school's success*. Cambridge, MA: Harvard Education Press.
- Rinaldi, C. (2001). Reggio Emilia: The image of the child and the child's environment as a fundamental principle. In Lella Gandini & Carolyn Edwards (Eds.), *Bambini: The Italian approach to infant-toddler care* (pp. 49–54). New York: Teachers College Press.
- Rogers, C. (1983). *Freedom to Learn for the 80s*. New York: Merrill Wright.
- Romberg, T. A. (1993). How one comes to know models and theories of the learning of mathematics. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education* (pp. 97–111). Netherlands: Kluwer Academic Publishers.
- Ropers-Huilman, B. (1998). *Feminist teaching in theory and practice: Situating power and knowledge in poststructural classrooms*. New York: Teachers College.
- Sewell, M., Marczak, M., & Horn, M. (2002). *The use of portfolio assessment in evaluation*. Retrieved January 10, 2002, from <http://ag.arizona.edu/fcr/fs/cyfar/portfolio3.html>.
- Shephard, K. (2009). Higher education for sustainability: Seeking affective learning outcomes. *International Journal of Sustainability in Higher Education*, 9(1), 87–98.
- Stiggins, R. J. (2005). *Student-involved assessment for learning*. Upper Saddle River, NJ: Pearson.
- Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment. *Education: Principles, Policies & Practice*, 5(3), 335–351.
- Stevenson, R. B. (1997). Developing habits of environmental thoughtfulness through the in-depth study of select environmental issues. *Canadian Journal of Environmental Education*, 2, 183–201.
- Vygotsky, L. (1978). *Readings for reflective teaching*. New York: Continuum International Publishing Group.
- Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA: MIT Press.
- Wilson, G. B. (1996). *Constructivist learning environments: Case studies in instructional design*. New Jersey: Educational Technology.

Provoking Potentials: Student Self-Evaluated and Socially-Mediated Testing

Tim Murphey

...the future of L2 developmental studies lies in the activity of teachers and learners.

(Lantolf & Poehner, 2014, p. 224)

Abstract This exploratory action research (Smith, 2015) describes a new conception of testing in which students are directed to evaluate themselves (give themselves grades) at two moments in time: the first after a certain amount of time filling in test answers that they can recall alone; and the second after asking others in the class for mediating help during a socially interactive time period. The first grade represents their own individual efforts, without utilizing their connections in the class. The second grade represents a situated person in a community with their connections in the class. Enacting self-evaluations and particularly the second stage of social testing seems to provoke potentials for expansive learning that may not normally emerge in traditional testing: potentials for self-appropriation of self-evaluation, agency, helpfulness, altruism, social learning, social construction, and the pedagogical learning of scaffolding and implicit mediation rather than explicit mediation (Nicholas, 2014a). I do not propose that these tests are valid for assessing each individual's competence (not that I believe many others are), but that these exploratory procedures enlighten students to different aspects of learning and evaluation, and help teachers to examine different aspects of classroom dynamics and learning potentials. I see these tests as a generative way of continuing student learning. While I do propose a way to test such tests more rigorously following conventional assessment guidelines, I am more concerned here with the expansive learning potentials provoked by the procedure and the parallels that seem to exist with dynamic assessment and socio-cultural theory, particularly the use of the zone of proximal development (ZPD) and the zone of proximal adjustment (ZPA). This social testing attempts to blend learning and assessment, which is an essential trait of dynamic assessment, and to blend theory with practice in praxis as described by Lantolf and Poehner (2014).

Keywords Testing · Assessment · Socially-mediated testing

T. Murphey (✉)

Kanda University of International Studies, Chiba, Japan

e-mail: mitsmail1@gmail.com

© Springer International Publishing Switzerland 2017

R. Al-Mahrooqi et al. (eds.), *Revisiting EFL Assessment*,

Second Language Learning and Teaching, DOI 10.1007/978-3-319-32601-6_17

1 Why Social Testing: Going Back in Time

Cozolino (2013) in his volume on *The Social Neuroscience of Education* proposes that a class should feel like a tribe, with everyone feeling like they belong. Without this feeling of belonging and attachment we spend much of our neurological energy asking ourselves who we are and defending ourselves from social exclusion. He also invites us to recognize that brains and neurons do not exist alone and are not isolated in the natural world.

Similarly, Lieberman's (2013) volume, entitled *Social: Why our brains are wired to connect*, brings attention to three main adaptations: first, that being well adjusted socially is as crucial to our well-being as food and water. He notes that the social pain of losing a loved one or breaking up or being an outcaste creates real physical pain directed by the same systems in the brain that tell us about physical pain. Secondly, we are a mind-reading species with mentalizing systems which are built to figure out what others are thinking. And finally, that our socially malleable selves often lead us to altruism (something we will see in the testing reactions below).

These neuroscience books came a decade after David Block's *The Social Turn in Second Language Acquisition* (2003) that highlighted the research in the field that supported the idea that we learn languages more easily the more we socially interact with each other. There has since been an even greater emphasis on the social turn in SLA (Atkinson, 2011; Ortega, 2009).

Dunbar (1998) claimed that the main reason humans had a big new brain part, the neocortex, was so that we could live in larger groups and be more actively social. The evidence that Dunbar and others found was impressive. This "sociality" led us to actually develop more intelligence. It seems that other species might have been more intelligent in the beginning, but it was our ability to be social that made us smarter and allowed us to survive longer in larger groups. Thus, it was "better to be social than smart" in the beginning. In other words, at first our brains got excited about social interaction, and that allowed us to improve our lives and our brains. The fact that Facebook is the most commonly visited website in the world (with over one billion accounts) also attests to our continual deep desire to connect and be social (Bower, 2013; Lieberman, 2013, p. 32), far beyond our mere desire for more information.

2 Vygotskian Socio-cultural Theory and Dynamic Assessment

Going back even further in history, Vygotsky wrote (circa 1930) that the teacher "has to become the director of the social environment, which, moreover, is the only educational factor" (1997, p. 339, cited in Lantolf & Poehner, 2014, p. 208). While a great part of our SLA educational endeavors are turning toward the social

understandings of learning, belonging, and creating, testing has been left mostly undeveloped in the shadows, except for a small SCT group of dynamic assessment researchers, inspired mostly by Vygotsky's zone of proximal development (ZPD) and Feuersteins's mediated learning experience (MLE), led by Lantolf and Poehner (2004, 2011a, 2011b, 2014), along with those who have proposed a "critical testing" movement (Kramsch, 1993; Pennycook, 1994; Shohamy, 2001).

Thus, I propose in this chapter, placing myself at the nexus of praxis where theory meets practice, that individual conventional tests are problematic as they attempt to measure a single person's abilities away from one's social networks which one naturally uses to solve problems in the real world. I think we need a deeper social turn in SLA as a whole, and in the testing field in particular, a "turn" that I think is being spurred by social neuroscientists and SCT dynamic assessment researchers. Freire (1985) writes acutely about the difference between inspection and evaluation:

Through inspection, educators just become objects of vigilance by a central organization. Through evaluation, everyone is a subject along with the central organization in the act of criticism.... In understanding the process in this way, evaluation is not an act by which educator A evaluates educator B. It's an act by which educators A and B together evaluate an experience, its development, and the obstacles one confronts along with any mistakes or error. Thus, evaluation has a dialectical character.... It's essential that members of the evaluating organization deeply believe that they have as much to learn from educators directly linked to popular bases as those who study at the bases. Without this attitude, the evaluators from an external organization will never admit to any gap between their view of reality and reality. By believing they possess the truth, the evaluators act out their infallibility. And with such a hypothesis, when they evaluate they inspect (pp. 23–25).

Swain, Kinnear, and Steinman (2011) introduce the problem well in saying, "We tend to take the use of tests for granted. However, underlying their use is a set of assumptions about the knowledge and abilities being tested that are different from those of SCT [socio-cultural theory]. For example, in general, we think of tests as something that must be done alone. It is considered cheating to ask a peer for help, to use a dictionary, or to search the Internet. Why?" (p. 118). Thus, our basic assumptions lead us down a path that ignores our sociality.

Lantolf and Poehner (2014) meticulously describe Vygotsky's zone of proximal development (ZPD) compared to Feuerstien's work with the mediating learning experience (MLE) and recognize that they have many commonalities. The most well cited definition of the ZPD is:

the distance between the actual developmental level as determined by independent problem solving, and the level of potential development as determined through problem solving under the guidance or in collaboration with more capable peers (Vygotsky, 1978, p. 86).

According to Lantolf and Poehner (2014), in much of the relevant literature, the ZPD has unfortunately been simplistically used to describe scaffolding procedures to achieve task completion rather than used more holistically to look at the internalization of psychological tools. Concerning Feuerstien's MLE Lantolf and Poehner (2014), state:

the *Mediated Learning Experience* (MLE) bears remarkable similarities to the view of ZPD as collective activity...both approach development as the internalizing of psychological tools that individuals use to organize and regulate their mental world...[and they both understand] mediated learning as a basic explanatory principle of development (p. 160).

However, both Vygotsky and Feuerstien were usually working with individuals, and much of the dynamic assessment literature is also about the ability of one teacher evaluating and adapting dynamically to a learner. This is very useful research but not usually what the majority of teachers actually experience. Of course they acknowledge that learning can and does happen between learners as well, especially in the case of Swain's collaborative dialogue (2006), for example, but their emphasis seems to be usually on an "expert" dynamically adjusting to the learning needs of a single learner. I contend that "experts" are just as well situated to learn and change as the students, of course in different ways and directions but nevertheless they, too, have abilities that change dynamically, or not, in interaction. I contend that every person has a zone of proximal adjustment (ZPA) as well as a zone of proximal development (ZPD). In other words, each person has a set of abilities to adjust well (ZPA), or not, to others that allow us to help them to certain degrees, and each person also has abilities to show others how they might adjust to us (Murphey, 1990, 1996, 2013a). I contend that learners can change experts as much as experts can change learners and that we are unknowingly often meditational means for each other's development. This seems to be accepted by SCT researchers when they are talking about learners learning together who are co-constructing, but rather ignored when they are talking about "experts".

Dynamic assessment grew out of educational psychology and counseling mostly, which is most often done in one to one situations. The fact that the great majority of the teachers in the world rarely actually interact with individual students, but rather spend most of their teaching time with a larger group, obviously dilutes the impact these experts might have. However, as Lantolf and Poehner (2014) show, one answer seems to be in training *students* to interact in ways that help each other develop through peer mediation (Shamir & Tzuriel, 2004; Tzuriel & Shamir, 2007) creating their own dynamic assessment and interactive learning. Because there is not enough of the teacher to go around and adjust qualitatively, teachers can teach students to adjust qualitatively to each other and, hence, create more opportunities for everyone to learn. Students are usually already more proximal to each other anyway. Teaching them to be sensitive to how each other learns can stimulate them even more in their everyday interactive classrooms. In addition to the social testing described below, activities like shared action logging and newsletters among students as well as student evaluated conversational videoing (Murphey, 1993, 2001a, 2001b) also can promote this sensitized mediation.

So the best qualitative feedback for students is probably that which comes from peers who are near the same level and experiencing similar things, are able to adjust to each other more qualitatively, and able to be near peer role models (Murphey, 2013a; Singh, 2010). There is a small but growing field of peer tutoring and learning that is recognizing this effect (Gafney & Varma-Nelson, 2008). In the

words of one of my students reflecting on the social testing: “Peers helped me. They knew many answers. I also helped them. I like this quiz, this does not make me feel isolated!”

Many years ago I asked students to make test questions and then later to actually evaluate their classmates’ oral skills in response to a set of questions (Murphey, 1996). The last few years I have been experimenting with what I call tentatively “social testing” with the procedures described below. This social testing attempts to allow students to learn further even while they are being assessed, which is an essential trait of dynamic assessment.

3 Student Self Evaluated Social Testing Procedures

My procedures start with a conventional test that slowly turns into a social collaboration, a contrast that is sharply noticed and commented upon in student feedback further below. The following steps have developed over the last 3 years of experimenting with this procedure (with a total of six semesters of university EFL classes, involving students from all four years with 20–80 students in a class).

1. Students take a regular style test (usually a fill in the blank, short answer, entailing recalling and reflecting on information) and after an appropriate amount of time (e.g. 20–30 min), they stop.
2. Then I tell them to put away their pencils and erasers, and to take out a pen (blue for best contrast) and give themselves an estimated score at the bottom of the test, say 50 % or 70 % or 86 %. I tell them I will take any erasers and pencils I see on their desks: they are only to have a pen from that point on.
3. Then I tell them they have 5 min (I usually lengthen it to 10–20 min depending on how active they are and how big the group is) to ask any of the questions to anyone in the room and to add to their answers or write down new answers on their tests. In order to make it as orally interactive as possible, I set a few more rules:
 - (a) *“You are not allowed to look at anyone’s paper or show your own paper to anyone”*. If I see this happening, I tell them they fail and I collect their papers. (Sometimes I need to explain that “copying” is not learning; whereas a dialogue can open ourselves up to an exchange of ideas and nuances.)
If I want them to interact with more people I tell them:
 - (b) *“You can only ask one or two questions per person”*.
 - (c) If they want to erase an old answer, they simply draw one line through it with the pen, ~~like this~~.
 - (d) I tell them that when the time is over, they will give themselves a second % score for the new state of the test and hopefully they have improved their tests a lot. [The change to ink allows the teacher to see approximately how

Table 1 Bottom part of the tests that students fill in

1st score _____/100 %; 2nd score _____/100 %; 3rd score _____/100 %
Who helped you?
Who did you help?
What do you think of this test?

much was answered with the help of others and how much was answered alone. They usually become intensively interactive during this time. I circulate and remind them loudly not to look or show their tests to anyone and to simply ask and dictate to each other. Many report actually constructing answers together during this time.]

4. After finishing the second part, I ask them to put in the second score and to write the *names of the people who helped them, the names of those they helped*, and to comment on *what they think of the test*. The bottom of my tests now look something like Table 1:

The third score, in Table 1, is for the teacher after the test and could be used in a variety of ways. For a final score, the first two scores can be averaged or calculated with different weights. Invariably with my overly humble Japanese students, I am raising the scores, but that may not always be the case.

Each time I do a test like this the students are in awe the first time, and after having done it and seen how much they learn from it and benefit from it they love it. The moment when a student wants to know the answers the most is likely to be on a test and trying to answer questions that they don't know the answer to. And chances are they will retain the information longer after learning them in an excited state (a hypothesis that needs verification). I fear my explanation cannot capture the excitement you will see when you start the second part of the test, and hopefully you will feel the positive provoking potentials of what you have started.

4 Analysis of Some Qualitative Findings

The first few years, I thought that this kind of test would only work one time a semester.

I thought if students knew they were always socially collaborative they would probably reduce their studying. However at least among some students it may make them study even harder so they can help their classmates, as UH below reports:

Because I had taken a test (#1) in this class and I knew how we would do the test #2, I tried to remember as much as possible not only for myself, but for my classmates. Last time I took the test, I was helped by others with answers, very helpfully. So, I wanted to help my classmates more than I did last time. In Test #2 it was interesting. I felt as if I was already working with classmates during my preparations for the test, and that motivated me to study. Although it was not so many people that I could help with the quiz, I was glad to hear "thank you" from them and to see their smiles. Showing thanks to people really makes them happy. (UH)

Comments below are from students in one class in 2013 who were asked if they wanted to have a third social quiz like the first two or no test. About two thirds of the students said they preferred no test while about a third actually wanted to have the test!:

Yes, because studying for quiz is hard, but thanks to quiz we can understand the class activities deeply. It means that quiz is good for us. That’s why I want to take a quiz. (AK ch)

I don’t want a test, but that’s OK if I do. Because if I do Tim’s quiz, I can memorize the activities of this class! (AN Ch)

Yes I want your test because it is so much fun to talk a lot. (YK Ch)

Table 2 shows some recent data from Spring 2014 concerning the last question at the bottom of the test: *What do you think of this test?* I had four classes that semester with from 20 to 68 students, and I sorted their comments into categories: Negative, Neutral, Positive, and Very Positive. Most of the negative comments merely stated that the test was difficult, hard, etc. and they did not have enough time to finish it. The neutral ones usually combined a negative with a positive such as “It was difficult, but I enjoyed collaborating.” The positive ones, 50 %, of the total, generally praised the social aspects of the test. And the 5 very positive I have placed below in Table 3 for your consideration. You can find all the evaluation comments for all 4 classes in Appendix 1.

These five comments (in Table 3) all come from my largest class that is mixed with all 4 years and all departments at my university. And while they are asked to find a new partner every class, the social testing mixes them up so much that they end up speaking to many for the first time (#1 below) and greatly enjoying themselves (#2 & #3). #4 below supports the idea that they are learning far more than merely test information (cf: expansive learning below); they are exercising “vital skills to live a real life!” such as asking, sharing, helping, and communicating (#5).

A few weeks after the tests I also gave each class their own comments back to them to read on a handout and talk about in a process known as critical participatory looping (CPL, Murphey & Falout, 2010). I also had enough room to put a few of the very thought provoking comments from other classes at the bottom of each

Table 2 Student Feedback Data on The First Social Tests in Spring 2014

# Students in class	65 ways	28 song	25 change	20 fresh	Totals	%
Negative comments	20	8	2	6	36	0.240
Neutral	13	6	7	7	33	0.22
Positive	37	14	16	7	74	0.50
Very positive	5	0	0	0	5	0.03

Answers to WDYTOTT (What do you think of this test?) at the end of the test

Table 3 Five very positive evaluations from the WAYS Class (unedited)

-
1. It was interesting because I think it is important to help each other and share information. I also had a chance to speak to strangers

 2. The most fascinated test that I've ever had

 3. I think this type of test is awesome!! We can help each other even in Quiz!! I think that is amazing! We can actually show our courage. We can co-construct our knowledge

 4. I really like this type of test. I've never done such a creative and interactive test, and I really think that I was required to get information and help people, and these are vital skills to live in real life!

 5. I've never had this type of test, but this is really fun. Talk to people and get answers from them. I learned how important to ask, share, communication are

handout so that they could consider what students from other classes were thinking (Appendix 1). As you will see in Appendix 1, most of the negative comments (24 % of the total comments, Table 2) are short, usually saying simply that the test was too *hard* and they needed more *time* and would *study harder* next time. A certain number of negative comments about a test being difficult are not necessarily bad as we do hope to challenge our students, and, especially for the purposes of social testing, they need to be encouraged to talk about the possible answers in what Swain calls collaborative dialogue (Swain, 2006).

There were a few practical difficulties that I encountered and in subsequent tests found ways to moderate them. One was emphasizing that they should only talk and listen and not look at each other's paper. Learning happens less well when we just copy what someone else has. When we listen and then write ourselves and talk about it, it sticks better and we might even improve the teller's answer. Now, before each test I warn them that if I do see them looking at each other's test, I will rip their test up and they fail. After telling them I rip a sheet of paper in half and say, "Don't make me do it. I don't like to, but I will." And I walk around the class, as they are doing the interactive part and remind them out loud, "Please do not look at the tests! Talk and listen and write."

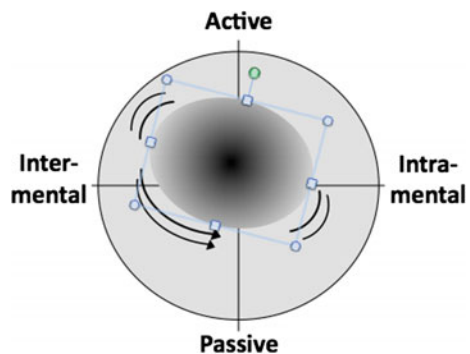
There is another mediating role, besides that of policing, that I enjoy playing in these test situations during the social testing portion. When I notice that most students do not know an answer to one of the questions, say #4, I go to the shyest person in the class, say Hiro, and I help him to answer #4 correctly, giving hints and clues usually. Then, I announce to the rest of the class that if they need the answer to #4 that "Hiro has it!" Hiro then becomes a very popular mediator of valuable knowledge for a while with new abilities and roles emerging in the process.

5 Discussion: Provoking Potentials and Expansive Learning

The fact that we bring people together in groups to learn creates great potential for social learning, however most teachers and schools do not treat students, nor classes, as active socially intelligent dynamic systems (SINDYS: Murphey, 2013c) which can learn from each other. Teachers often see the group as a threat to be controlled rather than a dynamic ever-changing mystery to be explored. And while we teach to the group, we evaluate the individual. Our present state of testing isolates students from what made us intelligent initially, each other. Our conscious mind, and the class mind, is like a cloud that is constantly moving between two sets of poles. A class together ideally would be able to be active and inter-mental, (Fig. 1) i.e. exploring each other’s thoughts. However, most often students are pushed into a passive role of listening to one professor talk and quite often drifting off to their own private worlds within their minds (i.e. intra-mental and passive) (Fig. 2).

What does SINDYS have to do with my Social Testing above? Well Dunbar (1998) and Leiberman’s (2013) arguments lead us to believe that we are first and foremost social animals meant to interact and learn from one another through socializing. The social testing procedures above and the students’ intuitive comments seem to support these findings. Most students also admit to enjoying being more active and intermental and learning a lot from the testing procedures, i.e. enacting an actively intermental SINDYS. In our normal everyday world, we often access information about the health or economic activity of people like ourselves so that we can hopefully live more intelligently, creating SINDYS. When a teacher returns information that she has gathered from the class, back to the class, also known as Critical Participatory Looping (Murphey & Falout, 2010), she can make

Fig. 1 Socially intelligent dynamic systems consciousness cloud in motion



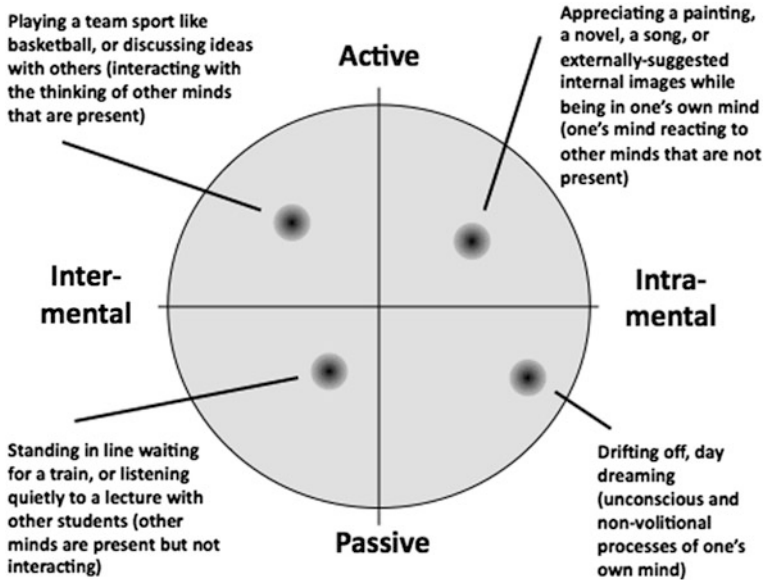


Fig. 2 SINDYS: examples of mentalizing in the quadrants

the group more intelligent about itself creating an active SINDYS in the classroom. Every class is a SINDYS, but some are unfortunately more passive and dormant and instrumental than others who receive more information about the group that they can use to engage and learn with. For Falout, this is the social crux:

Teachers in Japan endeavor to make personal connections with their students and inspire them. But in a culture of learning inclined primarily toward testing grammar knowledge on paper, the point soon comes when lessons become heavily reliant on explications of complicated textual phrases and passages that are largely unapplied, unwieldy, and uninteresting. ... English is then a school subject with specimens of the language captured somewhat randomly and pinned down for dissection with various parts studied as facts or rules (Falout, 2013, p. 145).

Social testing brings SINDYS into the assessment arena and allows social intelligence (the social crux) to play a greater role, which of course it already plays in the real world.

I do not naively believe that everyone will stop their individual testing of students in the near future and begin social testing only. But I would like to suggest social testing as an alternative way of doing testing for those who feel inclined and as a way to learn, perhaps for a “real” test done a class later. The evidence above suggests that students learn a lot more through teaching each other and interacting socially with material that fits nicely with task-based learning rather than being isolated and simply listening to a professor’s lecture.

6 Expansive Learning and Self Evaluation

This type of testing creates environments of emergent and expansive learning (Sannino & Ellis, 2014). While perhaps allowing less teacher control over the final product, it creates great un-imagined and emerging affordances:

At the beginning of a process of expansive learning the object is only abstractly mastered as a partial entity, separated from the functionally interconnected system of the collective activity. By ascending to the concrete, an abstract object is progressively cultivated into concrete systemic manifestations and transformed into a material object that resonates with the needs of other human beings as well. These phases often require the subject to struggle and break out of previously acquired conceptions in conflict with new emerging ones (Sannino & Ellis, 2014, p. 8).

In this new testing environment, students are at first often confused as they have never done a test in this way. As they venture to ask questions and answer the ones their partners are asking they ascend to the concrete and engage in teaching and learning. They become regretful when they are not able to help a classmate and decide they want to study harder for the next test, and they are thankful when a classmate helps them understand something. Some discover they can construct unique answers together and that talking about the questions helps them understand them far better than simply struggling to write them down. In this expansive learning, potentials for broader learning are provoked beyond the set curriculum and students take their learning on a journey through listening, helping, advising, and giving.

I have noted elsewhere (Murphey, 2010) that we often over-rate information, success, and teaching (telling) processes when we should give more value to questions, challenges and experiential learning as they allow more expansive learning to emerge. I believe these interactive testing procedures seem to create room for better questions, challenges and experiential learning. Graciously the universe appropriately constructs things with enough eternal complexity as to forever allow us challenges that help us develop. Facing these challenges together (e.g. collaborating on tests) make living more joyful than we will probably ever have the capacity to fully appreciate. And thus we continue to challenge the impossible, daring greatly, partially failing, and marvel at the miracles of tentative success and understanding, improvising and softly assembling our ways through chaos.

Furthermore, asking people to evaluate themselves for a real test can engage them in a process much deeper than most young people usually engage in. Students tell me at first that they don't know how to evaluate themselves and are often at a loss. Forced to do so by the rubrics of the social test, they estimate and struggle with the same questions that teachers struggle with as well. "Is this question worth the same as another? Is this answer sufficient? How much do I know that was not asked about? How hard did I study? Is it shown on the test?" Evaluating yourself can be an enlightening thing and part of expansive learning, taking you places you had not known you could go.

If I might be permitted a brief basketball analogy, the professional and more recently collegiate basketball teams in the US have started to use an interesting statistic to evaluate the contributions of players called the With or Without You Stat (Keating, 2012). They calculate simply how much the team scores as a whole for the time that a player is on the court. For example, it used to be if player-A played the first half and player-B took his place in the second half and we saw that player A scored 10 points and player-B scored only 4, we might decide that player A was much better. But if we looked at the With or Without You Stats it might show that the team actually scored more when B was in the game in the second half (with all his teammates scoring much more) and their opponents scored much less in the second half, then we might have to re-evaluate player B as being better for the team. He is probably blocking more shots, stealing more balls, passing better and rebounding better (the intangibles that are hard to count sometimes). I give this scenario to my language students and ask them how they can be like player B in the classroom and stimulate their classmates and help them learn more? When they realize that the way they act influences the wider group, they take more responsibility for their actions. In a testing situation, it leads them to want to help more and give good answers. And I believe their altruism becomes contagious.

Another interesting study involved students simply using “we” self talk rather than “I” in saying affirmations before competing in dart throwing (Son, Jackson, Grove, & Feltz, 2011):

Before completing a team-based dart-throwing activity, 80 undergraduates were randomly assigned to one of three pre-performance self-talk conditions: (a) self-talk statements that focused upon one’s personal capabilities, (b) self-talk statements emphasizing the group’s capabilities, or (c) a control condition where neutral statements were implemented. Participants in all conditions subsequently rated their confidence in their own (i.e. self-efficacy) as well as their team’s (i.e. collective efficacy) capabilities, before carrying out the task. Overall, self-efficacy, collective efficacy, and performance indicators were all greatest for individuals who practised self-talk focusing on the group’s capabilities, as opposed to individual-focused and neutral conditions. Findings are considered with respect to their novel theoretical contribution to the social cognition literature and their implications for fostering efficacy perceptions and team performance (p. 1).

The different five self-talk statements are given below:

(i.e. “I am a [We are] confident performer[s]”, “I [We] believe in my [our] ability”, “I am [We are] focused and ready”, “My [Our] skill will improve with every throw”, and “I [We] will perform well”). The neutral statements, on the other hand, made no reference to one’s skill, ability or performance on the upcoming task (i.e. “I live in [location]”, “I am a student”, “I am a male/female”, “I am [age] years old”, and “I have [colour] eyes”) (p. 3).

For a long time many have believed that there is nothing more competitive and individual than sport performance. However, the basketball and dart throwing examples are leading us to understand that individuals that see themselves not as individuals but as a group helping each other do better than those with a purely individual frame of mind.

7 Future Studies: Learning Tests Followed by Conventional Tests

One option worth investigating is doing this type of testing before a conventional final exam to spur more collaborative study by students outside of class before the test so they can see what they are not so sure about and discuss it with classmates in class during a social practice test and outside of class in preparation for the actual final exam. To test the impact of the social testing, we would need two similar groups studying the same content with the same teacher. One would do a social test the class before the final test, and the other given the same amount of time in class to study alone or with others before the final, in order to keep “time on task” relatively the same. Our hypothesis is that those who did the Social Practice Test would end up having higher average scores on their finals and probably socialize and bond better with their peers.

I would also like to see “self-evaluation” studied more rigorously as well. My guess is that when students are more comfortable evaluating themselves appropriately, they are more stable and dare to take on more challenging tasks that lead them to learn more and be more explorative in many domains. The trouble in most educational environments is that the evaluations are practically always done by others and thus children grow up thinking they have to please others, not themselves, and the extrinsic nature of the evaluation often dilutes the more natural intrinsic motivations that they may have.

I recently attended a presentation on dynamic assessment and pragmatics by Nicholas (Sept. 2014a, b), and I became more sensitized to the possibility of training students not to just give answers, but to help, and give hints to, and scaffold the learning of their peers more, i.e. not just give explicit mediation but implicit mediation, or what I choose to call *collaborative soft assembly*. My proposal is actually to make many more small social tests to help students learn more. This is touched on by Roediger, who describes how tests can make us smarter:

When my colleagues and I took our research out of the lab and into a Columbia, IL., middle school class, we found that students earned an average grade of A- on material that had been presented in class once and subsequently quizzed three times, compared with a C+ on material that had been presented in the same way and reviewed three times but not quizzed. The benefit of quizzing remained in a follow-up test eight months later...

This isn't just a matter of teaching students to be better test takers. As learners encounter increasingly complex ideas, a regimen of retrieval practice helps them to form more sophisticated mental structures that can be applied later in different circumstances. Think of the jet pilot in the flight simulator, training to handle midair emergencies. Just as it is with the multiplication tables, so it is with complex concepts and skills: effortful, varied practice builds mastery.

8 Limitations

The greatest limitation in this exploratory action research study is that I did not have time to evaluate the learning potentials that I saw emerge. I do hope that some graduate students might engage with these ideas as a Ph.D. dissertation and do them justice following the descriptions above (Future Studies). At present these are still very much just a part of exploratory action research practice, but I must confirm a practice I plan to continue doing and exploring.

9 Conclusion

Social testing with student self evaluations can create better learning conditions and expansive learning. I do not expect it to replace our conventional tests any time soon, however I offer it as a potential for creating more learning and helping in our classrooms for those teachers who are more concerned with student learning than student testing. It also ties in nicely with how students will most often work for the rest of their lives: they will not be taking conventional tests but most probably collaborating with others to create and improve products, services, and conditions for accomplishing essential tasks in the work force. Thus, exercising their communicating, across what Cozolino (2013) calls the “social synapse”, might be enriching their potentials in their future lives. (I know of no one who lists “test-taking” under the rubric *workforce skills* on their resumes. However, listing “helpful person-centered communicator” is about crossing that social synapse effectively and something that leads to more creatively collaborative communications.)

In the end, I may be developing students’ altruism rather than testing their English (cf. UH comment above) and developing their ability to learn more socially rather than testing an isolated brain unconnected to others. As one student commented: “I enjoyed the test, especially because I could help others with answers.” Actually a handful of students comment this way at the end of every test, saying that being able to help others, not their own test scores, was what made the test so exciting. Holistically, I see both of these outcomes as far more useful than simply learning or testing English. This is what might be called Value Added Language Learning (Murphey, 2013b), i.e. the learning and development of English through the learning of more important things that lend value to English. Thus, learning to help others learn by helping them find the answers they need is a noble quest. Infusing testing with helping habits and socially networked learning might just create better communities for students to learn with. My colleagues and I have started referring to these groups as Present Communities of Imagining (Murphey, Falout, Fukada, & Fukada, 2012) and SINDYS (Murphey, 2012).

It is about time we acknowledged that our minds are no longer, if ever they were, isolated, independent, and individual entities, but rather our minds and our brains

are interconnected and networked, and work best with other minds in collaboration. Both the philosopher Bache (2008) and the neuroscientist Cozolino (2013) concur that living in social collective creativity seems to be our calling.

This Wilga Rivers (1975) quote summarizes much of what I wish to say above and supports our social testing within exploratory action research and principled improvisation:

The essence of language teaching is providing conditions for language learning—using the motivation which exists to increase our students’ knowledge of the new language; we are limited only by our own caution, by our own hesitancy to do whatever our imagination suggests to us, to create situations in which students feel involved—individually, in groups, whichever is appropriate for the age level of our students in the situation in which we meet them. *We need not be tied to a curriculum created for another situation or another group. We must adapt, innovate, improvise...* (p. 96).

[bold added]

10 Postscript

Objectivity is useful when trying to evaluate others, but practically impossible when trying to evaluate yourself. And yet, the person who will evaluate you the most in your life is yourself, and that person also has the greatest impact on your life. So if possible it is probably a good idea to make friends with yourself and understand how you think and how you would like to think. Most people realize that if they evaluate themselves harshly, they kill their motivation. Thus, being kind and encouraging toward one’s self is a good idea. You will probably generate better behaviors and do more in your life. This comes from frequent self appraisal. It seems many young people are convinced that only others (and “objective tests”) can tell them how worthy they are. This is a great disservice to learning and identities-in-context.

The social testing procedure is not objective and does not wish to be. It is unashamedly subjective and personal. Many people never take the time to do self-evaluations consciously and thus this procedure brings one’s thoughts on self-evaluation to the surface for their evaluation. This topic deserves a complete article in and of itself, but for now suffice it to say, our subjective selves rule a much larger portion of our lives than we think. Thus, it behooves us to pay some attention to how we evaluate ourselves and to do so in low-risk situations. The social testing procedures should be done often, every few weeks if possible, because as Roediger (2014) says (cited above), “tests make us smarter”. I would amend it to say, “good tests make us smarter, especially when we can learn how to help others”.

Appendix 1

Feedback from 4 classes of students after their first social test May 2014 which were looped (Murphey & Falout, 2010) back to the individual classes for their further consideration. Note: a few comments from *other classes* were also looped back to each class at the end of each handout.

CLASS #1WAYS OF LEARNING (all 4 years, all departments) = 77 students

What do you think of this test? Test one May 19 2014

***ON BACK: Please write your opinion of these comments and the test.**

Circle comments you agree with a lot.

Negative evaluations – (20)

- 1. This test was a little difficult**
- 2. This test was very difficult. I could not write the answers.**
- 3. Enjoy and also a little bit confusing because there are a lot of answers.**
- 4. It was too hard to answer because I've not complete songs and could sing those. After testing, friends told me where I missed.**
- 5. I hope I will not get F grade because of this test.**
- 6. It's very hard for me to write correct answer.**
- 7. I need to review**
- 8. I could not answer back side of test.**
- 9. I should study more...regret.**

10. I'm confused when writing the lyrics. Did I need to write, "I'm in love" and Doitagain repeatedly?
11. It's hard to me to answer all questions.
12. I think this test is difficult for me.
13. It's a little difficult for me. I did well about songs but others were terrible.
14. It is so difficult. If I have a chance I would study hard. I can sing a song, but I can't explain the meaning of some story.
15. I should have prepared more ☹
16. It's difficult for me to remember the content of the textbook. I forgot the content, so I feel this test is difficult.
17. I need to summarize the contents of this class to enjoy learning English more.
18. It is so difficult. But I think I get English song score. I can't remember some story. Next time, I want to improve English skills.
19. This test was slightly difficult for me, especially PLU part was.
20. This test was so difficult for me that I couldn't solve easily. I should study more.

Neutral Evaluation (in the middle 13)

1. It's a good test. However, I didn't know where test is. So, some questions are difficult to remember.
2. It's very difficult but I can write some answers because student helped me.
3. I read the PLU chapters but I forgot the titles. I could confirm whether my answer is collect or wrong.
4. It is a little bit difficult. However it is good for me to remember.

5. I could fill in the songs, but it was difficult to know PLU or what LH means and so on.
6. Helping each other is good. There are many things that I forgot!
7. I enjoyed but I thought I have to read carefully the PLU.
8. This type of test was really new for me!
9. I felt so difficult but it's very useful and makes my skill up.
10. I didn't remember all of things so I need to read PLU more.
11. I thought this test is unique. I forgot many contents of PLU. Hh
12. The test was fun. However, time to work on by myself was not enough. I could talk with unknown classmates.
13. I could sing these things but when read I forgot the spell or made some mistake. I could notice this and next time I will memorize completely.

Positive Evaluation + 37

1. I hope the other's classes tests are like this! It was fun ☺
2. Some parts were a little bit difficult, but I could enjoy to answer, because in my head, so many songs.
3. It's really interesting and completely new for me. I want to do this again even if it's not test time.
4. It was fun. Score Two system is interesting!!
5. I liked to compare my answers to others. That was an interesting test.
6. Some questions let me think I have to read textbook more.
Socializing answer (!) is fun ☺
7. The test was a little bit difficult for me but I could enjoy teamwork.
8. It's a little bit difficult for me but teamwork is good.

9. I enjoyed the test because I could ask many people.
10. I'm so excited by the test. I have never told answers to my classmates! It's difficult to ask because... *[unreadable-to-typist]*
11. I like this type of test! I could talk many people and could get a lot of answers from them.
12. I was able to talk to new people and learn new knowledge! It was unusual but fun ☺
13. This test was pretty difficult because I forgot some things. But asking other students is really enjoyable and useful to fill blanks and understand.
14. I really enjoyed asking answers to others. However, I want to know the correct answers.
15. If we always do homework and enjoy class (songs, NLS), this kind of test is easy! It was fun.
16. It's fun.
17. This test was fun at the last part because you get to interact with people to ask answers around the classroom.
18. It's fun because I can communicate with lots of people I've never talked to. I've never done this type of test though.
19. It was a little bit difficult. However, I felt helping each other is very fun and important!!
20. This test is fun and to learn a lot!
21. It is a good way to be helped by new friends.
22. That was good because test reminds me what I learned.
23. Everyone I asked helped me, and also I could help some people. It was good!!

24. It was enjoyed! I could make time to talk to old and new classmates. It was fun ☺ But I have to take everyday class more seriously than before!
25. I think it was interesting. Helping with other people really works and I think it is good way to learn new thing.
26. I remember the lyrics of songs with melody. I re-recognized the 5 memorizing ways are great!
27. It was fun! Next time, I will be more ready and read PLU more.
28. Even the test makes me help communicate with someone.
29. The test is a little difficult for me but I could talk many people. It's fun.
30. Interesting, I've never done the test like this!
31. The type of test is interesting because I haven't taken test such the way.
32. It was fun!!! Sharing information we have was good!
33. Fun. Useful. Difficult.
34. Interesting! Finding the answers from students is important for us.
35. I enjoyed the test and helping friends to find answers.
36. It's fun.
37. It was good to talk with others.

++ Very Positive Evaluation WAYS Class

1. It was interesting because I think it is important to help each other and share information. I also had a chance to speak to strangers.
2. The most fascinated test that I've ever had.

- 3. I think this type of test is awesome!! We can help each other even in Quiz!! I think that is amazing! We can actually show our courage. We can co-construct our knowledge.
- 4. I really like this type of test. I've never done such a creative and interactive test, and I really think that I was required to get information and help people, and these are vital skills to live in real life!
- 5. I've never had this type of test, but this is really fun. Talk to people and get answers from them. I learned how important to ask, share, communication are.

One empty / One Inc omprehensible

Another Class's comments to consider

- 1. I like this kind of test because it has been interesting to share answers with peers. Japanese tests make me isolated, but today's wasn't.
- 2. I really enjoyed sharing the answer. My classmates helped me very well. But sometimes I couldn't answer, and I felt very sorry about it. So in the next time, I'll study more!

CLASS #2 Song (CBI) Test 1 What do you think of this test? May19 2014 3rd&4thEng.Dept 28 students

***ON BACK: Please write your opinion of these comments and the test.**

Circle comments you agree with a lot.

Negative Evaluation

- 1. Difficult. Memorizing the words is more difficult than I had thought. I need more time
- 2. It was difficult because I misunderstood that the test was only about call and response songs.

3. Paper test is difficult than sing a song. Maybe I have a lot of spelling mistakes.
4. It's difficult, in particular describing.
5. Difficult. I must study after the class.
6. It is difficult. I couldn't guess at all.
7. I need more time because I couldn't fill up the blanks which I know the answers for.
8. Too difficult

Neutral Evaluation

1. It's a good opportunity to review what I did in the class. I don't hate the type of test, but I don't like tests.
2. This type of test, I'm not good at this. It is helpful to learn because writing whole sentences is better than mark sheets.
3. It's a little hard thing for me, but sharing the ideas is a good way to understand more each other.
4. Song test is interesting, but others were difficult for me.
5. I like this way of the test. But time is little short for me. I did not have time despite that I knew the answer.
6. I could remember many song-lines I guess because of melodies. I wrote answers with singing inside of my heart.

Positive Evaluation

1. I like it. I think writing is a good way to learn.
2. This type of test is really good for me. I've never done like this one. It's really good way to communicate with classmates and review.
3. I like it.
4. I've taken this kind of test before, and I liked it because you can know how much effort the classmates put.

5. **It's a little hard thing for me, but sharing the ideas is a good way to understand more each other.**
6. **This test is very useful for me because I have to remember, so I could improve.**
7. **Social part is good.**
8. **Interesting.**
9. **It was [more] difficult than I thought. But classmates helped me a lot! It's a good way to learn things I don't know.**
10. **I like the style of the test**
11. **Great! But I helped a lot so I couldn't ask what I wanted to ask.**
12. **I had fun! Asking questions to my classmates and completing my test is very unique way to take a test.**
13. **The time is a little bit short to me. But asking others after the test is unique. I've never taken this kind of test.**
14. **If I don't know the answers, I could ask other students. It was good way to remember.**

Another Class' comments to consider

1. **It was interesting because I think it is important to help each other and share information. I also had a chance to speak to strangers.**
2. **The most fascinated test that I've ever had.**
3. **I think this type of test is awesome!! We can help each other even in Quiz!! I think that is amazing! We can actually show our courage. We can co-construct our knowledge.**
4. **I really like this type of test. I've never done such a creative and interactive test, and I really think that I was required to get information and help people, and these are vital skills to live in real life!**

5. I've never had this type of test, but this is really fun. Talk to people and get answers from them. I learned how important to ask, share, communication are.

CLASS #3 Change (CBI) Test1 What do you think of this test?

3rd&4thEng.Dept May 16 2014 /25 students

***ON BACK: Please write your opinion of these comments and the test.
Circle comments you agree with a lot.**

Negative Evaluation

1. A little difficult because I was absent three times.
2. It was more difficult than expected. I thought it was just about songs.

Neutral Evaluation

1. It was a little difficult to answer except SD. However, it would help me to understand the class and content.
2. It was very difficult for me, but it was also useful
3. It was more difficult than I expected so that I'm going to work harder!
4. This test made me see how I didn't understand this class. I try to study more.
5. In some questions, I couldn't remember part of the question. But some of them, I remembered after stretching.
6. I think it's a good test, but it wears my wrists down, thanks to many questions.
7. Some questions are difficult for me, but I enjoyed it.

Positive Evaluation

1. **It's difficult some parts, but I can learn more and more, so I like it.**
2. **This is a weird but interesting test. I've never taken this kind of test before.**
3. **I think this test makes sense because we can review what we learned by memorizing and we can ask help after the test.**
4. **Please don't change this way of testing.**
5. **It's fun, but a little bit difficult. Also, we can help each other.**
6. **It was enjoyable and helpful to ask answers to many people.**
7. **It is good that we can ask other people.**
8. **I felt that I should have reviewed more for the test! However this was a good chance for me to talk with classmates.**
9. **It was so much fun to have a conversation with many people.**
10. **It was interesting because I haven't taken this type of test.**
11. **I think this so good for us because we can review all classes.**
12. **I couldn't write some answers on my own, but it was fun to get help with friends.**
13. **Asking classmates the answers what I'm not sure was really helpful! I like this style.**
14. **I like this kind of test because it has been interesting to share answers with peers. Japanese tests make me isolated, but today's wasn't.**
15. **I really enjoyed sharing the answers. My classmates helped me very well. But sometimes I couldn't answer, and I felt very sorry about it. So in the next time, I'll study more!**

16. This test covered what I learnt in all lessons, so it's difficult and confused. However it's good for reviewing. This test became one of good materials.

Another Class' comments to consider

- 1. It was interesting because I think it is important to help each other and share information. I also had a chance to speak to strangers.**
- 2. The most fascinated test that I've ever had!**
- 3. I think this type of test is awesome!! We can help each other even in Quiz!! I think that is amazing! We can actually show our courage. We can co-construct our knowledge.**
- 4. I really like this type of test. I've never done such a creative and interactive test, and I really think that I was required to get information and help people, and these are vital skills to live in real life!**
- 5. I've never had this type of test, but this is really fun. Talk to people and get answers from them. I learned how important to ask, share, communication are.**

CLASS #4 FRESH Test1 What do you think of this test? May 14 2014 / 20 students 1st year

***ON BACK: Please write your opinion of these comments and the test. Circle comments you agree with a lot.**

Negative Evaluation

- 1. It's very difficult. And I can't fill in the empty. I'm not good at the memory ability. I don't want to try it again (just kidding!)**
- 2. It's difficult for me. But I didn't study very well. I try to study very well next test.**
- 3. I remembered only songs, so I couldn't answer the questions of PLU. I should have read PLU again.**
- 4. I think this test is a little bit hard. Because I didn't remember the correct spelling. So if I have another test, I should study more.**
- 5. This test is difficult for me. There are easy questions. I learned SSS but I can't answer. I forgot the answer.**
- 6. I thought my English skills are little and short of English vocabulary. I have to listen to teacher's what to say. I have to study harder.**

Neutral Evaluation

- 1. I think this test is a little bit difficult. I've never taken such test.**
- 2. This style of test is very interesting. I've never take any test like this. I thought I must look both of text and notebook carefully. And I want more time. It was short to answer all.**
- 3. I'm racking of effort! [lacking] I should usually study hard. This test makes me decide to make a effort.**
- 4. I thought this test is difficult, but it is fun. I could not remember No. 5 and No. 6. I remembered all the song.**

5. I could answer the questions that ask me about some song. I think it is because I talked and sung with my friend about it. I should study more about PLU to become LH!
6. I could answer only song's question. I couldn't answer clearly 4,5, 6!! So this test was difficult. And I found out that I must read PLU more times!!!
7. It was a little difficult. I forgot to study about 5 and 6. So I have no idea.

Positive Evaluation

1. I think this test is very good!! "Why do you think so?" Hahaha... the answer is very easy. A monkey even can answer this question. Through this test, we could know that I can't live without partner or companies.
2. It's very fun. I feel like playing game. Someone can help me and I can help someone. We can communicate each other.
3. I think this test helps me to study English. I didn't know how improvement my English skills. So I study this test, for example Language Hungry is fascinating to me!!!!
4. I think this test is fun. But I can't answer [some] questions, because I studied a little. So I want to study hard next time.
5. I think this test is difficult to me. But I enjoyed it because I can know classmates answer. I like this style of test.
6. This test was very interesting. Especially, songs were really fun. While sometimes there were some hard questions for me actually.
7. Very hard. But AWWWWesome!

Another Class' comments to consider

- 1. It was interesting because I think it is important to help each other and share information. I also had a chance to speak to strangers.**
- 2. The most fascinated test that I've ever had.**
- 3. I think this type of test is awesome!! We can help each other even in Quiz!! I think that is amazing! We can actually show our courage. We can co-construct our knowledge.**
- 4. I really like this type of test. I've never done such a creative and interactive test, and I really think that I was required to get information and help people, and these are vital skills to live in real life!**
- 5. I've never had this type of test, but this is really fun. Talk to people and get answers from them. I learned how important to ask, share, communication are.**
- 6. I like this kind of test because it has been interesting to share answers with peers. Japanese tests make me isolated, but today's wasn't.**
- 7. I really enjoyed to share the answer. My classmates helped me very well. But sometimes I couldn't answer, and I felt very sorry about it. So in the next time, I'll study more!**

References

- Atkinson, D. (Ed.). (2011). *Alternative approaches to second language acquisition*. Abingdon: Routledge.
- Bache, C. (2008). *The living classroom: Teaching and collective consciousness*. Albany: State University of New York Press.
- Block, D. (2003). *The social turn in second language acquisition*. Washington D.C.: Georgetown University Press.
- Bower, B. (2013). *Bigger numbers, not better brains, smarten human cultures*. Retrieved from <https://www.sciencenews.org/article/bigger-numbers-not-better-brains-smarten-human-cultures?mode=topic&context=49>.

- Cozolino, L. (2013). *The social neuroscience of education: Optimizing attachment & learning in the classroom*. New York: W. W. Norton.
- Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 6, 178–190.
- Falout, J. (2013). The social crux: Motivational transformations of EFL students in Japan. In T. Coverdale-Jones (Ed.), *Transnational higher education in the Asian context* (pp. 132–148). Basingstoke, UK: Palgrave Macmillan.
- Feuerstein, R., Feuerstein, S., Falik, L., & Rand, Y. (1979; 2002). *Dynamic assessments of cognitive modifiability*. Jerusalem: ICELP Press.
- Freire, P. (1985). *Pedagogy of the oppressed*. New York: Continuum.
- Gafney, L., & Varma-Nelson, P. (2008). *Peer-led team learning: Evaluation, dissemination, and institutionalization of a college level initiative*. New York: Springer.
- Keating, P. (2012, August 13). On the plus side. *ESPN: The Magazine*. Retrieved from http://espn.go.com/nba/story/_/id/7551067/nba-utility-plus-minus-statistic-espn-magazine.
- Kramsch, C. (1993). *Context and culture in language teaching*. Oxford: Oxford University Press.
- Lantolf, J., & Poehner, M. (2004). Dynamic assessment: Bringing the past into the future. *Journal of Applied Linguistics*, 1(1), 49–74.
- Lantolf, J., & Poehner, M. (2011a). Dynamic assessment in the classroom: Vygotskian praxis for L2 development. *Language Teaching Research*, 15, 11–33.
- Lantolf, J., & Poehner, M. (2011b). *Dynamic assessment in the foreign language classroom. A teacher's guide* (2nd ed.). Center for Advanced Language Proficiency Education and Research, Pennsylvania State University, University Park, PA.
- Lantolf, J., & Poehner, M. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York: Routledge.
- Lieberman, M. (2013). *Social: Why our brains are wired to connect*. New York: Crown Publishers.
- Murphey, T. (1990). 'You and I', adjusting in interaction to get comprehensible input. *English Teaching Forum USIA*, 28(4), 2–5.
- Murphey, T. (1993). Why don't teachers learn what learners learn? Taking the guesswork out with action logging. *English Teaching Forum*, 31(1), 6–10.
- Murphey, T. (1996). Proactive adjusting to the zone of proximal development: Learner and teacher strategies. Paper presented at the *2nd Conference for Socio-Cultural Research Vygotsky and Piaget*, Geneva, Switzerland, September 11–15.
- Murphey, T. (2001a). Tools of recursion, intermental zones of proximal development, and critical collaborative autonomy. *JALT Journal*, 23(1), 130–150.
- Murphey, T. (2001b). Videoing conversations for self-evaluation in Japan. In J. Murphy & P. Byrd (Eds.), *Understanding the courses we teach: Local perspectives on English language teaching* (pp. 179–196). Ann Arbor, MI: University of Michigan Press. (pp. 179–196). Cambridge: Cambridge University Press.
- Murphey, T. (2010). Creating languaging and agency. *The Language Teacher (JALT)*, 34(4), 8–11.
- Murphey, T. (2012). *In pursuit of WOW!*. Tokyo: Abax.
- Murphey, T. (2013a). Value added English learning. *Journal of Saitama Educators*, 3(5), 1–8. Retrieved from https://www.academia.edu/6951792/Value_added_English_through_CBI_and_CLIL.
- Murphey, T. (2013b). Value added English through CBI and CLIL. *Journal of the Saitama City Educators*, 3(5), 1–8.
- Murphey, T. (2013c). The Impact of self-information as feedback to socially intelligent dynamic (SINDY) systems. *PeerSpectives*, 10, 19–21.
- Murphey, T., & Falout, J. (2010). Critical participatory looping: Dialogic member checking with whole classes. *TESOL Quarterly*, 44(4), 811–821.
- Murphey, T., Falout, J., Fukuda, Y., & Fukuda, T. (2012). Group dynamics: Collaborative agency in present communities of imagination. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory and practice* (pp. 220–238). New York: Palgrave.

- Nicholas, A. (2014a). *Dynamic assessment and pragmatics*. Paper Presented at Kanda University of Studies for the Research Institute of Language Studies.
- Nicholas, A. (2014b). Using strategic interactions in the classroom. *PeerSpectives*, 13, 2–7.
- Ortega, L. (2009). *Understanding second language acquisition*. London: Hodder Education.
- Pennycook, A. (1994). *The cultural politics of English as an international language*. London and New York: Longman.
- Rivers, W. (1975). *Speaking in many tongues: Essays in foreign-language teaching*. Rowley, Mass: Newbury House Pub.
- Roediger, H. (2014, July 20). How tests make us smarter. *The New York Times Sunday Review*. Retrieved from http://www.nytimes.com/2014/07/20/opinion/sunday/how-tests-make-us%20smarter.html?utm_source=Dan+Pink%27s+Newsletter&utm_campaign=b3bde2b380%20October_Newsletter10_14_2014&utm_medium=email%20&utm_term=0_4d8277f97a-b3bde2b380-306083877&r=1.
- Sannino, A., & Ellis, V. (Eds.). (2014). *Learning and collective creativity: Activity- theoretical and sociocultural studies*. New York: Routledge.
- Shamir, A., & Tzuriel, D. (2004). Children’s meditational teaching style as a function of intervention for cross-age peer-mediation. *School Psychology International*, 25, 58–97.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. New York: Longman.
- Singh, S. (2010). Near-peer role modeling: The fledgling scholars education paradigm. *Anatomical Sciences Education*, 3, 50–51.
- Smith, R. (2015). Exploratory action research as workplan: why, what and where from? In K. Dikilitas, R. Smith, & W. Trotman (Eds.), *Teacher-Researchers in Action* (pp. 37–45). UK, IATEFL. ISBN 978-1-901095-71-5.
- Son, V., Jackson, B., Grove, J., & Feltz, D. (2011). “I am” versus “we are”: Effects of distinctive variants of self-talk on efficacy beliefs and motor performance. *Journal of Sports Sciences*. Retrieved from <http://dx.doi.org/10.1080/02640414.2011.593186>.
- Swain, M. (2006). Linguaging, agency and collaboration in advanced language proficiency. In H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). New York: Bloomsbury Academic.
- Swain, M., Kinnear, P., & Steinman, L. (2011). *Socio cultural theory in second language education: An introduction through narratives*. Multilingual Matters.
- Tzuriel, D., & Shamir, A. (2007). The effects of peer mediation with young children (PMYC) on children’s cognitive modifiability. *British Journal of Educational Psychology*, 77, 143–165.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds & Trans.)). Cambridge, MA: Harvard University Press.
- Vygotsky, L. (1997). *Educational psychology*. Boca Raton, FL: Nova Science.

Part V
Future Perspectives on Assessment

Electronic Intervention Strategies in Dynamic Assessment in an Omani EFL Classroom

Priya Mathew, Rahma Al-Mahrooqi and Christopher Denman

Abstract This paper explores the suitability of Dynamic Assessment (DA) as a method of formal testing when the intervention is both electronic and supported by face-to-face encounters. The principles of DA appear in Vygotsky's socio-cultural theory (SCT) which postulates that cognitive development occurs when there is productive interaction. In DA, as opposed to psychometric tests, the learner is offered mediation during or after assessment. Performance with the assistance of the mediator helps the assessor to determine the learner's progress in the "zone of proximal development" or ZPD. Vygotsky describes the ZPD as the distance between a learner's actual level of development without mediation and their level of potential development when interacting with an able mediator. Participants in this study were a group of 12 EFL learners enrolled in the foundation programme of an Omani university that was designed to equip them with the language skills required for English-medium tertiary education. Students emailed pre-specified academic essays during the course of a semester to the first author who then offered them feedback using a word processor's review function. Students were then assessed on their ability to incorporate the researcher's feedback which ranged from implicit to explicit. A focus group interview with participants was held in addition to a series of observations to explore emergent trends associated with DA. Overall results suggest that electronic forms of DA involving mediation attuned to participants' ZPD are more effective than pre-scripted prompts based on assessors' guesses about the kinds of intervention learners may require during assessment. The paper concludes by suggesting that electronic forms of DA ensure students get the best possible mediation when they are undertaking important assessment processes and therefore may be of benefit in Omani EFL tertiary contexts.

P. Mathew (✉)
Middle East College, Seeb, Oman
e-mail: Priya@mec.edu.om

R. Al-Mahrooqi · C. Denman
Humanities Research Centre, Sultan Qaboos University, Muscat, Oman
e-mail: Denman@squ.edu.om

Keywords Dynamic assessment (DA) • Zone of proximal development (ZPD) • Cognitive development • Electronic feedback • Mediation

1 Introduction

The inadequacy of assessing L2 proficiency using psychometric tests alone has been raised by a number of authors. For example, Luria (1961), one of Vygotsky's most influential colleagues, differentiated between "static" and "dynamic" assessment and stated that two crucial pieces of information are missing from the former: the learner's performance with assistance and the ability to transfer skills to other tasks. In static assessment, the examinee completes the test independently without mediation. Sternberg and Grigorenko (2002, p. vii) claim that in this approach to assessment:

At some point in time after the administration of the test is over, each examinee typically receives the only feedback he or she will get: a report on a score or set of scores. By that time, the examinee is studying for one or more future tests.

In DA, however, both instruction and assessment occur simultaneously, thus helping the instructor/assessor to better comprehend students' emerging skills and not just those already acquired. The final assessment score in this approach is considered on the basis of the students' performance with assistance, while their performance in transfer tasks lets the assessor determine if skills acquired during DA have been internalized. Despite its potential utility, however, research into DA in the classroom and in more traditional teaching frameworks has remained somewhat limited as have studies on its potential impact on L2 development (Poehner & Lantolf, 2005), and especially in Arab contexts. Within these contexts in particular, it is time for practitioners with an intimate knowledge of their students and their specific learning contexts to seek assessment processes that yield greater insights into learners' cognitive development. As Norton (2007, p. 91) states, "recognition of the centrality of assessment to the learning process means that all who teach and facilitate student learning need to reflect critically on assessment practices in higher education".

In assessing descriptive writing in particular, the danger of subjectivity can be avoided when there is "dialogic mediation"—the foundation of interactionist DA—between assessor and learner. Evaluation should be fair, especially when it is in formal and high-stakes situations. In the current study, the researcher/practitioner sought to suggest an approach to assessment incorporating DA that might be effective and fair in an Omani learning context. In doing so, the research sought to address the gap that Meihami and Meihami (2014, p. 37) describe:

To date, L2 DA research has not focused on implementation of the procedure during regular classroom instruction but has instead occurred in one-to-one sessions outside the classroom and has been implemented by a teacher/researcher with expertise in applied linguistics.

With this gap foregrounded, the current research employed a DA approach to provide electronic feedback to a group of 12 EFL learners enrolled in an Omani university's English-language foundation program. Participants emailed copies of academic essays they had written as part of course requirements throughout the third semester of academic year 2012/2013 to author 1 who then offered feedback using the revision function of a commonly-used word processing program. Participants were then assessed on their ability to respond to the researcher's feedback, while a focus group interview and classroom observations were also conducted to examine the potential utility of DA in the Omani tertiary-level context.

2 Literature Review

2.1 *Vygotsky's Sociocultural Theory (SCT) and Zone of Proximal Development (ZPD)*

The influence of SCT in Second Language Acquisition (SLA) studies has become increasingly important over the course of the past two decades. Lev Vygotsky, a Russian psychologist, first wrote about SCT in the 1920s. He postulates that cognitive development occurs through "the productive intrusion of other people and cultural tools in the [developmental] process" (Newman, Griffin, & Cole, 1989, p. 68). In both individual-based and input-processing models of development, the environment contributes to development, and the focus is on determining to what extent the change can be attributed to either the individual or environment. Conversely, in SCT, the individual does not directly interact with the environment; instead, cultural artefacts and other individuals mediate interaction with it.

Vygotsky (1978) also stressed the importance of examining the learner's "zone of actual development" and "zone of proximal development" (ZPD). He defines ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86). When this concept is applied to assessment, the implication is that learners need to receive instruction in the ZPD which is ahead of their current cognitive functioning in order to promote development. Meihami and Meihami (2014, p. 36) claim that:

The significance of the ZPD is that it provides a framework for the diagnosis of learner abilities and an orienting basis for intervention to support their development. In other words, it is a means of accessing and at the same time promoting the process of development rather than focusing on its product, as happens in more conventional approaches to assessment.

A skilled mediator is thereby able to promote development or "internalization", which Lantolf (2000, p. 14) defines as "the process through which a person moves from carrying out concrete actions in conjunction with the assistance of material

artefacts and of other individuals to carrying out actions mentally without any apparent external assistance”. This mirrors Feuerstein’s Mediated Learning Theory (MLE) which stresses the importance of the mediator in altering the relationship between the environment and the learner in order to bring about development. The deliberate pitching of instruction in the ZPD, the interactivity between the learner and the mediator, and the resulting development and ability to transfer skills, are described in MLE by the terms intentionality, reciprocity and transcendence.

3 Principles of DA Realized in ZPD Theory

The principal tenet of the DA approach is encapsulated in Vygotsky’s (1978, p. 204) statement: “We must not measure the child, we must interpret the child”. DA posits that assisted performance during assessment, as opposed to independent performance, gives better insight into the development potential of the examinee. According to the principles of DA which, as discussed above, are rooted in Vygotsky’s concept of the ZPD, any assessment that does not take into account the learner’s ability to modify performance after mediation is incomplete.

A leading advocate of interactionist DA, Feuerstein’s theory of Structural Cognitive Modifiability echoes ZPD theory. The author holds that traditional tests are responsible for “ignoring a possibility that the predicted destiny may not materialize if powerful intervention takes place” (Feuerstein, Rand, & Rynders, 1988, p. 83). Lidz and Gindis (2003, p. 103) support this stance by claiming that “traditional standardized assessment follows the child’s cognitive performance to the point of ‘failure’ in independent functioning, whereas DA in the Vygotskian tradition leads the child to the point of achievement of success in joint or shared activity”. When mediation is offered in the ZPD, learners’ maturing abilities are targeted. Mediation or intervention during assessment sometimes makes development possible within a single session, which Wertsch (1985) terms “microgenesis”. Traditional psychometric assessments, on the other hand, view changes in the performance of the learner during assessment as a danger to test reliability.

Two approaches to DA can be classified as interventionist and interactionist. Poehner and Lantolf (2005) describe the former as involving a pre-test-treatment-post-test experimental approach. In this approach, the assessor offers assistance on an item-by-item basis which is based on a pre-established menu of test hints. In the interactionist approach, a skilled mediator provides assistance through dialogue that is precisely attuned to the ZPD of the individual learner. Davin (2013) claims that it is this approach to DA that is the most commonly employed within existing L2 acquisition research. Researchers concede that the mediation offered in interventionist DA does not cater precisely to the ZPD of each individual learner because the interaction is not dyadic. As Poehner and Lantolf (2013, p. 324) point out, “in developing the tests with a focus on efficiency and breath of administration, it is necessary to compromise the more fine-grained and individually negotiated modes of mediation that have characterized most L2 DA

research to date”. One way to overcome this potential issue, however, is by using interactionist DA in an online format for a reasonably large number of students—the approach adopted in the current research.

4 Studies in DA

Many studies have employed interventionist and interaction approaches to DA in both layering and sandwich formats. For example, Lantolf and Aljaafreh’s (1995) research into writing, over a period of eight weeks, involved helping learners’ deal with instruction in the use of modals, verb tenses, prepositions and articles. The authors reported the progress of learners with mediation in addition to regression until the learner Fully grasped the concept. In addition, Poehner (2007) used DA with university students studying French as a second language. The researcher employed both dynamic and static pre-tests, followed by an L2 instructional period, and concluded with dynamic and static post-tests and two transfer tasks. Poehner reported that mediation performed in this manner was able to bring to the surface problem areas with listening comprehension that might have otherwise remained hidden. In addition, in order to make better placement decisions, Anton (2009) used the interactionist method of DA to differentiate among students who obtained similar scores in the static pre-test.

Despite the potential utility of DA as reported by these studies, Davin (2013) maintains that the form of DA administration involving assessor-learner dyads is time-consuming and can therefore limit the number of participants that the mediator can potentially work with. In order to overcome this obstacle, a number of authors have sought to implement DA principles in an online format. For instance, Poehner and Lantolf (2013) adopted an online format for L2 reading and listening. Learners were checked on their ability in transfer tasks, while a non-dynamic test score, a mediated test score, and a “learning potential score” were arrived at to be used as a basis for making instructional decisions. The authors state that, by taking into consideration these influential concerns, the online application of DA can be of great benefit in L2 learning contexts. Davin (2013) also adopted a cumulative interventionist DA approach and supplemented it with instructional conversation to provide learners with a more flexible mediation attuned to their ZPD. The author highlights the complementary nature of these approaches and recommends the conjoint use of these frameworks to avoid the potential pitfalls of interventionist DA.

As the above studies suggest, mediation in DA has taken many forms and, for this reason, it is important for mediators to decide whether DA should be pre-scripted or flexible (Lantolf & Poehner, 2004) and where it should be placed. That is, this placement could be within the “sandwich” format between the pre-test and post-test, or in the “cake” format layered throughout the process of instruction and assessment (Sternberg & Grigorenko, 2002, p. 27).

In addition to these considerations, much of the SLA research on corrective feedback has tended to focus on whether it should be implicit or explicit. For

example, Ellis et al.'s (2009) survey of 11 studies on corrective feedback reports that explicit feedback is more effective for improving performance. Moreover, the authors also claim this to be especially true of lower proficiency learners. In addition, Aljaafreh and Lantolf (1994) argued for the use of both implicit and explicit feedback that employs learners' ZPD for guidance orientation. They observed that explicit feedback obscured the development process from the instructor and even perhaps inhibited the process. Hence, the important difference between feedback in SLA theory and the SCT-guided mediation offered in DA.

5 Learning Potential Score

As discussed above, one of the fundamental tenets of Vygotsky's SCT theory is the ZPD, which requires the DA practitioner to take into account the difference between independent and mediated performance. This gives the assessor insights into the ZPD of the learner or the amount of cognitive development that happens when applying the intervention strategy. Both Feuerstein, Rand, and Rynders' (1988) MLE and Budoff's (1987) theory of "Learning Potential Assessment" (LPA) support this notion. Sternberg and Grigorenko (2002) also emphasize the need to quantify the process of assessing a student's learning potential or LPA. These approaches hold promise for practitioners who need to be accountable for the scores allotted in formal assessment, especially while grading non-multiple choice and true/false answers.

In response, Kozulin and Garb (2002) created a formula to quantify an individual's learning potential. They claimed that this formula offers a theoretical basis from which to determine those students with high and low levels of learning potential. When the authors applied this formula to their study of reading comprehension, they reported significant differences between actual and mediated performance. Kozulin and Garb claimed that learners who had identical scores in the non-dynamic pre-test showed variations in their performance after mediation. This suggests the effectiveness of the formula for differentiating between learners who might perform similarly in psychometric tests, while also helping the instructor/assessor to keep track of the learner's progress.

6 Methodology

6.1 Research Questions

In order to explore the potential utility of an electronic DA approach in an Omani tertiary-level EFL foundation course, the following research questions were posited:

1. How useful is DA through electronic intervention in helping the assessor to determine the cognitive progress of a learner?
2. Is DA through electronic intervention a potentially useful formal tool for continuous assessment?
3. From the learners' perspectives, what problems, if any, are associated with using online DA feedback?

7 Participants

The participants of the present study were 12 students enrolled in the Foundation Programme of the Middle East College (MEC), the largest private college in the Sultanate of Oman. The participants were female and aged between 18 and 21. All students applying for admission to undergraduate courses at the college need to have a language proficiency equivalent to an IELTS band of 5.5 in the four tested skills. Those students who do not achieve this upon entry to the college must enroll in the 1 year English foundation programme which is aligned with the standards prescribed by Oman's Ministry of Higher Education. MEC's Foundation Programme has a three-semester structure and three tiers of English which approximate elementary, pre-intermediate and intermediate levels.

The population of level 3 students at the research site numbered 170, while the cohort chosen for this study totalled 32. This particular cohort was chosen because the first author was also the writing instructor for this group. The 12 students from this class who volunteered to participate were asked to submit their essays by email from weeks 8 to 13 of the semester. In order to recruit the 12 participants, purposive sampling was used to help determine how DA influenced the performance of high, average and low scorers on non-dynamic pre-tests. Therefore, four students from each level (high, average and low) were chosen and their non-dynamic pre-test scores, post-test scores after mediation, and learning potential scores were analyzed. The categorization of students into the three different levels of proficiency was based on several factors: non-dynamic pre-test scores, observation of students in class, face-to-face interactions, the seriousness of their errors, and uptake of corrective feedback. For example, Learners 4 and 8 have been categorized as average and high scorers respectively based on a combination of these factors, in spite of their similar scores in the non-dynamic pre-test. Since DA was used as a tool for continuous assessment, it was decided to follow the performance of the L2 participants in three cycles of pre-test, mediation and post-test in order for the assessors/researchers to accurately gauge their ZPD. Thus, each learner was assessed three times using electronic intervention DA strategies.

As one of the pre-specified outcomes that learners are required to achieve, semester three MEC foundation program students learn how to write essays of a minimum length of 250 words adhering to prescribed academic writing conventions. Given the centrality of academic writing skills to the successful completion of

the foundation program, this study examines the use of online DA in the evaluation of three essays which were written in class. In terms of the focus group interview, the researcher explained at the commencement of the study that respondents would be participating in a study investigating the feasibility and suitability of using a DA approach to evaluation by integrating electronic resources. Participants were once again reminded of the objective of the focus group discussion, after which author 1 reassured them that their grades would not be affected by the opinions they expressed and that their choice of participating or not participating in the research would in no way impact upon their standing in the course.

8 Data Collection and Analysis

Electronic intervention strategies using DA principles were implemented in an intact level 3 writing class at the research site. One of the requirements for the intermediate or level 3 group based on the Omani standards is to write an academic text with a minimum of 250 words that displays control of layout, organization, punctuation, spelling, sentence structure, grammar, and vocabulary. The instructor offered participants mediation attuned to their ZPDs using the review function of a word processing program. The instructor employed an interactionist approach since this format allowed for feedback which is more finely tuned to the needs of each learner.

The open-ended dialogic mediation that occurs in the interactionist DA approach required more effort from the mediator as the responses were not pre-scripted. Feedback ranged from explicit to implicit depending on demonstrated uptake by the individual learner. For example, if the instructor found that implicit feedback was not as effective with some students, she made it more explicit to prevent learner frustration. In these cases, once the learner exhibited some progress in response to the explicit feedback, the instructor then employed more implicit feedback to check whether the concept had been internalized. Instruction on how to write academic essays in the prescribed academic genre was provided during the scheduled six hours of weekly in-class tutorials.

Conducting several DA cycles ensures that transfer tasks are attempted with every new essay that the learner writes. The instructor is thus able to not only determine the openness of the learner to mediation but also the learner's ability to apply internalized concepts to other tasks. Therefore, the DA cycle approach towards understanding a learner's development helped the instructor to evaluate the learner better. This promoted fairness in evaluation as DA affords insights into the learner's metalanguage during mediation in the ZPD.

Minick (1989) notes that interactionist DA might make it more difficult to quantify, analyse and draw comparisons between learners. Since these interactions are not easily quantifiable, the mediator draws up a profile of each learner, detailing cognitive development and responsiveness to mediation in order to arrive at conclusions regarding the learner's ZPD. However, an attempt was made to quantify

the process of interactionist DA applied in this study by using LPS scores as calculated using Kozulin and Garb's (2002) formula:

$$\text{LPS} = \frac{(S_{\text{post}} - S_{\text{pre}}) + (S_{\text{post}})}{\text{Max } S \text{ Max } S \text{ Max } S} = \frac{(2 S_{\text{post}} - S_{\text{pre}})}{\text{Max } S \text{ Max } S \text{ Max } S}$$

In this equation, *S post* stands for learner scores in the post-test, *S pre* for scores in the pre-test and *Max S* for maximum marks that can be scored. These researchers interpreted students' LPS scores in roughly the following manner: high ($\text{LPS} \geq 1.0$), mid-level ($0.88 \geq \text{LPS} \leq 0.79$) and low scorers ($\text{LPS} \leq 0.71$) based on their learning potential scores.

The same formula was also used by Poehner and Lantolf's (2013) study of computerized DA in second language reading and listening comprehension as a way of calculating examinees' LPS scores. Although the researchers' approach was interventionist and, in this way, differed from the interactionist approach employed in the current research, it was possible to quantify the latter by considering the scores before and after mediation. Scores were calculated using the writing in-house rubric developed by the university which focuses on the four features of structure, cohesion, grammar and vocabulary. The scripts were double-marked by author 1 and another instructor at the research site to ensure reliability.

9 Results and Discussion

9.1 Essay Scores and DA

Appendix 1 features a list of the 12 participants' scores before and after electronic DA in addition to their learning potential score. Learning potential scores found in this appendix reveal that all students scoring high and average marks on the pre-test displayed a greater responsiveness to mediation than those scoring low marks as their LPSs fell into the high and mid-level ranges. Interviews with participants and observation of classroom performance revealed high levels of motivation to improve writing skills, which participants felt were crucial for academic success.

However, the LPS of around half ($n = 2$) of the poor performers on the pre-test indicated a higher openness to mediation than the other two poor scorers. For example, learner 11, who performed poorly on the pre-test, exhibited more openness to mediation in the first DA cycle than in subsequent cycles. An informal discussion with this participant during the period of classroom observation revealed that she was unable to focus on the feedback given during DA because she was busy meeting other assignment deadlines. Learner 9 showed consistency in her LPS. Interviews and class observations confirmed her high levels of motivation. Interestingly, this participant improved her performance to such a degree after mediation that her LPSs were on a par with those who scored well on the first pre-test.

The first author's informal interview with learner 12 revealed that she did not feel motivated to improve her work. She felt that her language skills were too poor, and she lacked confidence to make the required changes to her essays. She also cited a lack of familiarity with computers—the means through which essays were submitted and DA offered—as a reason for her limited progress. Hence, the intrusion of affective variables, such as the lack of confidence and motivation, may impede mediation in low performers.

Interestingly, learners 1 through 9 and learner 12 indicated an ability to transfer the skills acquired in their first DA to subsequent tasks. However, transfer studies in L2 DA carried out by Poehner (2007) and Ableeva (2010) do not report consistent improvement in learners' performance after initial DA sessions. The consistent improvement in transfer task performance observed in this study was perhaps due to the similarities between the required written tasks. That is, in the second and third tasks addressed in each DA cycle, participants had to attempt essays of the same genre which were scored using the same rubric.

An examination of poor scorers' performance indicates that affective factors, such as confidence and motivation, in addition to academic factors like a lack of study skills, interfered with mediation. Hence LPSs are a useful indication of when other types of intervention like counseling might be necessary. Since there were three DA cycles implemented in the current research, early intervention was possible and the maximum benefit accrued to the student as each individual was given the kind of intervention that was most appropriate.

10 Focus Interview

The focus group was used as a method of data collection since the researchers felt that students would be more comfortable talking as a group and that participants would be prompted to reflect more deeply on the topic being discussed when engaged in a "social context". More specifically, as Patton (2002, p. 386) explains:

Unlike a series of one-on-one interviews, in a focus group participants get to hear each other's responses and to make additional comments beyond their own original responses as they hear what other people have to say. However, participants need not agree with each other or reach any kind of consensus. Nor is it necessary for people to disagree. The object is to get high-quality data in a social context where people can consider their own views in the context of the views of others.

The focus group interview lasted about 30 min. Eight students who were a mix of high, average and low scorers volunteered to participate. Many of the students who took part in the discussion felt that it was difficult to type their essays using word processing software and that they also did not have the technical know-how to deal with sending emails. However, all participants agreed that this process helped familiarize them with these electronic tools and displayed a sense of enthusiasm about their newly-achieved electronic skills. One participant reported that one of the

results of integrating technology into the DA process was to improve students' electronic literacy skills. This echoes Warschauer and Kern's (2005) finding that an outcome of using computer networks in L2 classrooms was that "students in these classes did not experience new technologies principally as an aid to second language learning; rather, they saw themselves as developing new literacy skills in a new medium of critical importance for their lives" (p. 45).

In terms of participants' perceptions about sending essays by email, many reported that they felt they were engaging in something appropriate for university-level students. One participant even claimed that she felt "important" in front of her little brothers and sisters when doing her work electronically and receiving feedback from her teacher through email. The illegible handwriting of teachers when marking students' essays on paper was also discussed as a point of concern with more traditional written methods of feedback. Participants claimed that online DA feedback helped them to make corrections and ask for help from others. However, they were not motivated to find out what their mistakes actually were when the corrections were done on paper in the teacher's handwriting. One respondent even claimed to feel depressed when she saw red pen marks on her paper essay. A number of respondents also pointed out that electronic forms of DA enabled them to finish their work at home and get immediate feedback from the teacher without having to wait for the next class.

In addition, electronic DA allowed participants to use the spelling and grammar tools available with word processing software to check for linguistic mistakes. They were able to use these tools and therefore found the process of typing essays easier and more rewarding. Warschauer and Kern (2005, pp. 1–2) articulate this point in the following explanation: "the fact that computer-mediated communication occurs in a written, electronically archived form gives students additional opportunities to plan their discourse and to notice and reflect on language use in the messages they compose and read".

An important point of DA is that it emphasizes the importance of mediation attuned to the individual's ZPD. Indeed, students actually acknowledged that less explicit feedback made them "figure it out"—that is, encouraged them to work out their errors and how to deal with these—themselves. Thus, a major advantage of using the interactionist approach can be seen from the viewpoint of the mediation the teacher was able to provide which prompted students to self-correct. Interestingly, low scoring participants in the pre-test said that they were encouraged to write more when submissions were made online. On the other hand, the number of essays written did not vary between higher and lower scorers, thus suggesting that the electronic mode of submission and feedback may be motivating even for lower level scorers.

The following extended excerpt from the focus group interview highlights a number of these themes. Pseudonyms have been used to protect the identity of participants.

Researcher: Let us talk about the way you submit your essays nowadays.

Mariya: Teacher, the letters of keyboard were difficult at first but now easier and quickly

Lujaina: Very easy

Maha: We can correct our mistakes.

Lujaina: Spelling

Afra: Also correct mistakes

Maha: Also grammar

Researcher: What do you think, Hafsa?

Huriya: (nodding) I agree.

Noor: Because I am first year here, I did not know how to send email. Now I know.

Raya: I hate red. When I see the red pen marks on the paper, I am depressed. With online feedback the teacher says only 'grammar'. She does not say my mistake directly so I can think. I feel with the paper she is finding fault but with online I can think of my mistake. I know there is a mistake and I can figure it out myself.

Haya: I can correct my own mistakes. Sometimes I go to the teacher but mostly I am able to correct myself.

Lujaina: Sometimes I get help from my friends who help me...because they are good in English.

Afra: Writing on computer is easy.

Researcher: Why is that?

Afra: Paper is more mistake but computer corrects mistakes.

Lujaina: Not only spelling but grammar. There are red and green lines.

Maha: Green is grammar.

Researcher: Did you correct your mistakes when I gave you feedback on paper?

Lujaina: Um...no.

Mariya: But when you give online we correct. Before it was difficult typing but now easy.

Raya: We don't have to write the whole essay again, we just need to correct the mistakes.

Maha: We can read easily now because the teacher types her comments but before teacher's handwriting is difficult.

Noor: Something new for us.

Raya: Yes, that is true. Gives us the feeling that we are doing something important. For 12 years we were writing on paper and we are bored with it. My little brothers and sisters are impressed when they see me writing my essays on the computer.

Noor: Yes. Same for me.

Maha: Can finish extra work at home and send it directly at that moment. So we don't waste time. We have more time to learn extra things in class.

Researcher: Do you write more essays now than you did when you had to write on paper?

Noor: No.

Mariya: Same as before.

Afra: (smiling) Yes...because teacher checks.

Lujaina: Yes.

11 Classroom Observation

Classroom observations tended to confirm a number of trends emerging from the electronic DA and focus group interviews. For example, for very low scorers in the pre-test, it was observed that implicit feedback offered by the teacher/researcher was not effective as students found it impossible to self-correct without a large amount of direct help. Hence, these participants had to wait for an opportunity to meet the instructor face-to-face so that the item in question could be explained. However, despite this potential limitation, with some low-scoring students the researchers observed that the need to wait for instructor face-to-face time actually resulted in more peer interaction. In addition, many participants appeared to take electronically available feedback much more seriously than corrections done by hand. Consequently, they scored higher marks in the second half of the semester when it was mandatory to submit their academic essays online.

As a way of encouraging engagement with the electronic feedback, the teacher saved all three of the participants' submitted academic essays under personalized electronic folders to enable them to accurately track student submissions. This storage and tracking system was displayed to participants during class time using the digital projector so that all students could see their work saved and documented. The researchers observed that this improved students' motivation to complete their work as it was very clear that the researchers closely monitored essay submissions. Previously with the essays submitted on paper, the teacher often found it hard to keep track of various learner submissions as the learners themselves were required to maintain their own essay portfolios and to submit these only at the end of the semester. This system often resulted in the scenario wherein, at the end of the semester when teachers asked students to submit their essay folders, there was often a mad scramble to finish essays, which many students tended to write haphazardly or to copy from their more punctual peers. Instances of plagiarism were also quite common.

12 Discussion

When teachers use more traditional methods of assessing students, there is often a bias towards the number of essays completed rather than their actual quality. This is especially the case when the teacher is responsible for marking a large number of essays within a short time period. Moreover, improvements witnessed in student second drafts are often based on the teacher's general sense of the student's abilities formed from their overall performance in class. When the class size is between 32 and 35 students, this tends to encourage highly subjective evaluations. The first author reflected on her own practice at the research site and also had numerous informal discussions with other teachers regarding these. When electronic interactionist DA is used over three cycles, the assessment is based on a more thorough

understanding of the student's abilities or "openness to mediation" as the researcher could easily compare the first and second drafts because of instant access to student work saved and catalogued electronically.

Face-to-face encounters also helped the first author to evaluate those students who did not respond to implicit electronic feedback. What is noteworthy about this point from the current study is the willingness from even low scorers to seek to understand and correct their mistakes. As the students in the focus group interviews and classroom observations expressed time and again, technology for them is a great motivator. This study clearly conveys the call from the "net generation" for their instructors to integrate technology into teaching and assessment.

This study explored the feasibility of using electronic DA as an instrument of continuous evaluation based on the premise that it gives the evaluator a better grasp of the learner's development over the course of a single semester. However, perhaps studying DA cycles longitudinally over the course of a few semesters with larger sample sizes might shed light on motivational, cultural, social and other factors which could account for the degree of learner modifiability. Having a quantifiable variable like the LPS, rather than qualitative learner profiles and analysis, makes further research in this area much more feasible. In addition, an experimental research design might help researchers to understand more about the difference between traditional paper-and-pen feedback and electronic mediation based on DA principles.

13 Limitations and Conclusions

As with many such approaches to DA, the main drawback here concerns the limited number of students who can participate due to the labour-intensive nature of interactionist DA. Poehner and Lantolf's (2013) study involving interventionist DA in computerized testing does, in fact, highlight the nature of this disadvantage. It should nonetheless be noted that, in the present study, the researchers handled a sizeable number of participants over three DA cycles. However, if the number of cycles were reduced and if the cycles were more evenly spaced, with perhaps one of the cycles occurring towards the beginning of the semester and another towards the end, it would be much easier for the instructor to successfully manage the workload. Poehner and Lantolf discussed future developments in DA, and this study is a partial response to what they envisage for it—an attempt to explore more open-ended approaches to computerized DA:

Other models of DA have been proposed for computerized administration, and these may also serve as a point of departure for L2 DA researchers. We are in the beginning stages of what promises to be a lengthy and challenging process. For example, there is no reason to limit C-DA to the multiple-choice testing format. In fact, we are currently developing a language comprehension test that uses a cloze format with more open-ended types of mediation. We also see assessments that include language production tasks as an especially important, though challenging, context for C-DA (p. 337).

A number of researchers agree that mediation precisely attuned to the ZPD of the individual learner is more effective than pre-scripted prompts based on assessors' guesses about the kind of intervention learners may require during assessment. It is only fair that students should get the best possible mediation when they are going through an important assessment process. When they go through three DA cycles, it "minimizes the risk of an erroneous evaluation, by definition. It provides mediation that is constantly adjusted and attuned to the learner's or group's responsiveness to mediation. At the same time, it promotes the very development it seeks to assess in the first place" (Poehner & Lantolf, 2005, p. 252). Therefore, DA in this form can be used by the instructor to evaluate students fairly during formal assessment as it provides deeper insights into the precise stage of learner development and reveals whether there is development or regression. Moreover, all participants in the current study responded very well to the electronic mode of feedback utilized, and the researchers believe that the additional benefit of electronic literacy is another advantage of using this mode of feedback.

Appendix 1: DA Cycle Scores and LPS by Participant

Participant #1

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	67	74	1.08
2	63	75	1.16
3	70	75	1.06

High Scorer 1 in Pre-test 1

Participant #2

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	58	67	1.01
2	55	64	0.97
3	50	60	0.93

High Scorer 2 in Pre-test 1

Participant #3

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	65	70	1.00
2	70	75	1.06
3	74	75	1.01

High Scorer 3 in Pre-test 1

Participant #4

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	55	60	0.86
2	60	60	1.00
3	60	60	1.00

High Scorer 4 in Pre-test 1

Participant #5

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	48	58	0.90
2	55	65	1.00
3	58	68	1.04

Average Scorer 1 in Pre-test 1

Participant #6

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	40	50	0.80
2	45	55	0.86
3	48	58	0.90

Average Scorer 2 in Pre-test 1

Participant #7

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	48	58	0.90
2	62	65	0.90
3	65	70	1.00

Average Scorer 3 in Pre-test 1

Participant #8

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	55	60	0.86
2	58	62	0.88
3	62	65	0.90

Average Scorer 4 in Pre-test 1

Participant #9

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	25	45	0.86
2	40	50	0.80
3	48	55	0.82

Poor Scorer 1 in Pre-test 1

Participant #10

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	30	40	0.66
2	28	33	0.50
3	25	25	0.33

Poor Scorer 2 in Pre-test 1

Participant #11

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	15	40	0.86
2	30	40	0.66
3	25	40	0.73

Poor Scorer 3 in Pre-test 1

Participant #12

DA cycle	Non-dynamic pre-test	Post-test after mediation	Learning potential score
1	28	35	0.56
2	30	35	0.53
3	30	35	0.53

Poor Scorer 4 in Pre-Test 1

References

- Ableeva, R. (2010). *Dynamic assessment of listening comprehension in second language learning*. Unpublished doctoral dissertation, The Pennsylvania State University, University Park, PA, USA.
- Aljaafreh, A., & Lantolf, J. P. (1994). Negative feedback as regulation and second language learning in the zone of proximal development. *Modern Language Journal*, 78, 465–483.
- Anton, M. (2009). Dynamic assessment of advanced second language learners. *Foreign Language Annals*, 42, 576–598.

- Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 53–81). New York: Guilford Press.
- Davin, K. J. (2013). Integration of dynamic assessment and instructional conversations to promote development and improve assessment in the language classroom. *Language Teaching Research, 17*, 303.
- Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.
- Feuerstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am. Helping retarded performers excel*. New York: Plenum.
- Kozulin, A., & Garb, E. (2002). Dynamic assessment of EFL text comprehension of at-risk students. *School Psychology International, 23*, 112–127.
- Lantolf, J. P. (2000). Introducing sociocultural theory. In Lantolf, J. P. (Ed.), *Sociocultural theory and second language acquisition* (pp. 1–26). Oxford: Oxford University Press.
- Lantolf, J. P., & Aljaafreh, A. (1995). Second language learning in the Zone of Proximal Development: A revolutionary experience. *International Journal of Educational Research, 23*, 619–632.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics, 1*, 49–72.
- Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, V. S. Ageev, S. Miller, & B. Gindis (Eds.), *Vygotsky's educational theory in cultural context* (pp. 99–106). Cambridge: Cambridge University Press.
- Luria, A. R. (1961). Study of the abnormal child. *American Journal of Orthopsychiatry. A Journal of Human Behavior, 31*, 1–16.
- Meihami, H., & Meihami, B. (2014). An overview of dynamic assessment in the language classroom. *International Letters of Social and Humanistic Sciences, 5*, 35–43.
- Minick, N. (1989). *L.S. Vygotsky and Soviet activity theory: Perspectives on the relationship between mind and society*. Newton, MA: Educational Development Center.
- Newman, D., Griffin, P., & Cole, M. (1989). *The construction zone: Working for cognitive change in school*. Cambridge: Cambridge University Press.
- Norton, L. (2007). Using assessment to promote quality learning in higher education. In A. Campbell & L. Norton (Eds.), *Learning, teaching and assessing in higher education: Developing reflective practice* (pp. 92–101). Southernhay East: Learning Matters.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Poehner, M. E. (2007). Beyond the test: L2 dynamic assessment and the transcendence of mediated learning. *The Modern Language Journal, 91*, 323–340.
- Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*, 233–265.
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research, 17*(3), 323–342.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing. The nature and measurement of learning potential*. Cambridge: Cambridge University Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Warschauer, M., & Kern, R. (2005). *Network-based language teaching: Concepts and practice*. Cambridge: Cambridge University Press.
- Wertsch, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

The Future of E-assessments in the UAE: Students' Perspectives

Racquel Warner

Abstract One of the core values in education is the need to align assessments with content, skills, and knowledge in order to maintain validity and reliability. There is good research evidence to show that well designed assessment systems lead to improved student performance and ensure success of students. Electronic assessment, which is regarded as the flip side of the e-learning coin, is acclaimed by some stakeholders in UAE higher education institutions as a possible magic bullet or saviour for the evaluation of learning. Others argue that e-assessment might herald the death of assessments with high levels of reliability and validity. This qualitative study investigates teachers and students' perceptions of e-assessment within a private higher education institution (HEI) in the UAE. Through the use of a questionnaire and interviews, perceptions of virtual learning environments, e-assessment methods on virtual platforms, and the process of giving feedback on performance on e-assessments are analysed. The participants offered unique insights into the conduct of e-assessment and most were concerned about the prospect of electronic feedback replacing verbal feedback and face-to-face interaction between the lecturers and students. Most participants indicated some benefits of e-assessments to the pedagogical processes in the university, but were reluctant to express wholehearted agreement with a transition to e-assessments as a sole @ method of summative evaluation. The study concludes by recommending ways of promoting the idea of e-assessment to lecturers and students including new codes of practice, training and assurances to both stakeholders that these new methods are an improvement on previous practice and that e-assessment can actually increase reliability and validity.

Keywords E-assessment • Computer assisted assessments (CAA) • Feedback

R. Warner (✉)

Mohammed Bin Rashid School of Government, Dubai, UAE

e-mail: warnerdxb@gmail.com

© Springer International Publishing Switzerland 2017

R. Al-Mahrooqi et al. (eds.), *Revisiting EFL Assessment*,

Second Language Learning and Teaching, DOI 10.1007/978-3-319-32601-6_19

1 Introduction

The quality of the student learning experience is very central to education in the UAE because of the competitive nature of the higher education sector. Educating, engaging and satisfying an increasingly diverse student population have become a necessary part of the teachers' job description. The move toward e-learning in the Emirates has been seen as a step in the right direction toward student satisfaction. The next logical step in this trajectory is the incorporation of e-assessments in HEIs across the UAE. In the wider educational research corpus, there is extant research on the adoption of e-learning from the perspective of instructors and educational technologists (Bull & Mckenna, 2004; Stephens & Mascia, 1995; Warburton & Conole, 2003), but a dearth of research on students' perception of e-assessment. In the UAE there is little research in either area. It is apparent that students are the primary stakeholder in this matter of e-assessments and therefore it is imperative to gauge their perceptions of the matter since this has a direct impact on the internal validity of the assessment (Anastasi, 1982). Domino and Domino (2006) assert that if students lack the confidence in a particular assessment, their level of engagement and cooperation can be negatively affected.

This research was intended to explore this matter in a private higher education institution in Dubai, by asking students from one programme to participate in an online survey to gauge their perceptions of e-assessment in 3 dimensions: reliability, pedagogy and affective factors. The guiding research questions were:

- What do you consider to be e-assessment tasks?
- What assessment format do you prefer?
- What are the features of e-assessment that students like?
- What is the future of e-assessment in HEIs?

2 E-assessment in the UAE context

One of the core values in education in general is the need to align assessments with content, skills, and knowledge in order to maintain validity and reliability. There is good research evidence to show that well designed assessment systems lead to improved student performance and ensure the success of students. If this is the case, it is imperative to open a discussion about the use of e-assessments in educational institutions in the UAE because the latest ICT initiative in UAE schools has been the roll out of I-pads and other e-learning tools for all students. The UAE has taken pride in being on the cutting edge of education by using ICT to support learning. This trend is not unique to the UAE because it is a global development to embrace the latest tools in the field of information and communication. Increasingly students use tools such as word processors, graphic software, statistical software tools and online research tools as an essential part of learning. However, when it comes to assessments students are restricted to the use of the traditional pen and paper model

of assessment. While the ramifications of this bizarre practice are being explored by researchers and experienced by teachers, it is necessary that all stakeholders consider the option of using ICT to create a bridge between learning, teaching and assessment. E-assessment is considered by some as just the flip side of the e-learning coin. Gipps (1994) reasons that “if teaching and its associated resources become electronic, then assessment too will need to take that route, to ensure alignment between the modes of teaching and assessment” (p. 26).

This research was conducted at a branch campus of a UK university in Dubai at which e-assessments have been slowly integrated into most disciplines either through formative or lower weighted summative tasks. The specific group of students among whom the data was collected are new students at the university who come from diverse secondary school backgrounds. Lecturers who use e-assessments report many positive benefits such as increased involvement by students, and increased participation and submissions in online tasks. Convenience and flexibility of marking remotely and providing feedback are additional benefits cited by staff. Students’ perceptions have scarcely been investigated, hence the purpose of this research is to find out what students know about e-assessments and how they feel about them.

3 Literature Review

A pivotal component of any educational process is an assessment process which can be used to draw conclusions about the curriculum, what adjustments need to be made in the classroom, and most importantly inform students about their progression. In best practices, assessments reflect the core values of an educational institution because well designed assessments have been correlated with improved student performance in numerous research studies (Conole, 2003). The emergence of ICT, and by extension, exponential technological advancement has created a link between teaching, learning and assessment. This development has also been fuelled by internationalization of education, which has made the need for e-assessments more urgent.

Apart from the aforementioned developments at a technical level, there are scathing and critical reports about the gap between what employers expect and the skills graduates are presenting upon application. This warrants the rethinking of the entire assessment process in higher education to ensure that students are being equipped with transferable skills which will be useful for the future.

Technological advancement throughout this decade has raised the expectations of most stakeholders in Higher Education and changed the pedagogical process immensely. It is evident that the continuous innovation in technology has impacted teaching and learning and is likely to continue to do so in the future. One noticeable result of advanced technology was the availability of Massive Open Online Courses which competed with traditional universities for students. Universities are now coming to terms with the fact that they need to be more flexible in their evaluation and assessment processes and are now exploring and realizing that there is some

value in e-assessments. Traditional paper based assessments have been superseded by e-assessments and students regard traditional forms of paper based tests as being obsolete (Parshall, Spray, Kalohn, & Davey, 2002).

4 Definition of Electronic Assessment (E-assessment)

The commonly held understanding of electronic assessment is the use of computers and computer software to assess learners' work. JISC (2007) define e-assessment as "the end-to-end electronic assessment processes where ICT is used for the presentation of assessment activity and the recording of responses" (p. 6). The E-assessment Association Scotland (2014) defines e-assessment as:

a single term [which] describes a range of learning and assessment activities that have distinct meanings in their own contexts e.g. electronic marking, online assessment, computer-aided assessment and direct on-screen testing are all referred to as e-assessment.

At present e-assessment is currently used in most HEIs for the delivery of online tests and examinations, online submission and marking, plagiarism detection, e-Portfolio assessment, and assessment of contributions to asynchronous and synchronous discussions. Benson (2003) believes that the underlying principles of assessments do not change in an electronic environment and that the same issues of validity, reliability, fairness and flexibility are all still relevant factors to be considered when using e-assessments.

Higher education institutions globally are beginning to realize not just the value of e-assessment but the vital need for it in order to keep abreast with growth, changing demographics of students and pedagogical needs. The traditional paper based forms of assessments are quickly fading into obscurity as e-learning advances into classrooms, providing digitally literate students with greater interactivity and creativity. Pangali (2003) suggests this increasingly computer literate student population will drive up demand for e-assessment in education and training.

5 E-assessment Enhances Quality Assurance Through Timely and Constructive Feedback

Nicol (2010) posits that "Assessment and feedback practices should be designed to enable students to become self-regulated learners, able to monitor and evaluate the quality and impact of their own work and that of others". E-assessment has the potential of increasing the quality of support students receive from their lecturers because it creates an easier means of providing timely and constructive feedback. The literature of computer assisted assessment (Bull & McKenna, 2004; Gilbert,

Gale, Warburton, & Willis, 2009; Pachler, Mellar, Daly, Mor, & Wiliam, 2009; Whitelock & Brasher, 2006) indicates that automated immediate formative feedback can be one of the key benefits of e-assessment. According to the QAA ‘Institutions should ensure that appropriate feedback is provided to students on assessed work in a way that promotes learning and facilitates improvement’ (QAA Code of Practice for the assurance of academic quality in higher education, Section 6 May 2006). Gipps (2008) believes the importance of feedback as part of the pedagogical process in education cannot be overstated. Research has shown that by increasing the provision of formative feedback, performance standards in assessment show a positive increase and the concepts and content of lectures are more clearly understood (Clarke, Lindsay, McKenna, & New, 2004; Black & William 1998a, 1998b).

Extant research about feedback suggests that for it to be effective feedback should be timely, provide constructive information to help with learning, and be related to explicit learning outcomes and assessment criteria that have previously been discussed with students (Black, Harrison, Lee, Marshall, & Wiliam, 2002; Cowan, 2009; Sadler 1998).

E-assessment gives lecturers the capacity to provide feedback to students on their learning in a timely manner. Comments can be written throughout the text and links to useful resources that can help students improve in the area of weakness can be appended to an electronically submitted paper. Students do not need to wait on an appointment to see their lecturer anymore and can simply use the online resources suggested. This however does not reduce teacher student interaction, but on the contrary lecturers can more actively monitor students’ use of electronic feedback through discussion boards or social networking platforms.

6 Regulation of E-Assessment

One concern expressed by many educators about e-assessments is usually the need to ensure fairness, validity and reliability. These features of assessments do not change in an electronic environment (Booth & Ainscow, 2002). E-assessments are highly regulated and there are industry standards set to govern the administration of these types of assessments. The Qualifications and Curriculum Authority in the UK is just one quasi governmental body that publishes e-assessment regulatory principles that governs UK HEIs (QCA, 2007). The degree to which an institution follows this guideline, however, is out of the immediate control of the regulators. The development of policies and mechanisms that govern e-assessment will undoubtedly require collaboration among all stakeholders in order to ensure the integrity of e-assessments when used either for formative or summative tasks.

7 Practical Considerations of E-assessments Stir Debate

Any discussion about assessments will always explore the benefits and the drawbacks. The same, by extension, is true of the e-assessment debate which triggers very contrasting views. Supporters of e-assessment have posited that the advantages include immediate feedback to students and staff, enhanced learning opportunities through knowledge tracking, real time evaluation of the strengths and weaknesses of a course, and convergence with other computer-based or online materials. It has also been pointed out by James, McInnis, and Devlin (2002) that e-assessment can provide more complex scenarios such as computer simulations, images and sounds, with which students can interact. Paper based tests do not have this component. According to Hamilton and Shoen (2005), web-based testing has significant advantages in the areas of cost, ease of use, reliability, replicability, scoring, aggregating results, and data management.

Some who are opposed to e-assessment argue that interactive assessment activities are time-intensive to produce in much the same way as interactive learning activities are and make additional demands on institutional resources and support. They contend that entertainment and assessment are mutually exclusive and regardless of how interactive e-assessments are, a lack of computer skills among some learners could negatively influence results. Brosnan (1999, pp. 48–49) posits that:

computer anxiety can lead to simplification of conceptual reasoning, heightened polarisation and extremity of judgement and pre-emption of attention, working memory and processing resources. Individuals high in computer anxiety will therefore under-perform in computer-based versions of assessment.

Another disadvantage of e-assessments is the time it takes to set up the assessments to ensure reliability, validity and ease of administration. To achieve these standards, it is essential to equally monitor the quality of question design and assessment scoring methods with the same rigour as regulation of student conduct in the assessment. As with any assessment, poor question construction, inconsistent grading of tasks and questions, or assessments that are inaccessible to some candidates, can adversely affect the results. Effective procedures for e-assessment begin by ensuring that the assessment design is fit for the purpose of the assessment, revising examination regulations and information given to candidates, and then providing appropriate training for all staff involved in invigilation and technical support. This huge investment of preparation time is viewed as a negation of the efficiencies e-assessment claims to bring to the teaching and learning environment, and something which Brown, Bull, and Pendlebury (1997) see as a profound change in working practices for academics.

Another point raised by detractors of e-assessment is the issue of authentication of students' work for electronic coursework submissions. This could add to the interval of time between submission and giving feedback, as lecturers will have to spend time holding viva voce with students who they suspect have not submitted their own work. According to the Joint Information Systems Committee, some

universities in the UK are experimenting with biometric technology to verify students' identities, but this is still in its infancy (JISC, 2010).

At a TESOL Arabia conference in 2013, Dr. Michelle Estable of the Higher Colleges of Technology in Dubai pointed out the practical considerations such as IT infrastructure, accessibility and security, which had to be taken into account to enable successful use of e-assessments in HEIs in the UAE. Institutions wishing to have e-assessments must have a sufficiently robust information technology system available for students who are required to engage in e-assessment. They must also, as a priority, ensure the reliability of e-assessments. In so doing the confidence of lecturers and students in using the available technology in assessments will increase. To run a reliable e-assessment, the university needs to have on hand IT specialists, system back-ups and contingencies in case the technology fails.

The system must be able to simultaneously handle large amounts of data traffic and large numbers of users without crashing. Ensuring that students have sufficient access to the facilities at a given time for the e-assessments is important if the e-assessment is going to be fair to all test takers. To ensure security of assessment content, summative, or high stakes assessments, should not be placed on the IT infrastructure where students could access them. Additionally, the results need to be posted securely, thus, ensuring privacy and confidentiality for each student. It is obvious that formative or diagnostic assessments require less institution wide support and can be managed by individual academics or departments. A growing body of evidence indicates that well designed and well-deployed diagnostic and formative e-assessments can foster more effective learning for a wider diversity of learners (Nicol, 1997; Sharpe, Benfield, & Francis, 2006). However, summative e-assessment requires a larger degree of institutional contingencies in case of technical glitches, hacking and system shutdowns. In case of these occurrences there could be legal ramifications which would have to be written into policy documents and contracts with external providers.

E-assessments done in most HEIs are currently run using a VLE software such as Blackboard Vista or Moodle. This software is familiar to students and lecturers because they are also used in regular classes. Many students however, report systems crashes during submission times when the VLE software cannot handle the heavy traffic and sometimes it does not allow submission. This becomes a problem when the submissions are for high stakes summative tasks. More frequent low stakes uses of VLE for e-assessment, however, involves asynchronous and synchronous online discussions, online submissions and marking, plagiarism detection and e-Portfolio assessment.

8 Cost Effectiveness of E-assessments

Hall and LeCavalier (2000) summarized the cost of firms converting their traditional training delivery methods to e-learning. IBM saved US \$200 million in 1999, providing five times the learning at one-third the cost of their previous methods.

Ernst and Young reduced training costs by 35 % while improving consistency and scalability by using a blend of Web-based (80 %) and classroom (20 %) instruction. The same can be extrapolated for the education sector if universities switched to sustainable methods of e-learning and e-assessments. Barron (2001) observes that e-learning technology providers have been increasingly able to provide cost-savings and broader benefits, develop integrated offerings, and propose innovative ways of applying e-learning.

In addition to generally positive economic benefits, other advantages such as convenience, standardized delivery, self-paced learning, and variety of available content, have made e-learning a high priority for educational institutions. Increasingly, more students are learning with the use of software, word processors and digital media. Assessment systems need to take this into account and allow students to use ICT to demonstrate their knowledge and skills. Assessments that fail to do so are maladaptive. Much of the discussion about implementing e-assessment has focused on the technology, but as Driscoll (2002) points out, e-learning is not just about the technology, but also many human factors. The demographic descriptors for students have changed and assessments need to be accessible to a range of students on different sites and for their different rates of progress. With e-assessment these variables can be accounted for.

9 The Research Methodology

The aim of this research is to establish the generally held perceptions of students about e-assessments. The research is largely descriptive and data was collected without an established hypothesis. This was done because, as educational technology is still in its infancy, there was little objective information about perceptions of e-assessment on which to form a hypothesis. Given the degree of subjectivity surrounding this topic, care was exercised in designing a survey that avoided personal assumptions. The research approach that most ideally suits this study can best be described as qualitative in nature. However, the data extracted is also expressed quantitatively using descriptive data.

The main data collection instrument was an online survey because of the convenience with which students in the population could access the survey once it was available online. In addition, given the topic of e-assessment, it was felt that this medium of conducting the survey was congruent with the topic.

The items for the survey can be divided into five dimensions of questions which aimed at operationalizing the topic of students' perceptions of e-assessment:

1. Knowledge of e-assessment
2. Comparison of e-assessment and traditional paper-based assessments
3. Challenges with e-assessment
4. Student assessment preferences
5. Predictions about the role of e-assessment in education

10 Sample

The online survey was completed by 154 pre-university students. The sample contained a fairly equal distribution of 75 male and 79 female respondents. Having both genders represented in the sample gives an opportunity to see the gender balance in the perceptions about e-assessment. The age range of the sample was 17–23 years old with a mean of 19 years and a median value of 20. This 1-year difference between the mean and median age suggests that the data is evenly distributed among foundation students who mostly fall between the ages of 18–20 years. The students on the foundation programme are all expected to participate in e-assessment across all modules, so it was useful to gauge their perception.

11 Results and Discussion

11.1 Knowledge of E-assessment

In terms of students' knowledge of e-assessment, it is obvious from Fig. 1, that students have a very general idea of what e-assessment is. The data shows that equal numbers of students regard all options as e-assessments as those who regard only computer based test and coursework graded online as e-assessment. This is perhaps a perception that has been transferred from traditional assessment where tests, presentations and coursework are considered assessment, but journals and portfolios are not. Additionally, this data might reflect the current types of assignments which students submit as part of their e-assessment (Fig. 2).

In the graph above there is only a 7 % difference between students who have participated in e-assessments before and those who have not. This data is interesting as it shows that increasingly students from all over the world are using e-assessment before entering university. The respondents in this study represent a diversity of cultural backgrounds from Africa, India, Pakistan, Russia, UK and USA. Prensky (2001) describes the generation of students who are now in university as digital natives.

What should we call these “new” students of today? Some refer to them as the N-[for Net]-gen or D-[for digital]-gen, but the most useful designation I have found for them is Digital Natives. Our students today are all “native speakers” of the digital language of computers, video games and the Internet (p. 1).

Figure 3 indicates that students feel that e-assessments are best used for low stakes formative and practice tasks for in course grades but not for summative high stakes assessments. It is clear from the data that students generally have a very positive attitude towards e-assessment when used for formative tasks because these usually provide feedback which according to the literature positively correlate to improved performance (Black & William, 1998a, 1998b). So, in terms of face validity and from students' self-reporting, it can be argued that formative tasks are

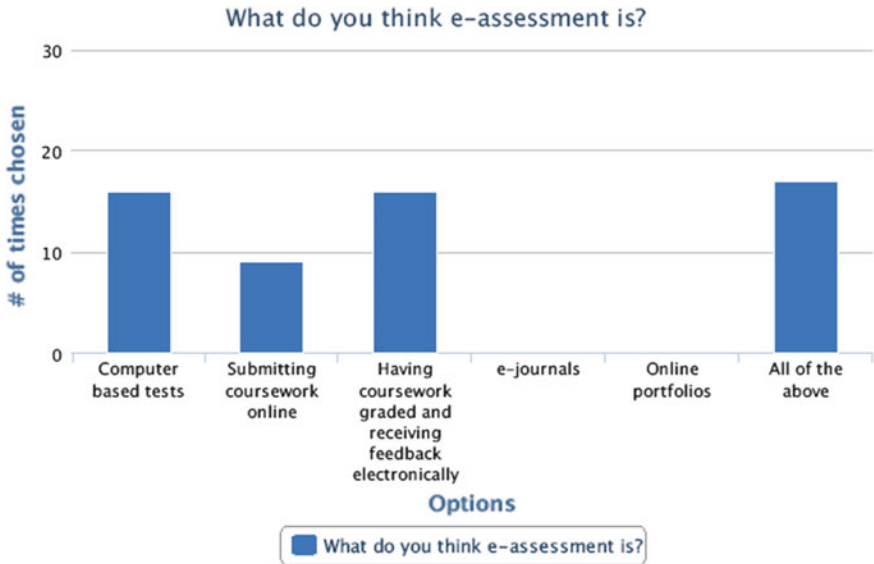
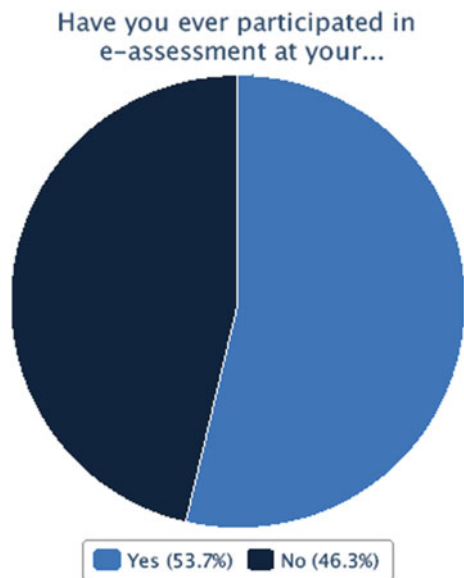


Fig. 1 Perceptions of e-assessment

Fig. 2 Prior experience with e-assessment



perceived to help with studies, and the feedback is thought by the students to be useful. Pellegrino, Chudonsky and Glaser (2001) suggest that formative assessments prioritize performance and focus less on student responses. As a result, many institutions are anchoring their e-assessment activities into meaningful scenarios so

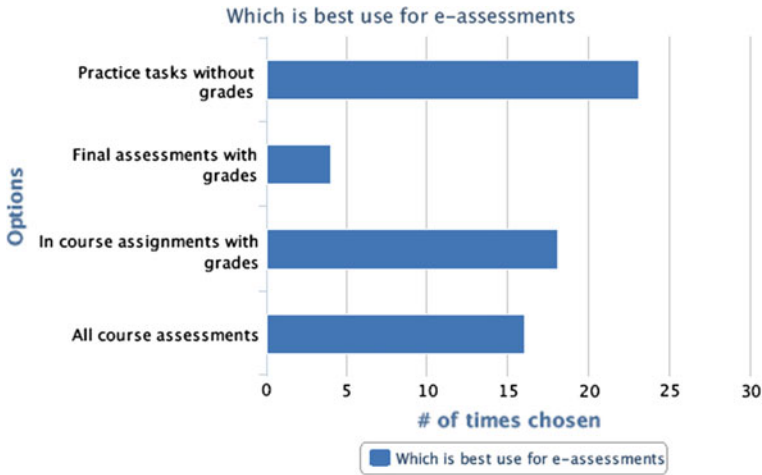


Fig. 3 Tasks best suited for e-assessment

that students are being assessed on their abilities to apply learning in authentic situations. When formative assessments are offered in interactive online formats, there is evidence that students voluntarily resit the tests and that the flexibility of being able to complete these formative tasks anywhere and at any time can help students to establish more regular patterns of study. Innovations in formative assessments on mobile devices designed for disaffected learners or those studying in workplace environments have proven to be a valuable means of engaging learners (Attewell, 2005).

12 Comparison of E-assessment and Traditional Paper Based Assessment

The uncertainty indicated in Fig. 4 about the suitability of e-assessment as an evaluation tool in university is a reflection of the wider sentiments about this topic. Only 25.93 % of the respondents were categorically positive about the use of e-assessments as an evaluation tool. The majority of the respondents (59.26 %) share the sentiments of the wider academic community of scholars, researchers, practitioners who think there is value but are hesitant to adopt e-assessment as a holistic approach to assessments.

In Fig. 5, 64 % of the respondents believe e-assessment is a better tool for providing feedback to students than traditional paper based assessment. This sentiment is a popular one across the Higher and Further Education sectors, and institutions need to pay attention to this increasing student demand for better assessment experiences and improved quality and efficiency in academic

Fig. 4 E-assessment as an evaluation tool

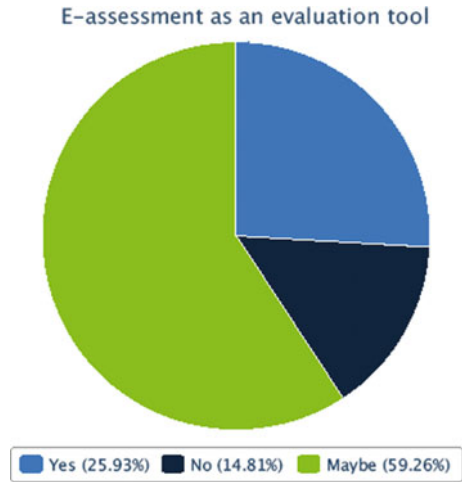
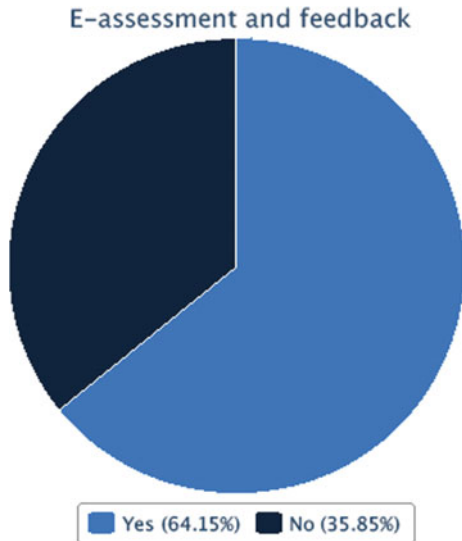


Fig. 5 E-assessment and feedback



administration. As mentioned in the literature review of this study providing timely and constructive feedback is a concern of HEIs which must be addressed as part of quality assurance and quality enhancement, especially in the context of fee paying international institutions and branch campuses. Most of these institutions use e-submission software such as Turnitin, which provide a Grademark tool option.

In terms of comparison between paper based and electronic assessments, the data in Fig. 6 shows students perceived paper based assessments to be more costly and more time consuming to prepare and administer. On the issue of cost, the initial expenditure for electronic equipment (e.g. computers), reliable software, bandwidth, specialized

	Cost	Time to prepare	Time to administer	More interactive
Paper based assessments	76%	77%	59%	31%
E-assessments	24%	23%	41%	69%

Fig. 6 Comparison table of students opinions

manpower, training, maintenance, security, trouble-free operations and the time involved in developing test questions can be very high (Scottish Qualifications Authority, 2004). So, e-assessment tools can be expensive to implement; however, once created, e-assessment applications become less expensive, being easy to operate and score, and re-useable from year to year in different combinations (Ridgway & McCuster, 2003). Moreover, students think e-assessments are more interactive. This clearly puts e-assessment ahead in students’ mind for being more engaging. Positive perception has long been established in psychological theory as a necessary precursor to positive behavior. If students perceive e-assessments to be more beneficial when compared to paper based ones, they will be more receptive to the former. If properly integrated into the curriculum and optimized for interactivity, e-assessments will be a preferred option for students. Research has shown that students prefer e-assessment to paper-based assessment because “the users feel more in control; interfaces are judged to be friendly; and because some tests use games and simulations, which resemble both learning environments and recreational activities” (Richardson et al., 2002, p. 635).

As the capabilities of ICT increase, new interactive options are emerging for students’ knowledge, as well as practical and critical skills to be assessed. In addition, computers are now able to undertake aspects of the marking of these sophisticated responses which enables timely and constructive feedback. “Such innovations push the boundaries of technology and pedagogy and inevitably result in assessments which have no paper-test equivalent” (Winkley, 2010, p. 20). The current technology allows students to interact with simulations and submit the results and analysis for assessment. Webpage design and peer evaluation is another interactive option created by the use of ICT in assessment. The introduction of a voice or sound element widens the reliability of e-assessments to include auditory learners. Some of these assessment activities can be offline assessment of material delivered through e-learning and could include the assessment of a presentation on the web or Prezi. Students’ online skills could be assessed by allowing them to demonstrate the use of particular software packages such as databases and spreadsheets (JISC, 2006). Is it any wonder then that students perceive

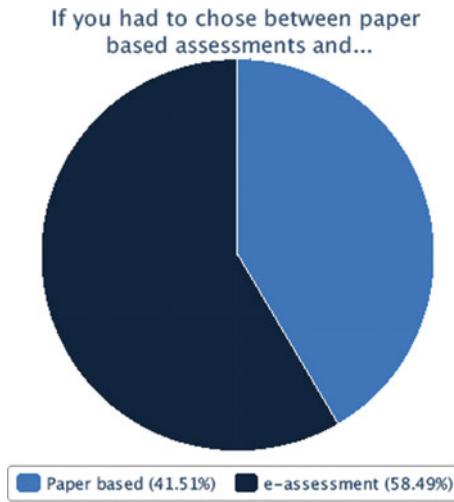


Fig. 7 If you had to choose between paper based assessments and e-assessment

e-assessments to be potentially more interactive and therefore as seen in Fig. 7 58 % of them indicated a preference for it as an assessment option?

As mentioned before, e-assessments have the ability to incorporate many interactive elements, but the primary reasons, as seen in Fig. 8, for students preferring e-assessment are the quick feedback and ease of correcting errors.

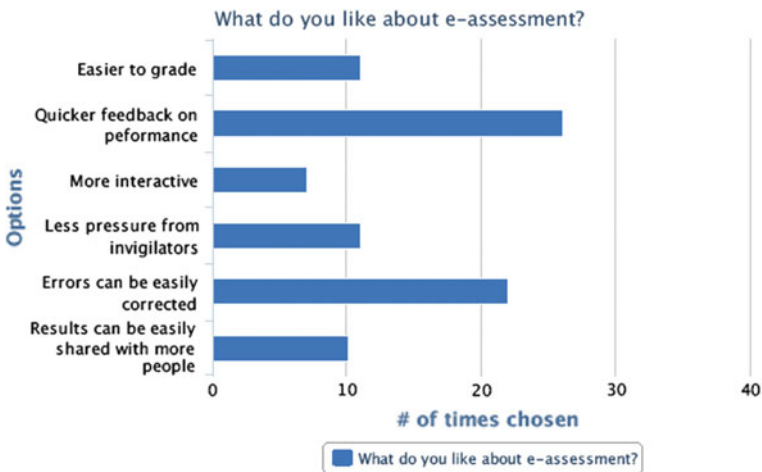


Fig. 8 Reasons for e-assessment preference

13 Challenges with E-assessment

In responding to questions about the challenges they perceive in e-assessments 35 % of the students felt internet connectivity was the greatest obstacle to the reliable delivery of the assessment. While this might be true from a purely practical standpoint, there are a number of other challenges that moving to e-assessments will bring; switching to an electronic mode of assessments will inevitably change the pedagogical processes in education and also challenge existing assumptions about the didactic process in formal education. Students will, no doubt, gain more control of many aspects of the assessment process if this is offered on demand and at the rate of progression or level of academic standard the students possess. This development could represent a considerable challenge to the formal education system, but may increase the motivation of more students to persevere (JISC, 2007) (Fig. 9).

Another challenge that will increase with e-assessment becoming more popular is the need to ensure a fully inclusive and fair experience. Objective tests must provide built-in accessibility features that are customizable for each candidate in advance. This is especially the case when teaching staff are responsible for question authoring. Item bank questions also require careful scrutiny and indexing so that each candidate experiences a test of equal measure to their ability, regardless of the combination of questions presented to them in a randomly generated test. This of course will be time consuming. E-assessment may reduce marking time dramatically, but the overall time spent on administering the assessment will not be reduced since the focus of effort and time will only shift to before the assessment period.

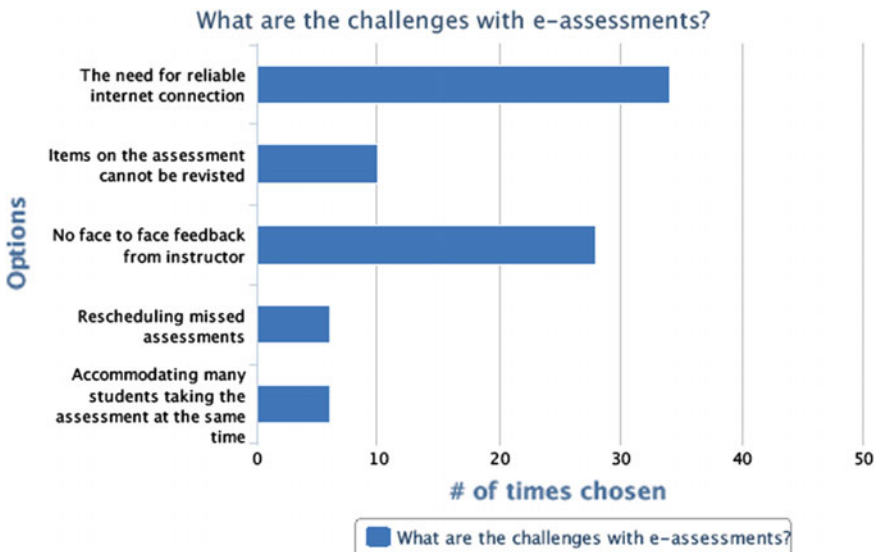


Fig. 9 Challenges with e-assessments

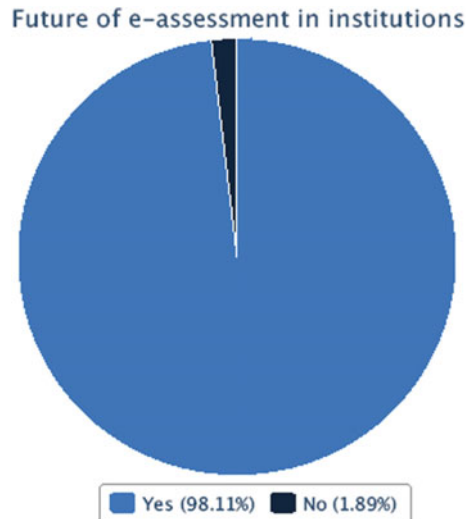
What is clear is that understanding of the potential of e-assessment tools and systems needs to be matched by awareness of their pedagogic benefits and pitfalls. Perhaps the greatest challenge to be faced over the next decade is choosing the best and most appropriate ways of using these tools (JISC, 2007).

14 Predictions of Future Use

In Fig. 10, 98 % of the respondents felt it was inevitable that most institutions would be using e-assessments at some point in the future, and by extension, they seem to be predicting the end of paper-based assessment. This transition will require the education sector to find new ways of assessing 21st Century skills such as creativity, problem solving, communication and collaboration, which are now all part of the PISA framework (OECD 2014).

Additionally, Bennett (2002) suggests that technology is central to learning and, as a result, is going to prove in the future, to be central to the assessment process. He further explains that technology will not only facilitate testing but also support authentic assessment. He refers to e-learning as part of the equipment of 21st Century scholarship and cites the success of online universities and virtual high schools in the United States. It is evident that the type of assessments required to diagnose and assess whether students have acquired 21st Century Skills will not be paper based, so it seems to be a reasonable perception that students have about the future increase in the use of e-assessments.

Fig. 10 Future of e-assessment in institutions



15 Conclusion

The present study has yielded some useful information about students' perceptions on e-assessment and while the findings may not be generalizable, they are certainly transferable and can be used to guide lecturers as they explore more efficient methods of assessment. Given the formative nature of this study in the UAE context, further research into other stakeholders' perceptions of e-assessment could generate a more complete picture from which a proper hypothesis could be drawn. The current study only surveyed a small group of the overall student population in the UAE so a more expansive study which includes government and other private institutions of Higher Education could yield some very useful data about this topic.

E-assessments will remove the tedium associated with traditional examinations for all types and age groups of learners. For many, including those with cognitive and some physical disabilities, e-assessment can offer a richer, more engaging, and a potentially more valid assessment experience than paper-based testing. The challenge is to make more use of this motivational potential in high-stakes examinations. The key benefits of e-assessment as reported by students in this study are the immediacy of feedback, improved assessment validity, increased flexibility, and more efficient and sustainable administration of assessments; all of which resulted in students' overwhelmingly positive perceptions towards this type of assessment.

E-assessment has a transformative potential in 3 key areas.

1. Coping with large student numbers on different geographical sites

While e-assessment skilled teachers do not have the limitations compared to e-assessments, the practicalities of teacher–student ratios and the availability of immediate targeted feedback make formative e-assessment a very powerful tool for practical individualization of assessment and learner empowerment.

2. Creation of improved assessment modes

One intractable debate in the education sector is whether e-assessments provide more detailed and immediate information about assessment modes. This uncertainty has prompted institutions to be more responsive and proactive in managing their assessment cycles. One undeniable variable about e-assessment is that data can be readily generated for reporting purposes, improving the quality of the assessment, and providing statistical support for a particular approach to student evaluation. Certainly the immediacy with which this data can be generated can improve response times for determining modes of assessment that are not reliable or valid.

3. Expanded role of e-assessment

It is expected that Higher Education institutions will set the pace for all other sectors in education in terms of e-assessments. There will, almost certainly be a major shift in its use, from being used occasionally to becoming the standard for assessment at all stages, from primary to postgraduate. The prevalence and relatively low cost of e-assessments will help to promote better standards and make

educational institutions more responsive to changes in technology, knowledge and society.

At present, most assessments are teacher created. In the future, teachers will use ICT tools to create and customize their own assessments, designed with their own specific purpose and students in mind, thereby ensuring that assessment is fully embedded as a formative resource in the curriculum. When e-assessment first started, it was seen as a low-grade means of automating yes/no and similar question types. Now, it has a major role in all areas of education—placement, diagnostics, assessment for learning, diversity and inclusiveness, pedagogy, summative assessment and awards, quality assurance, and individualization—bringing immediacy, thoroughness, reliability and validity. This expanded role of assessments needs to be responsive to educational and industry needs. E-assessments present an adaptive component to evaluation in order to meet the expanded demands of society.

E-assessment has developed very rapidly over the last decade but, one remaining barrier is wider adoption of ICT in schooling. Students have perceptually made the shift but institutions are lagging behind. Technology is changing the assessment paradigm not only by lowering the cost but also opening the possibility of global e-assessment. Examination Boards are no longer geographically limited and are now able to offer their services to anyone, anywhere in the world. Assessments are a universal experience in people's lifetime, so the prospects of this age-old process being made more fair, accurate, helpful, available and engaging are very exciting indeed.

References

- Anastasi, A. (1982). *Psychological testing*. New York: Macmillan Publishing.
- Attewell, J. (2005). *Mobile technologies and learning: A technology update and m-learning project summary*. Retrieved from <http://www.m-learning.org/docs/The%20m-learning%20project%20-%20technology%20update%20and%20project%20summary.pdf>.
- Barron, T. (2001). *An e-learning industry update. ASTD learning circuits*. Retrieved from www.learningcircuits.org.
- Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment*, 1(1). Retrieved from <http://www.jtla.org>.
- Benson, A. (2003). Assessing participant learning in online environments. *New Directions for Adult and Continuing Education*, 100, 69–78.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2002). *Assessment for learning: Putting it into practice*. London: Oxford University Press.
- Black, P., & William, D. (1998). Inside the black box: Raising standards through Classroom assessment. *Phi Delta Kappan*, 80(2), 139–144.
- Booth, T., & Ainscow, M. (2002). *Index for inclusion: Developing learning and participation in schools* (Rev ed.). Bristol: CSIE.

- Brosnan, M. J. (1999). Modeling technophobia: A case for word processing. *Computers in Human Behaviour*, 15, 105–121.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Bull, J., & McKenna, C. (2004). *Blueprint for computer assisted assessment*. London: Routledge Falmer.
- Clarke, S., Lindsay, K., McKenna, C., & New, S. (2004). Inquire: A case study in evaluating the potential of online MCQ tests in a discursive subject. *Journal of the Association of Learning Technology*, 12(3), 249–260.
- Conole, G. (2003). Time to grow up—Discipline and methodological issues in learning technology research. Keynote speech at *Shock of the Old Conference—Designing and Developing for the Disciplines*, Oxford University, 24th July.
- Cowan, E. M. (2009). Implementing formative assessment: Student teachers experiences on Placement. *Teacher Development*, 13(1), 71–84.
- Domino, G., & Domino, M. L. (2006). *Psychological testing*. New York: Cambridge University Press.
- Driscoll, M. (2002). Blended learning: Let's get beyond the hype. *LTI Magazine*. Retrieved from <http://elearningmag.com/ltimagazine/article/articleDetail.jsp?id=11755>.
- E-assessment.com. (2007). *Guide to effective practice*. Retrieved from http://www.e-assessment.com/wp-content/uploads/2014/08/e-assessment_-_guide_to_effective_practice_full_version.pdf.
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Routledge Falmer.
- Gipps, C. (2008). Socio-cultural aspects of assessment. *Student Assessment and Testing*, 1(8), 252–291.
- Gilbert, L., Gale, V., Warburton, B., & Willis, G. (2009). *Report on summative e-assessment quality (REAQ)*. Joint Information Systems Committee (JISC): Southampton.
- Hall, B., & LeCavalier, J. (2000). *E-Learning across the enterprise: The benchmarking study of best practices*. Retrieved from <http://www.brandon-hall.com/elacbenstu.html>.
- Hamilton, D., & Shoen, E. (2005). Same song, second verse: Evaluation and improvement of an established assessment program. In K. Martell & T. Calderon (Eds.), *Assessment of student learning in business schools: Best practices each step of the way* (pp. 138–153). Tallahassee, Florida: Association for Institutional Research.
- James, R., McInnis, C., & Devlin, M. (2002). *Assessing learning in Australian universities*. Canberra: Australian Universities Teaching Committee. Retrieved from <http://www.cshe.unimelb.edu.au/assessinglearning>.
- Joint Information Systems Committee (JISC). (2006). *E-assessment: An overview of JISC activities*. Retrieved from www.jisc.ac.uk/uploaded_documents/ACFC6B.pdf.
- Joint Information Systems Committee (JISC). (2007). *In their own words: Exploring the learner's perspective on e-learning* [online]. Retrieved from <http://www.jisc.ac.uk/media/documents/programmes/elearningpedagogy/iowfinal.pdf>.
- Joint Information Systems Committee (JISC). (2010). *Effective assessment in a digital age: A guide to technology-enhanced assessment and feedback*. Bristol: author.
- Nicol, D. J. (1997) Research on learning and higher education teaching. UCoSDA Briefing Paper 45, Sheffield, Universities and Colleges Staff Development Agency.
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education*, 35(5), 501–717.
- OECD.org. (2014). *PISA 2012 results in focus: what 15 year olds know and what they can do with what they know*. Retrieved May 9, 2011 from <https://www.oecd.org/pisa-2012-results-overview.pdf>
- Pachler, N., Mellar, H., Daly, C., Mor, Y., & Wiliam, D. (2009). *Scoping a vision for formative e-assessment*. Joint Information Systems Committee (JISC). Bristol: JISC.
- Pangali, C. (2003). Using technology to measure learning. *Chief Learning Officer*, 2(7), 36.

- Parshall, C. G., Spray, J. A., Kalohn, J. C., & Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, 9(5), 1–6.
- Quality Assurance Authority (QAA). (2006). *Code of practice for the assurance of academic quality and standards in higher education section 6: Assessment of students*. Retrieved from http://dera.ioe.ac.uk/9713/2/COP_AOS.pdf.
- Qualifications and Curriculum Authority (QCA). (2007). *Meeting the challenge: achieving equality for all*. Retrieved April 16, 2013 from <https://core.ac.uk/download/files/161/4157615.pdf>
- Richardson, M., Baird, J., Ridgway, J., Ripley, M., Shorrocks-Taylor, D., & Swan, M. (2002). Challenging minds? Students' perceptions of computer-based world class tests of problem solving. *Computers in Human Behavior*, 18(6), 633–649.
- Ridgway, J., & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education: Principles, Policy and Practice*, 10(3), 309–328.
- Sadler, D. R. (1998). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Scottish Qualifications Authority. (2004). *SQA guidelines on e-assessment for schools*. Glasgow: Hanover House.
- Sharpe, R., Benfield, G., & Francis, R. (2006). Implementing a university e-learning strategy: Levers for change within academic schools. *Journal of the Association of Learning Technology (ALT-J)*, 14(2), 135–151.
- Stephens, D., & Mascia, J. (1995). *Results of a survey into the use of computer-assisted assessment in institutions of higher education in the UK*. Loughborough: Flexible Learning Initiative.
- Warburton, B., & Conole, G. (2003). CAA in UK HEIs—The state of the art. *Proceedings of the 7th International Computer-assisted Assessment Conference* (pp. 433–441). Loughborough: Loughborough University. Retrieved from <http://www.lboro.ac.uk/service/ltd/flicaa/conferences.html>.
- Whitelock, D., & Brasher, A. (2006). Roadmap for e-assessment. *Joint Information Systems Committee (JISC)*. Bristol: JISC.
- Winkley, J. (2010). E-assessment and innovation. *Becta*. Retrieved from http://www.alphaplusconsultancy.co.uk/pdf/e-assessment_and_innovation.pdf.