ICME-13 Monographs

Denisse R. Thompson
Megan Burton
Annalisa Cusi
David Wright *Editors*

Classroom Assessment in Mathematics

Perspectives from Around the Globe

ICME13
Hamburg 2016

Springer

ICME-13 Monographs

Series editor

Gabriele Kaiser, Faculty of Education, Didactics of Mathematics, Universität Hamburg, Hamburg, Germany

Each volume in the series presents state-of-the art research on a particular topic in mathematics education and reflects the international debate as broadly as possible, while also incorporating insights into lesser-known areas of the discussion. Each volume is based on the discussions and presentations during the ICME-13 Congress and includes the best papers from one of the ICME-13 Topical Study Groups or Discussion Groups.

More information about this series at http://www.springer.com/series/15585

Denisse R. Thompson · Megan Burton
Annalisa Cusi · David Wright
Editors

Classroom Assessment in Mathematics

Perspectives from Around the Globe

Springer

Editors
Denisse R. Thompson
Department of Teaching and Learning
University of South Florida
Tampa, FL
USA

Megan Burton
Auburn University
Auburn, AL
USA

Annalisa Cusi
University of Turin
Turin
Italy

David Wright
Research Center for Learning and Teaching
Newcastle University
Newcastle upon Tyne
UK

ISSN 2520-8322 ISSN 2520-8330 (electronic)
ICME-13 Monographs
ISBN 978-3-319-73747-8 ISBN 978-3-319-73748-5 (eBook)
https://doi.org/10.1007/978-3-319-73748-5

Library of Congress Control Number: 2017964250

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

Preface

This monograph is an extension of the discussions and presentations shared in Topic Study Group (TSG) 40 on Classroom Assessment for Mathematics Learning that occurred during the 13th International Congress on Mathematical Education (ICME-13) held in Hamburg, Germany, in July 2016. Co-chairs of that Topic Study Group were Karin Brodie (South Africa) and Denisse R. Thompson (United States), with committee members Leonora Díaz Moreno (Chile), Nathalie Sayac (France), Stanislaw Schukajlow (Germany), and IPC liaison Elaine Simmt (Canada).

The goal of TSG 40 was "to share and build research relating to assessment *for* and *as* learning in mathematics classrooms" (http://www.icme13.org/files/tsg/TSG_40.pdf). With this goal in mind, contributions (papers, oral presentations, and posters) were solicited "relating to teaching mathematics in classrooms as well as practices in teacher education and professional development that address issues on assessment for learning and teaching mathematics. How teachers learn to assess for learning and how their learning is enacted is key to developing assessment for learning that enables teachers to gain insight into students' thinking to guide further instruction" (http://www.icme13.org/files/tsg/TSG_40.pdf).

Seven themes were identified as specific areas where contributions might be developed that would fit within the broad aims of the TSG:

- "The enactment of classroom practices that reflect current thinking in assessment for learning or assessment as learning in mathematics (for example, giving feedback, developing classroom conversations, peer or self-assessment);
- The development of pre-service and in-service teachers' professional knowledge or practices related to assessment for learning mathematics;
- The enactment of practices in teacher education and professional development that reflect current thinking relative to assessment for learning and assessment as learning;
- The development of assessment tasks that reflect the complexity of mathematical thinking, problem solving, and other important mathematical competencies;
- The design of alternative modes of assessment for learning (e.g., online, investigations, forms of formative assessment);

- The development of assessment practices that support equity or enhance access to the learning of mathematics;
- The enactment of practices to ensure that curriculum, instruction, and classroom assessment are well aligned." (http://www.icme13.org/files/tsg/TSG_40.pdf)

Not all themes were equally represented across the contributions. Within the sessions assigned to various aspects of the TSG during ICME, 12 papers, 12 oral presentations, and 14 posters were presented. Two of the papers were presented as part of a joint session with TSG 39 on Large-Scale Assessment and Testing in Mathematics Education. More information about the overall structure of the various sessions can be found in the *Proceedings of ICME-13*.

Development of this Volume

At the conclusion of the Congress, three Topic Study Group participants (Megan Burton, Annalisa Cusi, and David Wright) joined with one of the co-chairs (Denisse R. Thompson) to serve as the editorial panel responsible for overseeing the development of this post-Congress monograph. Given the relatively small number of overall contributions to the various themes of the TSG, a general call was made to authors of papers, oral presentations, or posters to expand their original contribution for the monograph. Fifteen of the original contributors chose to make that investment of time and submitted a revised and expanded version of their presentation made in Hamburg.

All submissions underwent a review process in which they were reviewed by two members of the monograph's editorial panel as well as one other potential contributing author. Reviews were returned to authors with guidelines and suggestions for revisions needed to strengthen the paper and make it acceptable for publication in the monograph. Two authors chose not to make the requested revisions, primarily because their research was not far enough along to enable the requested revisions to be made. Revised papers were reviewed again by members of the monograph's editorial panel and edited as needed.

Structure of the Volume

The remaining contributions, together with an introductory paper and a concluding paper, provide insight into various assessment practices from educators and researchers around the globe. Under no circumstances would we claim that the papers in this volume provide a complete picture of assessment practices in various countries. It is not even clear that they provide a representative picture of the types of practices that teachers use around the globe. Rather, they provide a glimpse into possible assessment practices, the types of information or data to be collected from those practices, and the potential for that information or data to inform further instruction.

The authors of the papers hail from eleven different countries. Thus, the papers provide a glimpse into the extent to which issues surrounding classroom assessment, particularly formative assessment, are increasingly important regardless of the schooling or cultural context.

The papers expanded from conference contributions have been grouped into four main categories:

- Three papers provide examples of classroom assessment in action. The paper by Swan and Foster focuses on designing curriculum and assessment lessons to encourage communication and problem-solving. Pai considers how teachers deal with in-the-moment assessment actions, and Straumberger investigates how self-assessment might be used by students to enhance their own individual mathematics practice.
- Four papers illustrate how technology can be used as a tool to facilitate formative classroom assessment, regardless of schooling level. For instance, the paper by Downton focuses on how digital flip cameras can help primary teachers capture assessment practices of young children so they can explore them in more detail with the children. The paper by Cusi and colleagues uses software, tablets, and interactive whiteboards to help teachers enhance formative assessment when working with fifth-grade children to engage them in problem-solving and explaining their reasoning. At the other end of schooling, the paper by Nagari-Haddif and Yerushalmy considers how online assessments can be used to understand the thinking of high school students in calculus, and the paper by Platz and colleagues uses an e-proof environment within tertiary education to assist students in developing their skills in constructing mathematical proofs.
- Two papers focus on how statistical models might assist with formative assessment; both were originally presented in the joint session with TSG 39 on large-scale assessment. Using Rasch modeling or Cognitive Diagnostic Assessment, the authors of the two papers explore how assessments related to problem-solving or concept development can inform teachers so that appropriate instructional interventions could occur.
- The final four papers address different perspectives to engage teachers in formative assessment. Sayac investigates the assessment practices of French primary teachers, under the assumption that one needs to understand teachers' assessment practices to develop professional development to enhance those practices. Andrade-Molina and Moreno illustrate how national curricular guides, together with national assessments, can send mixed messages to teachers about the nature of learning and the types of classroom assessments that can support that learning. Burton and colleagues describe five different pedagogical approaches used in preparing teachers and in professional development settings and how teacher educators might highlight the formative assessment practices that are naturally linked to the instruction within those pedagogical approaches.

Wright and colleagues share the work of a large-scale assessment project within Europe that has developed a toolkit to assist teachers as they work to integrate formative assessment into their regular classroom instruction.

The volume's introductory paper attempts to set the stage for the importance of classroom assessment by contrasting formative and summative assessment practices. Members of the two ICME-13 TSGs on assessment joined together to prepare a Topical Survey, *Assessment in Mathematics Education: Large-Scale Assessment and Classroom Assessment* (Suurtamm et al. 2016), that represents an overview of the state of assessment and the interactions of classroom and large-scale assessment as of Spring 2016. Rather than repeat the information in the introductory chapter, readers are referred to that volume for research issues related to (1) purposes, traditions, and principles of assessment; (2) design of assessment tasks; (3) classroom assessment in action; (4) interactions of large-scale and classroom assessment; and (5) enhancing sound assessment knowledge and practices. Each of the five sections in that volume concludes with a list of questions for possible future work.

The concluding paper in this volume looks across the various papers to consider what lessons can be learned from the various models of assessment practices and to consider how those lessons might suggest future areas of research. The fact that the papers are authored by researchers in many countries highlights the importance of cross-national and cross-cultural research studies so that we can learn from each other.

Potential Audience for This Book

This volume is applicable to a wide audience. Classroom teachers might read the volume for ideas about research initiatives and practices in other parts of the world that can be applied to their own context. Researchers might use the volume to contemplate areas for additional research. Mathematics teacher educators and professional development providers might use the volume, perhaps in conjunction with the Topical Survey on Assessment, as a supplement in a course in the preparation of teachers or the enhancement of teachers' instructional practice.

Acknowledgements

We thank all those authors who wrote and reviewed manuscripts for this publication. In addition, we appreciate the encouragement of Gabriele Kaiser, series editor and convener of ICME-13, to support the sharing of contributions from ICME with

the global mathematics education community, and especially with those unable to travel to Hamburg in 2016.

Tampa, USA — Denisse R. Thompson
Auburn, USA — Megan Burton
Turin, Italy — Annalisa Cusi
Newcastle upon Tyne, UK — David Wright

Reference

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., Silver, E., Ufer, S., & Vos, P. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. (ICME-13 Topical Surveys.) SpringerOpen.

Contents

Editors and Contributors

About the Editors

Denisse R. Thompson is Professor Emeritus of Mathematics Education at the University of South Florida in the U.S., having retired in 2015 after 24.5 years on the faculty. Her research interests include curriculum development and evaluation, with over 30 years of involvement with the University of Chicago School Mathematics Project. She is also interested in mathematical literacy, in the use of children's literature in the teaching of mathematics, and in issues related to assessment in mathematics education. She served as co-chair of Topic Study Group 40 on classroom assessment at ICME-13. In addition, she is a co-editor of the series *Research in Mathematics Education*, published by Information Age Publishing.

Megan Burton is an Associate Professor and the Elementary Education Program Coordinator at Auburn University, Alabama (USA). She teaches and advises undergraduate and graduate students in elementary education and conducts research related to elementary mathematics education, with focus on elementary teacher change, inclusion, and rural education. As a former elementary teacher with experience in inclusion and English Language Learners, she is committed to classrooms that allow all students to encounter strong mathematics instruction in meaningful, differentiated ways.

Annalisa Cusi graduated in Mathematics at Modena and Reggio Emilia University in 2001, where she obtained a Ph.D. in Mathematics in 2009. She has been teaching mathematics and physics in upper secondary school since 2001. She worked as a Research Fellow at the University of Turin from 2014 to 2016 within the European Project FaSMEd. Her main research interests are innovation in the didactics of algebra; the analysis of teaching/learning processes, with a focus on the role played by the teacher; methods to promote early algebraic thinking in young students; teacher professional development; and formative assessment processes in mathematics.

David Wright is Senior Research Associate: Research Centre for Learning and Teaching Newcastle University (United Kingdom) (now retired). He has 15 years of experience in teaching mathematics at secondary, further, and higher education as an Associate Lecturer in the Open University. He was Subject Officer for Mathematics for the British Educational Communications and Technology Agency (Becta) for 4 years and 10 years in initial teacher education and research at Newcastle University. He is the Scientific Director of the European Union research project: Formative Assessment in Science and Mathematics Education (FaSMEd).

Contributors

Melissa Andrade-Molina Faculty of Engineering and Science, Aalborg University, Aalborg, Denmark

Wanda Audrict TLLR, Inc, Stone Mountain, GA, USA

Jonathan D. Bostic Bowling Green State University, Bowling Green, OH, USA

Megan Burton Auburn University, Auburn, AL, USA

Jill Clark Research Centre for Learning and Teaching, Newcastle University, Newcastle upon Tyne, UK

Annalisa Cusi Department of Philosophy and Education, University of Turin, Turin, Italy

Ann Downton Faculty of Education, Monash University, Surrey Hills, VIC, Australia

Leonora Díaz Moreno Department of Mathematics, Faculty of Science, Valparaiso University, Valparaiso, Chile

Colin Foster School of Education, University of Leicester, Leicester, UK

Miriam Krieger Institute for Mathematical, Natural Sciences and Technology Education, University of Flensburg, Flensburg, Germany

Chap Sam Lim School of Educational Studies, Universiti Sains Malaysia, Gelugor, Penang, Malaysia

Valerie L. Mills Oakland Schools, Waterford, MI, USA

Francesca Morselli Department of Mathematics, University of Genova, Genoa, Italy

Galit Nagari-Haddif Faculty of Education, University of Haifa, Haifa, Israel

Engelbert Niehaus Institute of Mathematics, University of Koblenz-Landau, Landau, Germany

Jimmy Pai University of Ottawa, Ottawa-Carleton District School Board, Ottawa, ON, Canada

Marjorie Petit Ongoing Assessment Project, Moretown, VT, USA

Melanie Platz Institute for Mathematics Education, University of Siegen, Siegen, Germany

Cristina Sabena Department of Philosophy and Education, University of Turin, Turin, Italy

Nathalie Sayac Laboratoire de Didactique André Revuz, Université Paris-Est Créteil, Livry-Gargan, France

Carolyn Jia Ling Sia School of Educational Studies, Universiti Sains Malaysia, Gelugor, Penang, Malaysia

Edward A. Silver University of Michigan, Ann Arbor, MI, USA

Toni A. Sondergeld School of Education, Drexel University, Philadelphia, PA, USA

Waldemar Straumberger University of Bielefeld, Bielefeld, Germany

Marilyn E. Strutchens Auburn University, Auburn, AL, USA

Malcolm Swan School of Education, Centre for Research in Mathematics Education, University of Nottingham, Nottingham, UK

Denisse R. Thompson College of Education, University of South Florida, Tampa, FL, USA

Lucy Tiplady Research Centre for Learning and Teaching, Newcastle University, Newcastle upon Tyne, UK

Kathrin Winter Institute for Mathematical, Natural Sciences and Technology Education, University of Flensburg, Flensburg, Germany

David Wright Research Centre for Teaching and Learning, Newcastle University, Newcastle upon Tyne, UK

Michal Yerushalmy University of Haifa, Haifa, Israel

Part I
Introduction to the Volume

Chapter 1
Formative Assessment: A Critical Component in the Teaching-Learning Process

Denisse R. Thompson, Megan Burton, Annalisa Cusi and David Wright

Abstract This introductory paper to the volume contrasts formative assessment with summative assessment and describes the importance of formative assessment to classroom instruction. In particular, it argues that a task is formative to the extent that data from the task are used to enhance and inform further instruction rather than simply to provide an evaluation of a student or of instruction. The use of design research as a mechanism to develop sound classroom assessment is outlined because a design science framework provides a means to tie together varied exemplars of innovations in assessment. A cycle of task implementation and revision can lead to improved assessment practices.

Keywords Design research · Formative assessment · Summative assessment Evaluation

---

D. R. Thompson (✉)
University of South Florida, Tampa, FL 33620, USA
e-mail: denisse@usf.edu

D. R. Thompson
College of Education, EDU105, Tampa, FL 33620, USA

M. Burton
Auburn University, 5020 Haley Center, Auburn, AL 36849, USA
e-mail: megan.burton@auburn.edu

A. Cusi
University of Turin, Via Tamburini 45, 42122 Reggio Emilia, Italy
e-mail: annalo@tin.it

D. Wright
Research Center for Learning & Teaching, Newcastle University,
Newcastle upon Tyne NE1 7RU, UK
e-mail: wrightdavidg@gmail.com

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_1

1.1 Introduction

For much of the general public, including parents and politicians, assessment is often synonymous with tests. But assessment can and should be much more than just a test. In fact, one way to define *assessment* in mathematics is "as the process of gathering evidence about a student's knowledge of, ability to use, and disposition toward, mathematics and of making inferences from that evidence for a variety of purposes" (National Council of Teachers of Mathematics [NCTM] 1995, p. 3). In contrast, *evaluation* is "the process of determining the worth of, or assigning a value to, something on the basis of careful examination and judgment" (NCTM 1995, p. 3). Tests, then, are a means of evaluation, and evaluation is just one aspect of assessment.

The tension implicit in the previous paragraph reflects the fact that assessment has both formative and summative perspectives. A given assessment task can be either formative or summative, depending on how the information gathered from that task is used. If an assessment task is used for accountability purposes, at the individual student level or to make value judgments about the quality of education in a school or country, then that assessment task is *summative*; most large-scale external assessments or classrooms assessments used at the end of a unit of study fit within this category. However, when assessment tasks are used to collect insight into students' thinking that can inform the teacher or the students about their learning which is then used to guide further instruction, the assessment task is *formative*; tasks and activities that help move students' thinking forward and help guide teachers as they make instructional decisions fit within this side of the assessment coin.

Too often, assessment is viewed as something that occurs at the end of a unit of study or a specific time period. However, assessment "that enhances mathematics learning becomes a routine part of ongoing classroom activity rather than an interruption. … [and is] an integral part of instruction that encourages and supports further learning" (NCTM 1995, p. 13). The papers in this volume take this view of assessment—as an ongoing and integral part of instruction to enhance the learning of students.

1.2 The Role of Formative Assessment in the Classroom

Black and Wiliam (2009) describe formative assessment in terms of decisions made based on the assessment rather than on the actual collection of information from the assessment. Assessment is formative

> to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded than the decisions they would have taken in the absence of the evidence that was elicited. (p. 9)

As noted by Wiliam, this definition means that formative assessment necessitates "that one is clear about what it is that students are to learn, but it does not impose a particular view of the mathematics curriculum, nor does it entail any particular view of what happens when learning takes place" (2015, p. 250). That is, a determination of the nature of an assessment depends on how information from that assessment is used. A given task, even an end-of-unit test, could be formative if it is used to guide instruction or help teachers determine how to move students' learning forward, but could be summative if it is used solely to provide a grade.

The definition of formative assessment posited by Black and Wiliam poses a challenge for teachers, educators, and researchers. To gain the type of information needed to make effective instructional decisions, cognitively demanding tasks are needed that focus on conceptual understanding rather than just surface knowledge. Identifying and developing such tasks is not only a challenge for teachers, but is also a challenge for students who are asked to think mathematically in ways that involve more than just procedures and to explain their thinking in multiple ways—via pictures, words, symbols, or in some other format. Students and their teachers need many opportunities to engage with such tasks to develop an appreciation for the extent to which they can facilitate the learning process.

Over the last three decades, in particular, there has been a recognition around the globe of the need to engage many more students in mathematics, and to ensure that all students have an opportunity to be successful. As a consequence, mathematics educators in many countries have emphasized the importance of a student-centered classroom rather than just a teacher-centered or teacher-directed one. Formative assessment is a critical component of shifting to a student-centered perspective because it places the student at the center of the assessment process, through having students assess their own learning as well as supporting the learning of classmates. Black and Wiliam stress that, together with the teacher and the learner himself, fundamental agents in the assessment processes are the peers. Peers can challenge learners to reflect on their own thinking, helping them "to make unconscious processes overt and explicit and so making these more available for future use" (2009, p. 19). As Leinwand and colleagues note, "an important goal of assessment should be to make students effective self-assessors, teaching them how to recognize the strengths and weaknesses of past performance and use them to improve their future work" (2014, p. 95). Through both self-assessment and peer assessment of present and past performance, students become the center of the instruction and assessment cycle.

1.3 Design Research in Classroom Assessment

The report *Knowing What Students Know* (Pellegrino et al. 2001) identifies progress in the science of designing assessments as a key factor in enhancing classroom assessment. The report provides a range of assessment examples and steers the analysis of them towards a science of design:

> while it is important to carefully analyze each of the examples as a separate instance of innovative design, they also need to be analyzed as a collective set of instances within a complex 'design space.' The latter can be thought of as a multivariate environment expressing the important features that make specific instances simultaneously similar and different. (Pellegrino et al. 2001, p. 304)

Developments in design science in recent years (Barab and Squire 2004; Bereiter 2002; Burkhardt 2006; Cobb et al. 2003; DBRC 2003; Kelly 2003; van den Akker et al. 2006) provide a clearer view of what might be required for the design of effective assessments. The principles of design research can be described as:

> a formative approach in which a product or process (or 'tool') is envisaged, designed, developed and refined through cycles of enactment, observation, analysis and redesign, with systematic feedback from end-users. Educational theory is used to inform the design and refinement of the tools, and is itself refined during the research process. Its goals are to create innovative tools for others to use, to describe and explain how these tools function, account for the range of implementations that occur, and develop principles and theories that may guide future designs. Ultimately, the goal is transformative; we seek to create new teaching and learning possibilities and study their impact on end-users. (Wright et al. 2017, this volume as adapted from Swan 2014)

Examples within the papers in this volume provide windows into the different perspectives of the design process as researchers attempt to develop innovations in assessment occupying the complex design space identified in *Knowing What Students Know*. Teaching itself has also been characterized as a design science (Laurillard 2012) with technology and assessment playing crucial roles in improving practice. Hence, design research appears to provide a guiding framework for the development of assessment tasks and resources and might be adopted as a strategic approach for further research into assessment practices. A design framework provides one means to tie together different papers in this volume with their varied perspectives on formative assessment. As teachers take small steps in changing their assessment practice, reflect on the benefits and challenges of those changes, and then try again, they are actually engaging in aspects of design science (Suurtamm et al. 2016).

1.4 The Ongoing Nature of Formative Assessment

As noted in Suurtamm et al. (2016), the current climate in mathematics education encourages teachers to focus students' learning on both content and process and to ensure that students have robust mathematical proficiency consisting of appropriate skill proficiency, understanding of concepts, ability to reason, and productive attitudes towards learning mathematics. Research with Canadian teachers as well as with Finnish teachers has found that a focus on the use of formative assessment has encouraged teachers to view assessment as a social practice that becomes a natural part of the daily life of the classroom. As teachers move toward ongoing assessment practices that engage students in demonstrating robust mathematical proficiency, they often face a number of dilemmas: conceptual dilemmas relate to viewing assessment as more than an end-of-unit result; pedagogical dilemmas focus on how to develop and implement ongoing assessment opportunities; cultural dilemmas address challenges faced by teachers and students when assessment practices change from the established practices in a schooling environment; and political dilemmas arise as teachers' assessment practices interact with district or national assessment practices (Suurtamm and Koch 2014). Although not characterized as such, the papers in this volume reflect various ways in which teachers and researchers have addressed one or more of these dilemmas.

References

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1–14.

Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society, 9*(3), 321–327.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Burkhardt, H. (2006). From design research to large-scale impact: Engineering research in education. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 121–150). London, UK: Routledge.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

DBRC [The Design-Based Research Collective]. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Kelly, A. (2003). The role of design in educational research. *Educational Researcher, 32*(1), 3–4.

Laurillard, D. (2012). *Teaching as a design science*. London, UK: Routledge.

Leinwand, S., Brahier, D. J., Huinker, D., Berry, R. Q., III, Dillon, F. L., Larson, M. R., et al. (2014). *Principles to actions: Ensuring mathematical success for all*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Research Council, National Academy Press.

Suurtamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: Experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability, 26*(3), 263–287.

Suurtamm, C., Thompson, D. R., Kim, R. Y., Moreno, L. D., Sayac, N., Schukajlow, S., et al. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*, ICME-13 Topical Surveys. SpringerOpen.

Swan, M. (2014). Design research in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 148–152). Dordrecht, The Netherlands: Springer.

van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational design research*. London, UK: Routledge.

Wiliam, D. (2015). Assessment: A powerful focus for the improvement of mathematics instruction. In C. Suurtamm & A. Roth McDuffie (Eds.), *Assessment to enhance teaching and learning* (pp. 247–254). Reston, VA: National Council of Teachers of Mathematics.

Wright, D., Clark, J., & Tiplady, L. (2017, this volume). Designing for formative assessment: A toolkit for teachers. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe* (pp. 207–228). Cham, Switzerland: Springer International Publishing AG.

Author Biographies

Denisse R. Thompson is Professor Emeritus of Mathematics Education at the University of South Florida in the U.S., having retired in 2015 after 24.5 years on the faculty. Her research interests include curriculum development and evaluation, with over thirty years of involvement with the University of Chicago School Mathematics Project. She is also interested in mathematical literacy, in the use of children's literature in the teaching of mathematics, and in issues related to assessment in mathematics education. She served as co-chair of Topic Study Group 40 on classroom assessment at ICME 13. In addition, she is a co-editor of the series *Research in Mathematics Education*, published by Information Age Publishing.

Megan E. Burton is an Associate Professor and the elementary education program coordinator at Auburn University, Alabama (USA). She teaches and advises undergraduate and graduate students in elementary education and conducts research related to elementary mathematics education, with focus on elementary teacher change, inclusion, and rural education. As a former elementary teacher with experience in inclusion and English Language Learners, Burton is committed to classrooms that allow all students to encounter strong mathematics instruction in meaningful, differentiated ways.

Annalisa Cusi graduated in Mathematics at Modena and Reggio Emilia University in 2001, where she obtained a Ph.D. in Mathematics in 2009. She's been teaching mathematics and physics in upper secondary school since 2001. She worked as a research fellow at the University of Turin from 2014 to 2016 within the European Project FaSMEd. Her main research interests are innovation in the didactics of algebra; the analysis of teaching/learning processes, with a focus on the role played by the teacher; methods to promote early algebraic thinking in young students; teacher professional development; and formative assessment processes in mathematics.

David Wright is Senior Research Associate: Research Centre for Learning and Teaching, Newcastle University (United Kingdom) (now retired). David has fifteen years' experience teaching mathematics at secondary, further and higher education as an associate lecturer with the Open University. He was Subject Officer for Mathematics for the British Educational Communications and Technology Agency (Becta) for four years and ten years in initial teacher education and research at Newcastle University. He is the Scientific Director of the European Union research project: Formative Assessment in Science and Mathematics Education (FaSMEd).

Part II
Examples of Classroom Assessment in Action

Chapter 2
Formative Assessment Lessons

Malcolm Swan and Colin Foster

Abstract Formative assessment is the process by which teachers and students gather evidence of learning and then use this to adapt the way that they teach and learn in the classroom. In this paper, we describe a design-research project in which we integrated formative assessment into mathematics classroom materials. We outline two examples of formative assessment lessons, one concept-based and the other problem-solving, highlighting the important roles within them of pre-assessment, formative feedback questions, and sample work for students to critique.

Keywords Conceptual understanding · Formative assessment · Problem solving Mathematics task design · Teacher professional development

2.1 Introduction

High-quality *formative* classroom assessment has the potential to produce substantial student learning gains (Black et al. 2003; Black and Wiliam 1998, 1999, 2009). We follow Black and Wiliam's definition that:

> Practice in a classroom is *formative* to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions

---

Malcolm Swan: Deceased 24 April 2017.

This paper is based on a plenary given at the 37th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (Swan 2015).

M. Swan
School of Education, Centre for Research in Mathematics Education,
University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, UK

C. Foster (✉)
School of Education, University of Leicester, 21 University Road, Leicester LE1 7RF, UK
e-mail: colin.foster@leicester.ac.uk

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_2

> they would have taken in the absence of the evidence that was elicited. (2009, p. 9, original emphasis)

Designing mathematics lessons that embed high-quality formative assessment practices could lead to better learning in those lessons, and could also play a part in supporting teachers in developing their formative assessment practices more widely in other mathematics lessons.

In 2010, with funding from the Bill and Melinda Gates Foundation, we began the Mathematics Assessment Project (MAP) to support US middle and high schools in implementing the new Common Core State Standards for Mathematics (CCSSM).[1] These standards place a renewed focus on conceptual understanding and on the development of practices[2] (or processes) that should permeate all mathematical activity. In this project, we explored the research question: How can well-designed materials enable teachers to make high-quality formative assessment an integral part of the implemented curriculum in their classrooms, even where linked professional development support is limited or non-existent?

This ambitious goal was motivated by four empirical findings. First, professional development support is, in practice, in most places, sharply limited in quantity and in the quality of its leaders, and currently few have much deep experience of formative assessment. Second, the development of formative assessment expertise through professional development needs a program that lasts at least two years for significant impact (e.g., Wiliam et al. 2004). Third, most mathematics teachers rely on teaching materials, even when on familiar ground; thus, it is unreasonable to expect them to face the greater challenges of "adaptive expertise" (Hatano and Inagaki 1986) within formative assessment without well-engineered support. Finally, it is our experience that teachers, like students, learn strategies best through constructive generalization of principles from specific high-quality experiences. We see these lessons as supporting such experiences—as well as providing a 'protein supplement' to a generally carbohydrate curriculum diet. It was our goal that over time teachers transfer some aspects of these strategies into their existing practice, with or without the professional development support for which the project also developed materials. There is now some evidence of this happening (see Sect. 2.6).

The MAP project developed over 100 formative assessment lessons, called *Classroom Challenges*. Each lesson consists of student resources and an extensive teacher guide.[3] In this paper, we describe the research-based design of these materials and outline two examples, one concept-based and the other focused on

[1]See http://www.corestandards.org/Math/.

[2]The eight CCSSM Standards for Mathematical Practice are: (i) Make sense of problems and persevere in solving them; (ii) Reason abstractly and quantitatively; (iii) Construct viable arguments and critique the reasoning of others; (iv) Model with mathematics; (v) Use appropriate tools strategically; (vi) Attend to precision; (vii) Look for and make use of structure; and (viii) Look for and express regularity in repeated reasoning.

[3]These lessons are available free on the website, http://map.mathshell.org.

Table 2.1 Key strategies of formative assessment

	Where the learner is going	Where the learner is right now	How to get there
Teacher	1. Clarifying learning intentions and criteria for success	2. Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding	3. Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	4. Activating students as instructional resources for one another	
Learner	Understanding and sharing learning intentions and criteria for success	5. Activating students as the owners of their own learning	

supporting students' problem-solving strategies. We highlight across both kinds of lessons the important roles of pre-assessment, formative feedback questions, and pre-designed sample work for students to critique. We also articulate how the interaction between the three agents of formative assessment (teachers, learners, and their peers) and the three main aspects of formative assessment (identifying where learners are in their learning, where they are going, and how to bridge the gap) is manifest in the lesson structures (see also Cusi et al. and Wright et al. this volume). This interaction, or the "five key strategies" of formative assessment (Table 2.1), was first articulated by Wiliam and Thompson (2007). Finally, we offer evidence from early independent studies that suggest that these lessons can indeed enable teachers to redefine the classroom contract.

2.2 Theoretical Framework for Task Design

The CCSSM make it clear that the goals of the new curriculum are to foster a deeper, connected conceptual understanding of mathematics, along with the strategic skills necessary to tackle non-routine problems. In our work, we found it necessary to distinguish between lessons that are designed to foster conceptual development and those that are intended to develop problem-solving strategies. In the former, the focus of student activity is on the analysis and discussion of different *interpretations* of mathematical ideas, while in the latter the focus is on discussing and comparing alternative *approaches* to problems. The intention was that concept lessons might be used partway through the teaching of a particular topic, providing the teacher with opportunities to assess students' understanding and time to respond adaptively. Problem-solving lessons were designed to be used more flexibly—for example, between topics—to assess how well students could select already familiar mathematical techniques to tackle unfamiliar, non-routine problems, and thus provide a means for improving their strategic awareness. Importantly, for this to be effective, the technical demands of the task must be low, to allow processing capacity for students to focus on the strategic aspects of the problem. To this end, we recommend that these lessons should depend on content taught up to two years previously.

The tasks that we selected for the concept-based *Classroom Challenges* were designed to foster collaborative sense-making. Sierpinska (1994) suggests that people feel that they have understood something when they have achieved a sense of order and harmony, where there is a 'unifying thought' of simplification, of seeing an underlying structure and a feeling that the essence of an idea has been captured. She lists four mental operations involved in understanding:

> identification: we can bring the concept to the foreground of attention, name and describe it; discrimination: we can see similarities and differences between this concept and others; generalisation: we can see general properties of the concept in particular cases of it; synthesis: we can perceive a unifying principle. (p. 32)

To these, we add the notion of representation; when we understand something, we are able to characterize it in a variety of ways: verbally, visually and/or symbolically.

In light of this framework, we developed four genres of tasks for our concept-development lessons (Table 2.2). The first two rows refer to activities with mathematical objects (classifying and representing them), the third refers to making conjectures and statements about those objects, and the fourth refers to the identification of situations within which those objects may be found.

The problem-solving *Classroom Challenges* were designed to assess and improve the capability of students to solve multi-step, non-routine problems and to extend this to the formulation and tackling of problems from the real world. We define a problem as a task that the individual wants to tackle, but for which he or she does not have access to a straightforward means of solution (Schoenfeld 1985). One consequence of this definition is that it is pedagogically inconsistent to design problem-solving tasks for the purpose of practising a specified procedure or developing an understanding of a particular concept. In order to develop strategic competence—the "ability to formulate, represent, and solve mathematical problems" (Kilpatrick et al. 2001, p. 116)—students must be free to experiment with a range of approaches. They may or may not decide to use any particular procedure or concept; these cannot be pre-determined. Some task genres and sample classroom activities for problem solving are shown in Table 2.3.

We see problem solving as being contained within the broader processes of mathematical modelling. Modelling additionally requires the *formulation* of problems by, for example, restricting the number of variables and making simplifying assumptions. Later in the process, solutions must be interpreted and validated in terms of the original context.

Table 2.2 Task genres for concept development

Task genres	Sample classroom activities
Classify and define mathematical objects and structures	Identifying and describing attributes and sorting objects accordingly Creating and identifying examples and non-examples Creating and testing definitions
Represent and translate between mathematical concepts and their representations	Interpreting a range of representations including diagrams, graphs and formulae Translating between representations and studying the co-variation between representations
Justify and/or prove mathematical conjectures, procedures and connections	Making and testing mathematical conjectures and procedures Identifying examples that support or refute a conjecture Creating arguments that explain why conjectures and procedures may or may not be valid
Identify and analyse structure within situations	Studying and modifying mathematical situations Exploring relationships between variables Comparing and relating mathematical structures

Table 2.3 Task genres for problem-solving lessons

Task genres	Sample classroom activities
Solve a non-routine problem by creating an extended chain of reasoning	Selecting appropriate mathematical concepts and procedures Planning an approach Carrying out the plan, monitoring progress and changing direction, where necessary Reflecting on solutions; examining for reasonableness within the context Reflecting on strategy; where might it have been improved?
Formulate and interpret a mathematical model of a situation that may be adapted and used in a range of situations	Making suitable assumptions to simplify a situation Representing a situation mathematically Identifying significant variables in situations Generating relationships between variables Identifying accessible questions that may be tackled within a situation Interpreting and validating a model in terms of the context

2.3 Design-Based Methodology

Our method for lesson design was based on design-research principles, involving theory-driven iterative cycles of design, enactment, analysis and redesign (Barab and Squire 2004; Bereiter 2002; Cobb et al. 2003; DBRC 2003; Kelly 2003; van den Akker et al. 2006). In contrast to much design research, we worked to ensure that the products were robust in large-scale use by fairly typical end-users; thus, we engaged in what Burkhardt (2006) has termed "engineering research."

Each lesson was developed, through two iterative design cycles, and trialled in three or four US classrooms between each revision. This sample size enabled us to obtain rich, detailed feedback, while also allowing us to distinguish general implementation issues from more idiosyncratic variations by individual teachers. Revisions were based on structured, detailed feedback from experienced local observers in California, Rhode Island, and the Midwest. Overall, we obtained approximately 700 observer reports of lessons from over 100 teachers (in over 50 schools) using these materials. We also observed many of the lessons first-hand in UK schools. On this basis, the lessons were revised. These lessons have subsequently been researched by other independent organizations (see Sect. 2.6). It is worth noting that this engineering approach is more expensive than the 'authorship' model that is traditional in education. Nonetheless, even if widely adopted, the cost would be negligible within the overall running costs of an education system. We believe that it is the only approach—standard in other fields—that can reliably combine ambition of goals with robustness in practical use.

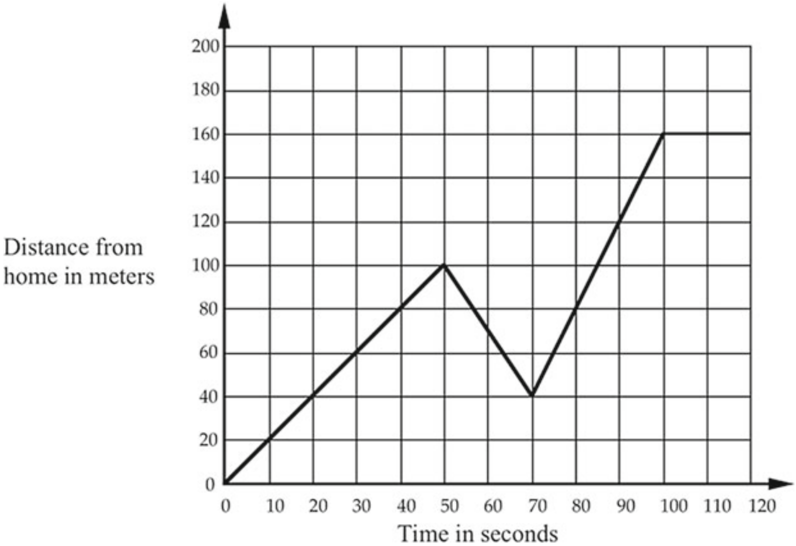
We now describe two examples of the formative assessment lessons that we developed, one concept-based and the other problem-solving (Swan and Burkhardt 2014). Each lesson was designed to occupy about 45 min of classroom time and to be worked on collaboratively. Complete lesson guides for these lessons may be downloaded for free from http://map.mathshell.org.

2.4 A Concept-Development Lesson

The objective of this lesson, *Distance-Time Graphs*, is to provide a means for a teacher to formatively assess students' capacity to interpret graphs. The lesson is preceded by a short diagnostic assessment, designed to expose students' prior understandings and interpretations (Fig. 2.1). We encourage teachers to prepare for the lesson by reading through students' responses and by preparing probing questions that will advance student thinking. They are advised not to score or grade the work. Through our trials of the task, we developed a 'common issues table' (Fig. 2.1) that forewarns teachers of some common interpretations that students may have, and suggests questions that the teacher might pose to advance thinking. This form of feedback has been shown to be more powerful than grades or scores, which detract from the mathematics and encourage competition rather than

Journey to the bus stop

Every morning, Tom walks along a straight road from his home to a bus stop, a distance of 160 meters. The graph shows his journey on one particular day.



1. Describe what may have happened.
   Include details like how fast he walked.
2. Are all sections of the graph realistic?
   Fully explain your answer.

Fig. 2.1 Initial assessment task: *Journey to the bus stop*, and an extract from the associated 'Common issues table'

collaboration (Black et al. 2003; Black and Wiliam 1998). Some teachers write their questions on the student work whereas others prepare short lists of questions for the whole class to consider.

The lesson itself is then structured in five parts:

1. **Make existing concepts and methods explicit**. The lesson begins with an initial task to clarify the learning intentions, create curiosity, help students become aware of their own intuitive interpretations and model the level of reasoning expected in the main activity (Strategy 1).[4] The teacher invites and probes explanations, but does not correct students or attempt to reach resolution at this point.

[4]The strategy numbers refer to the formative assessment strategies listed in Table 2.1.

Issue	*Suggested questions and prompts*
Student interprets the graph as a picture For example: The student assumes that as the graph goes up and down, Tom's path is going up and down *or* assumes that a straight line on a graph means that the motion is along a straight path.	• If a person walked in a circle around their home, what would the graph look like? • If a person walked at a steady speed up and down a hill, directly away from home, what would the graph look like? • In each section of his journey, is Tom's speed steady or is it changing? How do you know? • How can you figure out Tom's speed in each section of the journey?
Student interprets the graph as speed-time The student interprets a positive slope as 'speeding up' and a negative slope as 'slowing down'.	• If a person walked for a mile at a steady speed, away from home, then turned around and walked back home at the same steady speed, what would the graph look like? • How does the distance change during the second section of Tom's journey? What does this mean? • How can you tell if Tom is traveling away from or towards home?

Fig. 2.1 (continued)

2. **Collaborative activity:** ***Matching graphs, stories and tables***. This phase is designed to create student-student discussions in which they share and challenge each other's interpretations (Strategy 2). Ten distance-time graphs are to be matched with nine 'stories' (the 10th to be constructed by the student). When the cards have been discussed and matched, the teacher distributes a further set of cards that contain distance-time tables of numerical data. These provide feedback, by enabling students to check their own responses and reconsider the decisions that have been made. Students collaborate to construct posters displaying their reasoning. While students work, the teacher is encouraged to ask the prepared questions from the initial diagnostic assessment (Strategy 3).
3. **Inter-group discussion:** ***Comparing interpretations***. Students' posters are displayed, and students visit each other's posters and check them, demanding explanations (Strategy 4).
4. **Plenary discussion**. Students revisit the task that was introduced at the beginning of the lesson, and resolution is now sought. Drawing on examples of student work produced during the lesson, the teacher directs attention to the significant concepts that have arisen (Strategy 2).

5. **Individual work:** ***Improving solutions to the pre-assessment task***. Students now revisit the work they did on the pre-assessment task. They describe how they would now answer the task differently and write about what they have learned (Strategy 5).

2.5 A Problem-Solving Lesson

Our trials showed that teachers find it difficult to interpret and monitor students' extended reasoning during a problem-solving lesson, and very hard to select which students to invite to talk about it during whole-class discussion. We therefore decided to precede each lesson with a preliminary assessment in which students tackled a substantial problem individually. The teacher reviews a sample of the students' initial attempts and identifies the main issues that need addressing. The focus is on *approaches* to the problem. If time permits, teachers write feedback questions on each student's work, or alternatively prepare questions for the whole class to consider.

For example, one High School problem poses the following question:

A poster asserts that one female cat can have 2000 descendants in 18 months. Is this realistic?

This problem is accompanied by five pieces of information:

- The length of a cat's pregnancy is about 2 months;
- Cats can first get pregnant when they are 4 months old;
- Cats normally have 4 to 6 kittens in a litter;
- A female cat can have about 3 litters per year;
- Cats stop having kittens when they are 10 years old.[5]

The lesson is structured as follows:

1. **Introduction:** ***Responding to formative feedback***. The teacher re-introduces the main task for the lesson and returns students' initial attempts, along with some formative questions. Students have a few minutes to read these questions and respond to them individually (Strategy 3). 'Common issues' have been identified from trials and these are provided for teachers to use (Fig. 2.2).
2. **Group work:** ***Comparing strategic approaches***. In small groups, students are asked to discuss each person's work and then produce a poster showing a joint solution that is better than the individual attempts. Groups are organized so that

[5]This task was originally designed by Lesley Ravenscroft and appears courtesy of the Bowland Charitable Trust.

Issue	Suggested questions and prompts
Has difficulty starting	Can you describe what happens during the first five months?
Does not develop a suitable representation	Can you make a diagram or table to show what is happening?
Work is unsystematic	Could you start by just looking at the litters from the first cat? What would you do after that?
Develops a partial model	Do you think the first litter of kittens will have time to grow and have litters of their own? What about their kittens?
Does not make reasonable assumptions	What assumptions have you made? Are all your kittens born at the beginning of the year? Are all your kittens females?

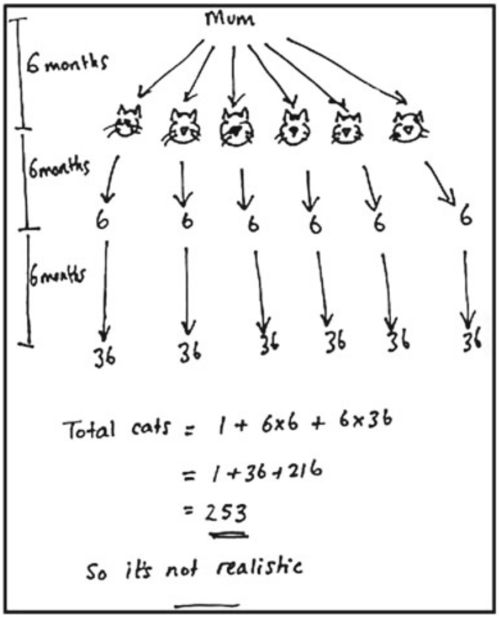
Fig. 2.2 An extract from the 'Common issues table' for *Having Kittens*

students with contrasting ideas are paired, thus promoting peer assessment (Strategy 4). The teacher's role is to observe groups and challenge students using the prepared questions to refine and improve their strategies (Strategy 2). The teacher may at this point ask students to review the strategic approaches produced by other groups in the class, and justify their own. Additionally, the teacher may introduce up to four pieces of "pre-designed sample student work" (Evans and Swan 2014), provided in the materials, which are chosen to highlight alternative approaches. Each piece of work is annotated with questions that focus students' attention (Fig. 2.3).

3. **Group work: *Refining solutions***. Students revisit the task and try to use insights to further refine their solution (Strategy 4).
4. **Whole-class discussion: *Reviewing learning***. The teacher holds a plenary discussion to focus on the processes involved in the problem, such as the implications of making different assumptions, the power of alternative representations and the general mathematical structure of the problem.

Questions for students:

Wayne's solution



- What has Wayne done correctly?
- What assumptions has he made?
- How can Wayne's work be improved?

Notes from the teacher guide:

Wayne has assumed that the mother has six kittens after 6 months, and has considered succeeding generations. He has, however, forgotten that each cat may have more than one litter. He has shown the timeline clearly. Wayne doesn't explain where the 6-month gaps have come from.

Fig. 2.3 Sample work for discussion, with commentary from the teacher guide

2.6 Conclusion

The two lessons we have described contain many features that are not common in mathematics teaching, at least in the US and UK. In both kinds of lessons, there is a strong emphasis on the use of preliminary formative assessment, which enables the teacher to prepare for and adapt interventions to the student reasoning that will be encountered. Students spend much of the lesson in dialogic talk, focused on comparing mathematical processes. The successive opportunities for refining the solution enable students to pursue multiple methods and to compare and evaluate them. Finally, pre-designed sample student work is used to foster the development of critical competence.

Early evidence of the impact of these lessons is encouraging. Drawing on a national survey of 1239 mathematics teachers from 21 US states, and interview data from four sites, Research for Action (RFA),[6] found that a large majority of teachers reported that the use of the *Classroom Challenges* had helped them to implement the Common Core State Standards, raise their expectations for students, learn new strategies for teaching subject matter, use formative assessment, and differentiate instruction.

The National Center for Research on Evaluation, Standards and Student Testing (CRESST) examined the implementation and impact of *Classroom Challenges* in 9th-Grade Algebra 1 classes (Herman et al. 2015). This study used a quasi-experimental design to compare student performance with *Classroom Challenges* to a matched sample of students from across Kentucky, comparable in prior achievement and demographic characteristics. On average, study teachers implemented only four to six *Classroom Challenges* during the study year (or 8–12 days); yet, relative to typical growth in mathematics from eighth to ninth grade, the effect size for the *Classroom Challenges* represented an additional 4.6 months of schooling. This remarkable gain cannot have come entirely from the particular topics focused on in those few formative assessment lessons, suggesting that there was significant 'seepage' of the pedagogy that these lessons exemplify into the teachers' other teaching—the goal that we set out at the beginning.

Although teachers felt that the *Challenges* benefited students' conceptual understanding and mathematical thinking, they reported that sizeable proportions of their students struggled, and it appeared that lower-achieving students benefited less than higher achievers. This, they suggested, may have been due to the great difference in challenge and learning style required by these lessons, compared with students' previous diet of largely procedural learning.

Finally, in 2014, Inverness Research (IR 2015) surveyed 636 students from 31 trial classes (6th Grade to High School) across five states in the US. They found that the majority of students enjoyed learning mathematics through these lessons and reported that they understood it better, had increased in their participation, and had improved in listening to others and in explaining their mathematical thinking. About 20%, however, remained unaffected by or disaffected with these lessons. This was because they did not enjoy working in groups, they objected to the investigative approach, and/or they felt that these lessons were too long or too difficult. It is our hope, with some evidence in support, that greater exposure to the

[6]RFA is a non-profit research organization; see http://www.researchforaction.org/rfa-study-of-tools-aligned-ccss/.

Classroom Challenges over a longer period will enable lower-attaining students to benefit more, as their teachers learn to broaden their adaptive expertise.

In conclusion, *Classroom Challenges* appear to provide a model for teachers as they attempt to introduce formative assessment into their everyday classroom practice, but they require a radical shift in the predominant classroom culture. How far teachers transfer this approach into the rest of their teaching is the focus of ongoing research. We are also currently looking at building on this work to design a suite of *Classroom Challenges* for elementary school ages.

Acknowledgements We would like to thank our Shell Centre designer-researcher colleagues at the University of Nottingham, Nichola Clarke, Rita Crust, Clare Dawson, Sheila Evans and Marie Joubert, along with David Martin and Daniel Pead; the observer-researchers around the US, led by David Foster, Mary Bouck and Diane Schaefer; Alan Schoenfeld, Hugh Burkhardt and Phil Daro, who led the project; and the Bill and Melinda Gates Foundation for the funding that made possible this core contribution to their 'College and Career Readiness' strategy for Mathematics.

The 'Having Kittens' task was originally designed by Acumina Ltd. (http://www.acumina.co.uk/) for Bowland Maths (http://www.bowlandmaths.org.uk), and appears courtesy of the Bowland Charitable Trust.

References

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences, 13*(1), 1–14.

Bereiter, C. (2002). Design research for sustained innovation. *Cognitive Studies, Bulletin of the Japanese Cognitive Science Society, 9*(3), 321–327.

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

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London, England: King's College London, School of Education.

Black, P., & Wiliam, D. (1999). *Assessment for learning: Beyond the black box*. Cambridge, England: University of Cambridge, Institute of Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Burkhardt, H. (2006). From design research to large-scale impact: Engineering research in education. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 121–150). London, England: Routledge.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9–13.

Cusi, A., Morselli, F., & Sabena, C. (this volume). The use of digital technologies to enhance formative assessment processes. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from across the globe* (pp. 77-92). Cham, Switzerland: Springer International Publishing AG.

DBRC [The Design-Based Research Collective]. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5–8.

Evans, S., & Swan, M. (2014). Developing students' strategies for problem solving in mathematics: The role of pre-designed "Sample Student Work". *Educational Designer, 2*(7). Retrieved from http://www.educationaldesigner.org/ed/volume2/issue7/article25/.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. W. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York, NY: W. H. Freeman/Times Books/Henry Holt & Co.

Herman, J., Epstein, S., Leon, S., La Torre Matrundola, D., Reber, S., & Choi, K. (2015). Implementation and effects of LDC and MDC in Kentucky districts. *CRESST Policy Brief No. 13*. Retrieved from http://www.cse.ucla.edu/products/policy/PB_13.pdf.

IR [Inverness Research]. (2015). *Mathematics assessment program (MAP): Project Portfolio*. Retrieved from http://inverness-research.org/mars_map/xindex.html.

Kelly, A. (2003). The role of design in educational research. *Educational Researcher, 32*(1), 3–4.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. London, England: Academic Press.

Sierpinska, A. (1994). *Understanding in mathematics*. London, England: Falmer.

Swan, M. (2015). Designing formative assessment lessons for concept development and problem solving. In T. G. Bartell, K. N. Bieda, R. T. Putnam, K. Bradfield, & H. Dominguez, (Eds.), *Proceedings of the 37th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 33–51), November 5–8, 2015. East Lansing, Michigan, USA.

Swan, M., & Burkhardt, H. (2014). Lesson design for formative assessment. *Educational Designer, 2*(7). Retrieved from http://www.educationaldesigner.org/ed/volume2/issue7/article24.

van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational design research*. London, England and New York, NY: Routledge.

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice, 11*(1), 49–65.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Wright, D., Clark, J., & Tiplady, L. (this volume). Designing for formative assessment: A toolkit for teachers. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from across the globe* (pp. 207-228). Cham, Switzerland: Springer International Publishing AG.

Author Biographies

Malcolm Swan has, since the 1980s, led the design team in many Shell Centre projects at the University of Nottingham, most recently the Mathematics Assessment Project (MAP). In 2008, he received the first 'Eddie' Prize for Educational Design from the International Society for Design and Development in Education (http://www.isdde.org/isdde/index.htm). In 2016, he became the first recipient, with Hugh Burkhardt, of the Emma Castelnuovo Medal for impact on practice from the International Commission for Mathematical Instruction. He passed away in Spring 2017 after a long illness.

Colin Foster is an Associate Professor in the School of Education at the University of Leicester, United Kingdom. He is interested in the design and use of rich mathematics tasks and lessons, and was a designer-researcher on the Mathematics Assessment Project.

Chapter 3
Observations and Conversations as Assessment in Secondary Mathematics

Jimmy Pai

Abstract In-the-moment decisions are important for teachers in a busy classroom. In this study, I explore secondary mathematics teachers' experiences in working with ephemeral assessment opportunities derived from observations and conversations within a classroom. The study reflects a phenomenological approach, and involves multiple interviews with three secondary mathematics teacher participants, as well as my own reflective journal as a secondary mathematics teacher. In this paper, I describe the complex nature of assessments that are embedded in classroom practice. Three phases of ephemeral assessment cycles (eliciting, interpreting, and acting) as well as factors that influence these phases of the cycle are discussed.

Keywords Assessment · Ephemeral · Observations · Conversations
In-the-moment decisions

3.1 Introduction

Assessment is a powerful process with which the teacher may facilitate learning. Wiggins (1993) noted that the word *assessment* derives from the Latin word *assidere*, meaning "to sit beside or with." Its origins (Klein 1966) suggest that assessment needs to be a process done *with* students, not simply *to* students. Baird and colleagues go so far as to state that "the intersection between assessment and learning is of utmost importance for the promotion or hindrance of quality in education" (2014, p. 3). Many have suggested that assessment strategies need to be

J. Pai (✉)
University of Ottawa, 75 Laurier Ave E, Ottawa, ON K1N 6N5, Canada
e-mail: Jimmy.Pai@ocdsb.ca

J. Pai
Ottawa-Carleton District School Board, 133 Greenbank Rd, Ottawa, ON K2H 6L3, Canada

J. Pai
Ottawa, ON K2P 0J4, Canada

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*, ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_3

varied—to include, for example, "observations, clinical interviews, reflective journals, projects, demonstrations, collections of student work, and students' self-evaluations" (Shepard 2000, p. 8). Research also indicates that many inservice mathematics teachers are using a variety of assessment practices (e.g., Suurtamm et al. 2010), including observations and conversations. These are, by their nature, in-the-moment and fleeting. They are *ephemeral*. These ephemeral forms of information are important, as these quick exchanges are "consequential—what [teachers] see and don't see shapes what [teachers] do and don't do" (Schoenfeld 2011, p. 228) in day-to-day activities.

In this paper, I share explorations of observations and conversations as assessment based on a study with three secondary mathematics teachers, as well as my own reflective journal as a secondary mathematics teacher. The central research question is: what is it like for a secondary mathematics teacher to consider and use observations and conversations as assessment opportunities within the classroom? The guiding questions include: (a) how do teachers engineer situations that are suitable for observations and conversations? (b) how do teachers interpret information gathered from observations and conversations? (c) what do teachers do with what they have interpreted from the observations and conversations?

3.2 Theoretical Framework

To better understand ephemeral forms of assessment, I draw from thinking about summative and formative functions of assessment (e.g., Harlen 2005, 2012; Wiliam 2010) as well as mathematics teacher noticing (e.g., Mason 2002; Sherin et al. 2011) in my theoretical framework.

An assessment serves summative functions when the assessor attempts to *sum* up the evidence of student understandings "by drawing inferences about [his or her] students' learning" (Harlen 2005, p. 213). Assessment could serve formative functions when it helps students *form* understanding. Leaning on Wiliam (2010):

> An assessment functions formatively to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of that evidence. (p. 148)

The descriptors *summative* and *formative* are not defining characteristics of any particular assessment process (e.g., Wiliam and Leahy 2007). Instead, Harlen suggests that it is possible for assessment information to be used "for both summative and formative purposes, without the use for one purpose endangering the effectiveness of use for the other" (2005, p. 215). These descriptors, then, only apply retroactively to an assessment process after it has taken place. In other words, whether or not an assessment process has functioned formatively depends on its interactions with specific students and how one utilizes the related information. This consideration of the functions of assessment relates to my third research question concerning what teachers do with interpretations of observations and conversations.

The three component skills of professional noticing of children's mathematical thinking summarized by Jacobs et al. (2010) also contributed to my exploration of the moments in the classroom. These three component skills included attending to children's strategies, interpreting children's understandings, and deciding how to respond on the basis of children's understandings. Of these three skills, Jacobs et al. (2010) believe that attending and interpreting are purposefully developed expertise, and argue that teachers must "execute these three skills in an integrated way, almost simultaneously" (p. 192) in order to build on student thinking. These perspectives on noticing inform my study with respect to my first and second research question regarding how teachers engineer situations for observations and conversations, and how they subsequently interpret the information.

Mason's (2002, 2011) conceptions of noticing have been influential in how I think about assessment as well as how I approached my study. He describes noticing as "an act of attention, and as such is not something you can decide to do all of a sudden" (Mason 2002, p. 61), and that it involves levels of awareness. This view of noticing both encompasses the entire assessment process and, at the same time, is subsumed in every element in the assessment process. It may encompass the entire assessment process if one considers questions such as why do we assess, for whom do we assess, and how might we assess. In this perspective, noticing is also specific to the different phases of assessment, since noticing is about "[working] on becoming more sensitive to notice opportunities in the moment; to be methodical without being mechanical" (p. 61). This involves being prepared to respond in the moment as opposed to reacting habitually.

3.3 Visual Representation of an Assessment Cycle

I have developed a visual representation of an assessment cycle in Fig. 3.1 to help me better understand the process of assessment, as well as to situate my study concerning ephemeral moments in the mathematics classroom. The phases have been influenced by Wiliam and Black's (1996) general descriptions of elicitation, interpretation, and action, as well as conceptions of noticing (e.g., Mason 2002; Sherin et al. 2011).

Eliciting is when the teacher gathers information about student learning, whether *purposive* (planned), *incidental* (unplanned), or a combination of the two. Information from observations and conversations can be both *purposively* or *incidentally* elicited. The resulting information may be *permanent* (i.e., long-lasting and can be revisited) or *ephemeral* (i.e., fleeting and need to be seized). This study focuses on assessment opportunities that are *ephemeral* in nature, and may be *purposive* or *incidental*. The eliciting process includes the idea of attending to children's strategies described by Jacobs et al. (2010). Attending is an important aspect of eliciting, since, for the teacher, elicited information fades away if

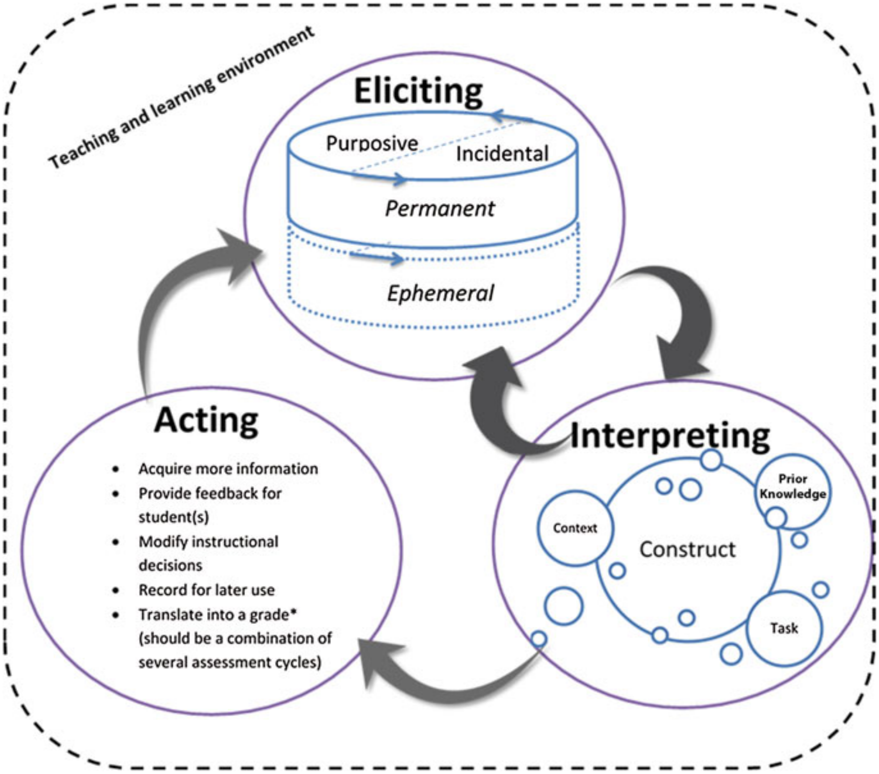


Fig. 3.1 Visual representation of an assessment cycle

unattended. Eliciting, therefore, involves how teachers design tasks and activities, facilitate student-peer conversations, and engage students in elaborating their thinking.

Interpreting is when the observer attempts to make sense of what the students mean, and considers what the students understand. The extent to which a teacher toils over interpreting is related to how and what was elicited, as well as the kinds of action she intends to take. If the teacher purposively elicits information about student understanding of integers, and she plans to use this opportunity to provide feedback on students' thinking, then this frame of mind would impact how she interprets what the students say and do. Jacobs et al. (2010) believe that interpretation is closely tied to teacher's "understanding of the mathematical landscape to connect how [children's] strategies reflect understanding of mathematical concepts" (2010, p. 195). Making sense of student understanding also includes deciding how to respond, which is "intended responding" (Jacobs et al. 2010, p. 173) and not the actual execution of the response. With ephemeral assessment opportunities, there is not a lot of time available for interpreting—all events that follow the elicitation are seen as actions, even if it is silence or to walk away from the students.

Acting is what the teacher does after eliciting and interpreting information from students. A teacher's actions may function formatively (e.g., provide feedback) or summatively (e.g., translate into a grade), or both. Some examples of actions have been listed in Fig. 3.1, but this is not intended to be an exhaustive list. Actions deriving from ephemeral assessment opportunities often serve formative functions, while tests remain the primary form of assessment that serves summative functions. This may be due, in part, to teacher perceptions of pressures from stakeholders who discount reports of student achievement that are not based on tests (Black et al. 2010).

Ephemeral assessment cycles may inform the teacher on next steps to take in class. In other words, information and interpretations may feed back into a new or existing assessment cycle. They may also inform more effective teacher actions in the future, as these cycles help develop "a collection of alternative actions [to habitual reactions] and an awareness of situations in which these actions would be preferable" (Mason 2011, p. 38). This means that interpretations and actions from one assessment cycle have the potential of feeding back into other assessment cycles. These compounding of interpretations and actions contribute to the assessment processes which may serve formative, summative, or both functions.

3.4 Methodology

This paper reports on aspects of my phenomenological study that aimed to better understand how teachers use observations and conversations within the classroom. It is phenomenological because of its focus on weaving together participants' lived experiences and interpretations. The study includes two rounds of individual teacher interviews with three secondary mathematics teachers, a focus group interview with the three teachers, as well as my reflective journal as a secondary mathematics teacher.

3.4.1 Teacher Interviews

The participants are secondary mathematics teachers, each with between 17 and 24 years of teaching experience: Cadence, Casey, and Fernanda. (Pseudonyms are used for teacher participants and any students they mentioned.) The purpose of the first round of semi-structured individual interviews was to re-construct participants' experiences of ephemeral assessment moments in their class. This first interview began with questions regarding background information to understand potential influencing factors in the teachers' perspectives. I then asked about how the participants create situations that allow them to observe, ask questions about, and listen to students' mathematical thinking. I continued by asking the participants to describe a specific moment in recent memory, and sought details about the set up of the classroom, details of the tasks, the structure and context of the lesson, student reactions and responses during the conversations and more. I also asked participants to recall what they considered during the moment and about short and long-term decisions that were impacted by the moment they described. I concluded the interview by exploring supports and challenges in observations and conversations.

During the focus group interview with all three participants, we explored and questioned themes that had emerged during the first round of interviews. The second round of individual interviews allowed for opportunities to reflect on the ideas that had been discussed, as well as to clarify participant statements from previous meetings. The process of interpreting emerged as difficult to recall and capture. As a result, at this final interview there was an emphasis placed on further exploring how we, as teachers, interpret events in the moment.

3.4.2 Reflective Journal

Reflective journals are useful for "[digging] deeper… into the heart of words, beliefs, and behaviours" (Janesick 1999, p. 513), and provide a good position from which to examine my own presuppositions, choices, experiences, and actions. In this way, it is helpful for developing bracketing skills and improving my awareness as a researcher. Schön (1983) coined the terms *reflection-on-action* (thinking back over what happened) and *reflection-in-action* (awareness of thinking as one acts in the moment). The first 19 days of the journal were completed before beginning the study. These initial journal entries were *reflections-on-action*, involving portfolio and observation assessments I used during my classes. As my journey progressed, my inquiry focused on ephemeral assessment opportunities, and I identified a need to explore my thoughts during these moments. Therefore, the entries for days 20–23, again before the study, also included *reflections-in-action* as well as dialogues of classroom conversations with students as context. I continued my reflective journal throughout the study, including interview, transcription, and analysis processes. During the interviews, I *reflected-in-action* and noted down thoughts I had during the conversations. While transcribing the interview, I fleshed out these descriptions of what I thought about in-the-moment. At the same time, I reflected on the actions that I have taken and related these to my experiences as a secondary teacher.

3.5 Analysis

Analysis was influenced by the work of Van Manen (1990, 2014). I began the analysis of the teacher interviews by reading through the transcriptions in their entirety to get a general sense of what the participants were talking about. For each participant, I then went through the transcripts paragraph by paragraph and identified key ideas related to how teachers *elicit* ephemeral information from students. I repeated the process for ideas related to *interpreting*, and again for *acting* on ephemeral information. Emergent themes were then compiled from the three participants under *eliciting*, *interpreting*, and *acting*. As I worked with these themes, factors that influence eliciting, interpreting, and acting also became apparent. As a result, I found it useful to explore a second layer of analysis that examined *factors* affecting the three phases of eliciting, interpreting, and acting.

The reflective journal was helpful in two ways. First, as I reflected on my own lived-experience, I became more aware of possible experiential meanings. This helped me reflect on conversations and observations, and identify key ideas. Second, this heightened sensitivity also helped me in my attempt to reach epoché (Van Manen 2014), and enabled me to understand each participant's view on assessment. Experiences are personal, and thus it is important to be able to understand the participants' experiences.

3.6 Discussion of Results

I have found it useful to consider two layers in discussing the data. For the purposes of this paper, I focus on discussing the first layer of my analysis, which involves what teachers do and think about during the ephemeral assessment process; the second layer involves the factors that affect what teachers do and think. A summary of the first layer of analysis (Fig. 3.2) categorizes what the teacher does and thinks about during the ephemeral assessment process. This was organized under eliciting, interpreting, and acting. During the eliciting phase, "how we elicit" describes what teachers do, and "what we elicit" describes how students respond. The interpreting phase involves what teachers think about. It begins with "making sense" of information obtained in the previous phase, and continues onto "building impressions" and "deciding on responses." Lastly in the acting phase, teacher actions serve three functions of formative, summative, and interpersonal.

3.6.1 Eliciting

We can consider two stages in eliciting: *how* we elicit and *what* we elicit. As mentioned earlier, ephemeral sources of information may be elicited purposively or incidentally. A teacher may plan ahead of time to elicit specific information about a student's understanding, or happen upon it incidentally.

***How* we elicit during the moment in the classroom**. The first stage involves how the teacher goes about eliciting information through conversations and observations. The teacher may do this through activities or through questions and prompts. These ephemeral sources of information can be elicited both purposively and incidentally.

<table>
<tr><td colspan="6">Eliciting</td></tr>
<tr><td colspan="6">How we elicit</td></tr>
<tr><td colspan="3" rowspan="2">Questions and prompts
(Boaler and Brodie 2004)</td><td colspan="3">Activities involving student(s), teacher, or both</td></tr>
<tr><td colspan="2">Ongoing activities</td><td>New activities</td></tr>
<tr><td colspan="6">What we elicit</td></tr>
<tr><td colspan="3">What the student says, writes, or does</td><td colspan="3">How the student says, writes, or does</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6">Interpreting</td></tr>
<tr><td colspan="6">Making sense of what and how the student says, writes, or does</td></tr>
<tr><td colspan="3">Building impression of the student(s)</td><td colspan="3">Deciding how to respond</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td colspan="6">Acting</td></tr>
<tr><td colspan="2">Formative function</td><td colspan="2">Summative function</td><td colspan="2">Interpersonal function</td></tr>
</table>

Fig. 3.2 What the teacher does and thinks about during the ephemeral assessment process

First, the teacher may elicit information through activities that involve individual students, groups of students, teacher, or any combination thereof. These may be ongoing or new activities. For example, Cadence had a cross-curricular (English and Math) banking project where students worked on understanding different concepts involved in banking, such as down payments and financing. Through this project, students had conversations with each other and the teacher. Cadence established opportunities for natural and ongoing elicitation of student learning through this project.

Second, the teacher may also say something to the students through the form of questions and prompts that elicit a response. Boaler and Brodie proposed categories for questions and "utterances that had both the form and function of questions" (2004, p. 777). One category of questions involved teachers gathering information or leading students through a method.

In the following sections, I point out two implications of how we elicit during moments in the classroom. The first is that incidental eliciting requires a level of awareness on the part of the teacher. The second is that there exists complexity in how we question and prompt.

3.6.1.1 Incidental Eliciting Requires Awareness

In an earlier section, I differentiated between purposive and incidental eliciting. My conversations with the participants revealed that incidental eliciting requires awareness in the moment. A teacher may have purposively devised many opportunities to elicit information about students' learning. However, having rigid 'look-fors' prevents the teacher from accessing the myriad of information that student words and actions may provide. For example, during the banking project, Cadence engaged with a student who had trouble understanding "16 out of 16" as a percentage. The conversation helped move the student's understanding forward, as well as provide information on what the teacher might be able to do next with the rest of the class. This would not be possible if Cadence was only listening for[1] conversations about the mechanics of compound interest.

In another example, Fernanda engaged the class in reviewing linear relations by providing students with five lines on a grid. Students were instructed to ask each other questions and to explore subsequent elaborations. As students worked, she came across a student who asked another student to identify parallel lines. Fernanda then took this idea, and asked the rest of the class to also engage in developing equations for parallel lines. This activity was created on the spot, specifically to elicit information about students' understandings of rates of change.

The difference between incidental and purposive eliciting is that one of them is necessarily born out of the moment. A teacher is unable to access incidental opportunities of noticing if she is not in a frame of mind that welcomes them.

3.6.1.2 There Is Complexity in How We Question and Prompt

I note three interesting findings in using the previously mentioned categories of questions proposed by Boaler and Brodie (2004) in my analysis. First, not many gathering information types of questions were described by the teacher participants when asked to consider the ephemeral assessment opportunities. This contrasts with findings from Boaler and Brodie (2004) where they found a high percentage of teachers used questions that involved factual responses. This may be due, at least in part, to the fact that participants were asked to consider moves that create opportunities for student conversations. Questions that involve factual responses, by their nature, are closed questions that are not conducive to continuing student conversations.

Second, the categories seem to be closely interrelated, and multiple types can be attributed to the participants' questions or prompts. In the following example, Casey described a conversation between her and two students when the students had just finished creating a sine graph with spaghetti:

> I said "okay ladies, predict for me, cosine, before you actually start [building the graph with spaghetti]" and Callie said "[pause] well […] I think it's going to look like this" but Callie, not confident, going "but I'm not really sure. I'm not really sure what it looks like but I think it kind of looked sort of similar." And Elizabeth, who's more confident, [said] "well no… cosine is supposed to start up here… and then it needs to go down" […] Elizabeth was starting to measure off and place the points for the cosine curve and then I said "okay, so what do you predict after that." And Elizabeth said… "okay, well, it's going to keep going down and it's going to keep going on, but it's moved over, it's like this curve moved over."

[1]The work on listening from several authors (e.g., Crespo 2000; Davis 1997; Suurtamm and Vezina 2010) has been influential in my thinking about how we elicit.

> So, she also got this connection. Callie wasn't so clear. (Casey, Initial interview, December 1, 2015)

As Casey and I explored this conversation further, we found that her request for a prediction served multiple purposes:

- Linked back to the sine graph that the students were working on
- Extended the students' thinking to a different situation involving cosine
- Probed and got the students to explain their thinking with respect to the cosine and sine function.

This differed from how Boaler and Brodie (2004) categorized questions and prompts, where questions and prompts served only one purpose each. I believe that the idea that teacher questions and prompts may serve multiple purposes creates a more complex picture of how we elicit information in the classroom.

Finally, the teacher may not even fully understand or appreciate the purposes that her questions and prompts might serve, until after the moment has passed. In the previous excerpt, as Casey and I discussed the multiple purposes in her questions and prompts, we discovered that she was unsure of the primary purpose. With the three purposes listed above, she explained that she did not know whether:

(a) She began with wanting to probe student thinking, by extending their thinking to a different situation, and subsequently to prompt the students to link back to the sine graph.

or perhaps

(b) She began by extending their thinking to a different situation, in the process got the students to explain their thinking, and subsequently involved the students in linking back to the sine graph.

As Casey and I reconstructed the moments in her classroom and explored her in-the-moment decisions on how she elicited information, we realized that often we are not completely clear with why we say what we say. The purposes may only emerge as a product of the interaction involving both the question/prompt and the response.

3.6.1.3 *What* We Elicit from the Moment in the Classroom

The second stage in eliciting involves what is generated by the student during the moment. During ephemeral assessment opportunities, the teacher is listening to what the student is saying and watching what the student is doing. At the same time, the teacher is taking in other cues, such as body language (e.g., eye contact, head scratches) and speech disfluencies (e.g., pauses, repetitions).

As Casey pointed out, it is important during conversations to "get at [students'] thought processes to actually understand what they're thinking" (Casey, First Interview, December 1, 2015). She explained that this may emerge naturally from student conversations with each other while engaged in a task. Other times, the teacher may elicit more information by prompting students to compare representations or asking students to elaborate on their mathematical thinking behind their actions. For example, during the same task related to trigonometric functions, Casey asked students to make connections between a sine graph on the *x*-*y* plane to the representation of a unit circle.

Teachers often do not recognize that they have already incorporated a consideration of body language and speech disfluencies, or even that they have been made available. For Casey, she remarked that our explorations of her conversations with students were eye-opening because we unpacked many subtleties in the interactions. For example, when Casey interacted with students about the trigonometric functions, she also paid attention to their body language. This included their gestures as they attempted to trace the graph with their fingers. This also included where they were looking as they explained their reasoning. Paying attention to speech disfluencies (such as pauses, repetitions) can also be important (e.g., Corley et al. 2007; Gordon and Luper 1989) in paying attention to student learning. For example, as Casey's student, Callie, engages with a problem that she is uncertain of, she produces speech disfluencies as she explains her thinking. The hesitations therefore heightened Casey's attention to what Callie said immediately following the hesitations. This then helped Casey identify the task as being more complex for Callie.

3.6.2 Interpreting

Not surprisingly, the ephemeral assessment process is complex and difficult to capture. Notably the most difficult aspect of reconstructing the ephemeral moment for the participants is not how they created the environments (eliciting), or what they did after the moment (acting). It is what they were thinking during the moment itself (interpreting). The teachers I interviewed often had difficulty reconstructing those interpretations, and in my analysis, I found that turning to my reflective journal helped to describe those moments. I have sorted interpreting into two stages: *making sense* of information and *thinking about* the information, which may be simultaneous and done subconsciously.

3.6.2.1 Making Sense of *What* and *How* the Student Says, Writes, or Does

The first stage is when the teacher integrates as many pieces of information as possible involving what and how the student says, writes, and does. There were four aspects that I saw through data analysis. I use Casey's example to elaborate on how these aspects contribute to her making sense of students' mathematical thinking about trigonometric functions. All of these aspects contribute to Casey's in-the-moment interpretation of how her students understood trigonometric functions.

First, teachers can only make sense of what they have taken in as information. Casey, for example, indicated that she scans the class all the time while teaching, and subsequently decides on areas of the room where she may be the most helpful. This means her attention to elicited information is opportunistic; with each decision to attend to a specific area, other possibilities vanish as it is impossible for the teacher to be everywhere at the same time.

Second, as I pointed out earlier, beyond simply what a student said, wrote, or did, body language and speech disfluencies can also be interpreted. This is also opportunistic; once again, a teacher cannot interpret what they did not sense. For example, Casey lamented that when she used an iPad to capture conversations in the classroom, she "focused on the iPad and…so [she] couldn't see [the expression of the student]" (Casey, First Interview, December 1, 2015).

Third, interpretation often requires information beyond simply what is available in the moment. In order to interpret student thinking in the moment, it is also helpful to relate to prior examples of what the teacher had seen or heard before such a moment. Casey, for example, elaborated on other interactions in the past that contributed to how she interpreted Callie's hesitations. She explained that she believed these may have led her to believe that Callie didn't quite get the concepts yet.

Finally, the process of making sense is neither algorithmic nor transparent. We are constantly interpreting happenings around us. Any attempted analysis of how we interpret cannot be during the moment. Analysis would always be retrospective because we cannot return to the same moment. Even if we begin to analyze how we interpret during interpretation, we would no longer be interpreting the happenings in the moment. This was the case for all three participants as, during the second interview, we began to investigate how we interpret in the moment decisions. However, even if the process of making sense cannot be algorithmic or transparent, there is still great value in examining how we interpret. Mason (2002) believes that it is important to reflect and build capacity for becoming aware of possibilities during the moment. Even though my interviews with the three participants failed to identify how they interpreted their moments without a shadow of a doubt, they believed that the attempt to understand how we interpret was a powerful exercise for thinking about our practice. Casey, for example, elaborated on how the interviews have helped her think about thinking:

> You've made me think a lot more about… this whole thing about how do you think in the classroom… good god, right!? When you've been teaching for almost 20 years, it's like "wow I don't think about it, you just do it"… so, you've made me think a lot about my teaching, which is actually very useful, because we have a tendency to get in our classrooms, close the door, and not think about our thinking [laughs] which is a bad plan [laughs] (Casey, Second Interview, April 22, 2016).

3.6.2.2 Building Impressions and Deciding How to Respond

The second stage can involve two elements: *building impression* of the student(s) and *deciding* how to respond. First, the teacher may be building an ongoing or new impression of how a student understands a particular topic. The teacher is also building an ongoing or new impression of the circumstances that are most appropriate for a specific student to learn. Across many interactions with the student, a teacher builds a better understanding of how a particular student interacts with her learning environment. This process of building an interpretation applies to individual students, but it also applies to the class as a whole. The teacher is also constantly developing a better idea as to what the class understands, and what strategies are most effective. The second element is deciding how to respond. This is the same phrase used by Jacobs et al. (2010) and reflects intended responding, and not the actual execution of the response. This is where teachers may think about the information that they have processed and decide on what to do next.

3.6.3 Acting

Teacher actions help facilitate mathematical learning in the classroom. Fernanda, for example, helped the class focus on parallel lines by redirecting a question from students, and Casey drew attention to different representations of trigonometric functions to further student understanding. In this section, I focus on several suggested additions to existing conversations regarding teacher actions during an ephemeral assessment process.

3.6.3.1 Teacher Actions also Serve Interpersonal Functions

Conversations with my participants revealed the importance of considering interpersonal functions of teacher actions, as these can improve, or make more difficult, the possibilities for the assessment process to serve formative or summative functions. For example, Cadence described a student who became very defiant about doing a written assessment related to the Pythagorean theorem. Cadence believed that the fact that she had built connections with this student before that incident allowed her to send the student to a student support worker. She commented on how "if [she] did that to [other students], and if [she] didn't have connection, they're done… and it would take months [to regain their trust] and maybe [even then] you won't get that connection back" (Cadence, Initial interview, November 23, 2015).

As mentioned in the theoretical framework, interpretations and actions feed back into the assessment process as teachers engage in more assessment cycles. As Cadence pointed out, interpersonal functions affect all ongoing and future assessment processes with a student, and therefore impact teachers' ability to support students' learning of mathematics. Although there has been work on the types and frequencies of feedback (e.g., Voerman et al. 2012), or effects of feedback on performance (e.g., Kluger and DeNisi 1996; Hattie and Timperley 2007; Shute 2008), I believe the interpersonal function helps frame another layer of complexity.

3.6.3.2 Assessment Functions Depend on How It Was Done and Who the Assessment Process Involved

In the example where Cadence sent a student to a student support worker, she commented on how this may not have gone well if it were a different student. This means it is important to consider not only what the teacher does, but also *how* s/he does it and *who* the student is, in order for it to serve possible functions. This subsequently indirectly affects the teacher's ability to support a student's learning of mathematics.

There were many different ways that Cadence could have had the student go to the student support worker. She could have called the student support worker to come retrieve the student, she could have asked the student to go to the student support worker, or she could have walked the student down herself. *How* she sent the student to the student support worker may impact how it was perceived by the student as well as how the rest of the situation played out. This could then have an impact on current or future assessment processes, and, for example, made it difficult to elicit, interpret, or act on more information.

It is also important to consider *who* the student is when acting during assessment processes. As Cadence pointed out, she had a good understanding of what worked with the defiant student. Her connection that she built in the previous year with a non-mathematics course also helped make the necessary decisions to send the student away.

3.6.3.3 Unintended Actions also Have an Impact

Even actions that are not easily captured may have an impact on the assessment process. This includes short acknowledgements such as 'mm-hmms' during a student's explanations of her/his thinking. For example, Casey was responding to student explanations about the trigonometric functions made out of spaghetti. Throughout the students' explanations, Casey responded with vocalisations such as 'yeah', 'mm-hmm', and 'hmm's.' In reflection, Casey believed that these unintended brief responses helped make students more comfortable in continuing their elaborations. This means that these vocalisations indirectly supported possible teacher actions that facilitate the learning of mathematics.

3.6.3.4 Moments May Serve Summative Functions Directly or Indirectly

Acts from ephemeral assessment cycles may also serve summative functions. These summative decisions may be directly or indirectly impacted by ephemeral sources of information. At Cadence's school, students who attend are those who have been unsuccessful in a regular school program. For a variety of reasons, including test anxiety, "the majority of [Cadence's] marks [come from] observations and conversations" (Cadence, Initial interview, November 23, 2015). When she makes summative decisions such as grading, this is compiled from the results of multiple interactions. In this way, no particular moment would constitute the entirety of a student's mark. Rather, she would make records about several instances of classroom interactions, and make summative decisions based on those. During the interviews, she also explained that she had been exploring online portfolios in order to better manage recordings of learning.

Actions that serve summative purposes can also be indirectly impacted by observations and conversations. For example, Casey describes instances where marks from written tests did not correspond with her observations about a particular student's understanding. As a result, she sought out other opportunities to explore this discrepancy. These examples demonstrated that not only is it possible for ephemeral sources of information to contribute to summative decisions, they also serve the function directly and indirectly.

3.6.4 Factors Affecting the Three Phases

As I explored the ephemeral assessment cycle, I began to notice many factors that appear to influence and are influenced by how and what teachers think and do. I conceptualize these influencing factors as four interrelated domains: teacher (e.g., identities, experiences, beliefs, frame of mind), students (e.g., identities, experiences, beliefs), relationships (e.g., teacher-student, student-student), and contexts (e.g., goal, time, accountability). A teacher may integrate these factors of influence consciously or subconsciously. In this paper, I briefly describe one aspect of the teacher domain that affects, and is affected by, the different phases in an assessment cycle: consideration of the teacher's frame of mind.

In the following exchange, Casey noted that her interactions with students changed when she was recording the conversation:

> Casey: I realized, this is true… when you are… videotaping… you then become an observer. And… you don't interact in… because the teacher as you know, goes from spot to spot to spot to spot to spot to spot
>
> Researcher: right of course
>
> Casey: as the observer, I had to stay here… for an extended period of time to actually observe this particular group (Casey, First Interview, December 1, 2015).

Besides having to stay with the group longer, she continued to explain that she found that she was asking the students to elaborate for the purposes of the recording, and not for the sake of better understanding student learning. She also noted that she was often distracted by the management of the iPad and did not pursue conversations as she normally would.

Fernanda believed that her mood may also impact how a lesson unfolds, or how a conversation with a student progresses. For example, if she had just disciplined a student, she explained, that may affect how she paid attention and how she responded during a conversation with another student. In a similar way, Cadence also suggested that there are rhythms that teachers and students get into. Sometimes the rhythm of the classroom is positive and productive, and other times the rhythm can be negative and destructive. Part of this rhythm can be attributed to the mental state of the teacher during the interactions of the classroom, since the rhythm is perceived by the teacher, as well as influenced by the teacher.

These frames of minds, moods, or rhythms indirectly affect the teacher in how she might introduce a task, how she might facilitate conversations with and between students, or how she might interpret elicited information. These subsequently impact how students engage in mathematical thinking.

3.7 Implications and Further Research

Teachers are constantly engaged in multiple assessment cycles. Processes involving in-the-moment assessment are difficult to capture, even when recognized as important. This paper sought to examine the intersection of formative assessment and noticing through teachers' lived experiences. Further research might also find it useful to explore both fields as we continue our conversations about assessment.

The visualization of the assessment process that I have developed is helpful in at least two ways. First, its development and continued modifications help me better understand the phenomenon of assessment strategies based on teacher observations of student actions and conversations in an environment that encourages group work and student discourse. Second, the framework itself, along with the various elaborations and implications from this paper, may be helpful for others in thinking about the complexities of in-the-moment decisions within the classroom. By sharing parts of my journey, I hope to participate in, and contribute to, the ongoing conversations on classroom assessment in secondary mathematics education.

3.7.1 So What? How Might This Be Helpful for Teachers?

As a secondary mathematics teacher, I have continued to ask myself these two practical questions throughout my research journey. Ephemeral assessment is a complex process. At the same time, it is also an important aspect of how mathematics teachers facilitate learning. My work has elaborated on some of the complexities involved in ephemeral assessment opportunities. This may be helpful for practitioners in two ways.

First, it may help teachers expand their awareness of possibilities in the moment. Becoming aware of the complexities involved in the moments of the classroom may be intimidating. However, this awareness can also be empowering. It provides teachers with different ways of reflecting *in* the moment as well as reflecting *on* the moment. The complexities illustrated are not to say that all teachers must consider every possible aspect of the situation, and act on them perfectly. Instead, it is to provide teachers with a sense of freedom and wealth because there are many possibilities that they could follow up on, and many opportunities after the moment to reflect and learn from the possibilities.

Second, acknowledging the complexities in the moment also may encourage teachers to seek out different ways of reflecting on their practice. During the interviews, all three participants pointed out that they thought it was helpful to reconstruct the moment with another teacher (who happened to also be a researcher). Casey, in particular, appreciated working together to unpack the complexities in the moments she described, and felt that it strengthened her decision making for the future. While teachers may not always have access to a researcher, they often have access to their colleagues. This study suggested not only that it is important to reflect, but also offers an example as to how one might reflect. It may be helpful to honor the moment that occurred by reconstructing it with another teacher—to describe not only what happened, but also how they made sense of what happened, why they interpret the situation the way they did, and more.

As a secondary teacher, I have enjoyed my research journey and plan on continuing in some capacity. I believe it may be helpful for professional development to provide opportunities for teachers to clarify and expand their definitions of assessment. Even a recognition that we, as individuals, have different working definitions for the concept of assessment may be eye-opening. The task of such a professional development session, then, would be to work with teachers (and to allow them to work with themselves) toward a growing definition of assessment and thinking about how this impacts their practice in meaningful ways.

3.7.2 Future Directions

At this point of my research journey, I can conceive of four possibilities for future research.

First, there may be modifications of, or elaborations on, various aspects of my visualization of the ephemeral assessment process, and influencing factors. Although influential factors may not impact teachers in the same way, it may be helpful to examine examples of how that may occur. During the interview process, the participants and I had difficulties identifying how they interpreted the situation and how they arrived at a decision to act. An exploration targeted at the interactions among influential factors and in-the-moment decisions may be illuminating.

Second, it may be helpful to simultaneously explore how a teacher and her students might experience the same moment in the classroom. These possible future explorations of student perspectives may provide more insight into the details of the ephemeral assessment process.

Third, it may be illuminating to better understand how awareness impacts the ways that teachers act and reflect. Perhaps a study where participants actively practice ways of improving awareness may be worth exploring. During the focus group, participants commented on how this study and its focus on in-the-moment thinking had really helped them think about how they thought in the classroom. It may be interesting for future studies to explicitly work with teachers on improving awareness.

Lastly, it may also be interesting to further explore the role of technology. The use of a recording instrument distracted Casey and altered her role in the moment. Because she was not used to recording students during the class, it may be interesting for future studies to explore the impact of technology on ephemeral assessment processes for teachers who have become accustomed to recording their students in some way.

A focus on ephemeral assessment helps identify the interactions in the classroom at the center of mathematics education research. I believe it is important to continue to emphasize these interactions. Not only is assessment a bridge between teaching and learning, I believe that better understandings of assessment processes (and their influencing factors) also provide roads to continuing improvements to teachers' professional practice.

References

Baird, J. A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. S. (2014). *State of the field review: Assessment and learning* (OUCEA Report 13/4697). Retrieved from University of Oxford, Oxford University Centre for Educational Assessment website http://bit.ly/1sGrSpT.

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice, 17*(2), 215–232.

Boaler, J., & Brodie, K. (2004). The importance, nature and impact of teacher questions. In D. E. McDougall & J. A. Ross (Eds.), *Proceedings of the 26th Annual Meeting of the North American Chapter of the International Group for Psychology of Mathematics Education* (pp. 773–782). Toronto, Canada: PME-NA.

Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105,* 658–668.

Crespo, S. (2000). Seeing more than right and wrong answers: Prospective teachers' interpretations of students' mathematical work. *Journal of Mathematics Teacher Education, 3,* 155–181.

Davis, B. (1997). Listening for differences: An evolving conception of mathematics teaching. *Journal for Research in Mathematics Education, 28*(3), 355–376.

Gordon, P. A., & Luper, H. L. (1989). Speech disfluencies in nonstutterers: Syntactic complexity and production task effects. *Journal of Fluency Disorder, 14,* 429–445.

Harlen, W. (2005). Teachers' summative practices and assessment for learning—Tensions and synergies. *The Curriculum Journal, 16*(2), 207–223.

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (pp. 87–102). Thousand Oaks, CA: Sage Publications.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Jacobs, V. R., Lamb, L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202.

Janesick, V. J. (1999). A journal about journal writing as a qualitative research technique: History, issues, and reflections. *Qualitative Inquiry, 5*(4), 505–524.

Klein, E. (1966). *A comprehensive etymological dictionary of the English language: Dealing with the origin of words and their sense development thus illustrating the history of civilization and culture*. New York, NY: Elsevier.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254–284.

Mason, J. (2002). *Researching your own practice: The discipline of noticing*. New York, NY: RoutledgeFalmer.

Mason, J. (2011). Noticing: Roots and branches. In G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 35–50). New York, NY: Routledge.

Schoenfeld, A. H. (2011). Noticing matters. A lot. Now what? In G. Sherin, V. R. Jacobs, & R. A. Philipp (Eds.), *Mathematics teacher noticing: Seeing through teachers' eyes* (pp. 223–238). New York, NY: Routledge.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York, NY: Basic Books Inc.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29* (7), 4–14.

Sherin, G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teacher's eyes*. New York, NY: Routledge.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy & Practice, 17*(4), 399–417.

Suurtamm, C., & Vezina, N. (2010). Transforming pedagogical practice in mathematics: Moving from telling to listening. *International Journal for Mathematics Teaching and Learning, 31*. Retrieved from http://www.cimt.org.uk/ijmtl/.

Van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy*. Albany, NY: State University of New York Press.

Van Manen, M. (2014). *Phenomenology of practice: Meaning-giving methods in phenomenological research and writing*. New York, NY: Routledge. (First published by Left Coast Press).

Voerman, L., Meijer, P. C., Korthagen, F. A. J., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education, 28*(8), 1107–1115.

Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.

Wiliam, D. (2010). The role of formative assessment in effective learning environments. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning using research to inspire practice* (pp. 135–159). Paris, France: OECD.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537–548.
Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. A. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 29–42). New York, NY: Teachers College Press.

Author Biography

Jimmy Pai is a secondary mathematics teacher in Ottawa, Ontario, Canada. He has been teaching mathematics in the Ottawa Carleton District School Board for the past 8 years. He also recently completed a Master's thesis at the University of Ottawa.

Chapter 4
Using Self-assessment for Individual Practice in Math Classes

Waldemar Straumberger

Abstract In this paper, the format of the self-diagnosis sheet is presented as a means to help students self-assess their confidence using competencies for individual practice. The reported study investigates evidence of the accuracy of judgments based on self-assessments that were collected throughout a year in a classroom setting. The paper further presents research from classroom assessment, as general theory framework, and educational psychology, which provides the methods for analysing the judgments of self-assessment. First results show an improvement of accuracy and a decline of overestimation after time.

Keywords Formative assessment · Self-assessment · Self-diagnosis sheet Individual practice

4.1 Introduction

The findings of the first PISA-study, specifically for Germany, indicated that the ability of teachers to identify weak students is missing (Horstkemper 2006). As a consequence, interest in educational diagnosis increased among the teacher community (Winter 2006). Another consequence was a realization that teachers had an obligation to help each learner be successful within the German curriculum. Under the label, "Diagnose und Förderung," which could be translated as "diagnosis and promotion of learning," procedures were grouped to assess the achievement level of students and provide support for advancing their learning process in the classroom. In this area, different materials were created and introduced, like digital diagnosis tests (Hafner 2008) or other analogue material (Salle et al. 2011), in order to integrate the approach of "Diagnose und Förderung" into the classroom. This development is comparable to the advent of formative assessment in the Anglo-American arena in the last 20 years (Maier 2010; McGatha and Bush 2013).

W. Straumberger (✉)
University of Bielefeld, Universitätsstraße 25, 33615 Bielefeld, Germany
e-mail: wstraumberger@uni-bielefeld.de

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_4

Self-diagnosis became popular in Germany as an opportunity to reach the goals of individual promotion of each student in the classroom (Bruder 2008). What's special is that self-assessment is used as one possibility to assess the achievement level of students during individual practice. In addition, it lightens teachers' assessment workload and creates an opportunity to support students more individually during practice phases. Experience Reports about its utilization in classrooms described possible intervention scenarios using self-diagnosis as a structured element in practice phases (Achilles 2011; Bruder 2008; Fernholz and Prediger 2007; Reiff 2006, 2008). This study examines the self-diagnosis sheet as an aid to integrate self-assessment into classrooms. The focus is on embedding experiences in existing theory and examining these experiences in a natural classroom setting.

As can be inferred from the term *self-diagnosis*, learners themselves perform diagnosis of their level of performance instead of the teacher. The underlying assumption is that the learner knows best what he knows and what he does not know. The self-diagnosis sheets work as a tool by presenting currently important skills to learners and giving them the opportunity to self-estimate and provide exercise recommendations for each of the listed competencies (Straumberger 2014). In this way, all students have the opportunity to practice individually according to their level of progress (Reiff 2006).

The self-diagnosis sheets resemble the more familiar format of a rubric. Both are instruments of self-assessment, which could be used in the context of independent learning and individual practice (Achilles 2011; Meier et al. 2007). They are used as an aid to allow learners to judge their own level of performance. However, they also differ depending on their applications. A rubric describes different qualities of competencies. These different qualities are formulated as performance levels of competencies and are often organized in a grid as in Fig. 4.1.[1] Each cell describes minimum requirements, making it possible to classify one's own performance in relation to established guidelines and to have the requirements for the next level in sight (Depka 2007). Rubrics are used to guide learners through new content. So they could be used for independent practice of skills or for self-study on new content (Danielson and Marquez 1998), primarily on more complex content (Meier et al. 2007).

In contrast to rubrics, self-diagnosis sheets do not contain specifically formulated competency levels and the different performance levels of expertise are, for example, represented through icons. In the presented study, smile icons were used to express the students' confidence with the competencies (see Fig. 4.2). This allows their use for more fundamental competencies in which a formulation of different levels would be difficult or impossible for the teacher. The description of the competencies in a self-diagnosis sheet uses prototypical tasks in order to prevent any barriers that might arise from verbal descriptions (see Fig. 4.3). In addition to

[1]It is just a scheme presented, because in the study no rubrics are used. Examples of rubrics can be found under the following links: http://www.learnalberta.ca/content/ieptLibrary/lib04.html and http://gbdsacmath.blogspot.de/2014/03/the-role-of-rubric.html.

	Level 1	Level 2	Level 3	Level 4	Level 5
Competence 1					
Competence 2					
Competence 3					
Competence 4					
Competence 5					

Fig. 4.1 Scheme of a rubric

Was ich schon alles kann - Trainingsplan

Name: ____________

Datum: 31.3.2014 Farbe:
Datum: 3.4.2014 Farbe:

Wie sicher fühlst du dich?	☺☺	☺	😐	☹	Das kannst du Üben	Erledigt/ Kontrolliert
1. Ich kann die Punkte und Figuren in ein Koordinatensystem eintragen. Beispiel: Dreieck A (0|1), B (5|2) und C(2|4)		X			S. 134 Information; S. 134 Ü 2 ✓, 3, 4, 5, und 6, S. 149 Ü 4 AH: S. 40 Ü 1 ✓	
2. Ich kann die Seitenlängen und den Umfang von Vielecken bestimmen.	X		X		S. 130 Information; S. 130 Ü 2 ✓; S. 131 Ü 6; ✓ S. 132 Ü 11 AH: S. 39 Ü 1 ✓	
3. Ich kann zwei zueinander parallele Geraden mit einem Abstand von 3 cm zeichnen.			X		S. 144 Information; S. 144 Ü 3 und 4; S. 145 Ü 8, 9 und 10; S. 149 Ü 5 AH: S. 41 Ü 1 und 3	
4. Ich kann eine Senkrechte zu einer Geraden zeichnen, die durch einen bestimmten Punkt geht.			X		S. 139 Information; S. 140 Ü 2, 3, 4 und 6; S. 149 Ü 5 AH: S. 41 Ü 1 und 2a)	
5. Ich kann den Abstand zwischen einem Punkt und einer Geraden bestimmen.		X	X		S. 141 Information; S. 142 Ü 2, 3, 4, 5 und 6 AH: S. 41 Ü 2b) ✓	

Fig. 4.2 Example of the utilized self-diagnosis sheets from the study

avoiding language barriers, it is also important to make sure that the prototypical items are in accordance with the already known tasks from the classroom (Brown and Harris 2013), so that students can connect their experiences of the competencies with the prototypical tasks. Because of this, the prototypical tasks would differ depending on the classroom practice. In addition, to facilitate autonomous practice, the exercise recommendations of the self-diagnosis sheet refer to familiar material that is already used in class. In most cases, self-diagnosis sheets are used for practice lessons before class tests to direct the learners' work on their individual deficits (Achilles 2011).

Prototypical tasks	Rating scale	Exercise recommendations
I can sign in points and figures into a coordinate system. Example: Triangle A(0|1), B(5|2) and C(2|4)		
I can determine the side lengths and the circumference of polygons.		
I can draw two parallel lines with a distance of 3 cm.		
I can draw the perpendicular to a line passing through a certain point.		
I can identify the distance between two points.		

Fig. 4.3 Scheme of a self-diagnosis sheet with examples for prototypical tasks, containing translated items of Fig. 4.2

4.2 Literature Review

This section contains reports presenting initial findings about the use of self-diagnosis sheets. The general theoretical framework centers around the use of classroom assessment, which includes self-assessment, to classify experience reports into existing research and theory. Theories and research from pedagogical psychology present a detailed look at the judgments in self-assessment.

Reports about experiences with self-diagnosis sheets confirm positive benefits of self-diagnosis sheets to individualize phases of exercise in school (Achilles 2011; Reiff 2008). There is often a lack of accuracy of the learners' estimations in the beginning, which are supposed to improve over time. Furthermore, the autonomous practice is hypothesized to have a positive impact on the learning culture in the classroom.

Classroom assessment could be done by the teacher, peers, or the students themselves. Self-assessment thereby refers to all processes of the learners to identify or diagnose their own learning progress (Brown and Harris 2013). Self-assessment in the classroom can be based on rubrics, criteria, self-ratings, or estimates of performance (Brown and Harris 2013). Rubrics could be either formative or summative assessments in the classroom, depending on their use. One example for a summative assessment is the use by the teacher to rate student work based on the criteria formulated in a rubric. Rubrics used in combination with self-assessment are only conditionally suitable for summative assessment (Brown and Harris 2013), therefore they are more often used as a part of a formative assessment.

On the theoretical level, especially referring to theories of self-regulated learning, the use of self-assessment brings advantages to the classroom. It is argued that the practice of self-assessment increases students' responsibility for their own learning process as well as academic performance by the use of metacognitive processes as well as increasing motivation and engagement in learning (Brown and Harris 2013). Self-assessment can reduce the formative assessment workload of educators. Already available self-assessment material only has to be adapted to the class. The independent work of the students enables the teacher to provide individual support during practice. In their literature review, Brown and Harris (2013) identified various studies that report high and low effects of the use of rubrics based on differing study designs and possibilities to use self-assessment in the classroom.

To date there are no studies in mathematics education that have examined the effects of self-diagnosis sheets or the underlying assumption for the use of self-diagnosis sheets. In educational psychology, the general assumption that the learner knows best what he knows and what he does not know is criticized (Kruger and Dunning 1999). Within educational psychology, the fit of judgment about performance and shown performance is examined under the concept of calibration (Bol and Hacker 2012; Schraw et al. 1993; Stone 2000; Winne and Azevedo 2014). Calibration was first discussed by Flavell (1979) in his contribution to metacognition and cognitive control and has been increasingly examined since then (Bol and Hacker 2012). Kruger and Dunning first showed that incompetent people do not realize that they are incompetent (Kruger and Dunning 1999). But not only incompetent people seem to be unable to assess their performance properly, top performers also have problems. In comparison, within their peer group, top performers underestimate their own competence (Ehrlinger et al. 2008). The research is often focused on the accuracy or fit of judgments and performance (Schraw 2009) and on factors that affect the accuracy and methods for measuring accuracy.

Several factors have been identified as having an influence on judgments of self-assessment. The factors most frequently mentioned in the literature are the age of the learner, the individual level of performance, and the type of tasks used in the assessments (Bol and Hacker 2012; Brown and Harris 2013). Although the literature does not provide any specific age levels, children's ability to judge their own performance seems to increase in accuracy in relation to their age (Brown and Harris 2013; Garner and Alexander 1989). A reason for this might be that younger children seem to be more optimistic than older children (Brown and Harris 2013). There are similar findings with respect to the level of performance. Individuals at a higher level of performance are more likely to accurately estimate their performance. Individuals at lower levels of performance display less accurate estimations of their performance with a tendency to overestimation (Bol and Hacker 2012). This finding is confirmed by Kruger and Dunning (1999).

Most of the previous studies on self-assessment worked with college students in non-naturalistic learning settings, without authentic tasks (Bol and Hacker 2012). The existing findings are restricted by the primarily laboratory design of the studies. In addition to the focus on college students, most studies include only one or two points of measurement for data collection. There is only a rare use of mathematics as the content for the self-assessments.

Methods for measuring the accuracy of calibration are grouped into absolute accuracy and relative accuracy (Hacker et al. 2008; Schraw 2009). Methods of absolute accuracy examine differences between judgments and performance. Methods of relative accuracy, however, investigate the relation of judgments and performance. In this study, the emphasis is on absolute accuracy. To evaluate the data, the *absolute accuracy index* (*AAI*) and *Bias* are used. The *AAI* measures differences between judgments and performance by the sum of squared differences, divided by the number of competencies (Schraw 2009). Low scores indicate a high accuracy and high scores a low accuracy of the calibration. *Bias* measures the direction of the error by summing the differences divided by the number of competencies (Schraw 2009). In doing so, negative values indicate a tendency to underestimation and positive values indicate a tendency to overestimation. The *AAI* and *Bias* represent different perspectives to the construct of calibration and should, therefore, be interpreted in relation to one another.

Many of the research results from classroom assessment and calibration research were carried out under less naturalistic settings in laboratory studies with college students. These studies showed a relation between performance and accuracy of self-assessment, which is also reported in the experience reports using self-diagnosis sheets in the classroom. Taking these findings into account, it appears necessary to examine the effects of self-diagnosis sheets and to clarify how the experience reports about self-diagnosis sheets as part of classroom assessment are compatible with existing findings made in studies with a less naturalistic setting. Additionally, studies about the development of self-assessment in natural settings are needed (Bol and Hacker 2012), including the development of accuracy in classroom settings.

4.3 Study Design

The aim of the study reported here was to collect data about self-assessment and performance in a natural learning environment in order to examine the accuracy of judgments. The study included 48 students (25 female and 23 male) from the fifth grade (age 10–11) of a secondary school in a German city. The self-assessment was recorded with the self-diagnosis sheet and the performance with a test, a more formal assessment. The accuracy of the judgments should indicate the ability of the students to use the self-diagnosis sheets as a tool for practice phases.

The following questions arise from the experience reports. Does the overestimation of students decrease when using self-diagnosis sheets and does the self-assessment get more accurate? From classroom assessment, how is the accuracy of judgments affected by performance and are there differences in the accuracy of judgments for groups with different performance?

Table 4.1 Design of the practice phases

Lesson	Actions in classroom	Data-collection
1	• Self-assessment with self-diagnosis sheet • Individual practice based on self-diagnosis sheet	• Self-assessment before practice
2	• Individual practice based on self-diagnosis sheet	
3	• Self-assessment with self-diagnosis sheet • Test and self-checking	• Self-assessment before test • Performance test
4	• Individual practice based on self-diagnosis sheet	

To ensure the practicability of the study design for application in regular classes, no additional lessons or materials were used, except for the self-diagnosis sheets and corresponding tests. During the year, parallel lessons with four practice phases at intervals of two months took place in three classes. Each practice phase consists of four lessons, followed by a class test, with the aim for the learner to prepare independently and individually for the class test. At the beginning of the practice phase, the students evaluate their confidence with the competencies and then independently start to practice (see Table 4.1). In the third lesson, the learners assess their confidence again and perform a test immediately after that, which they also correct on their own. The test examines the skills of the self-diagnosis sheets and was designed to be completed in less than half an hour. Finally, the fourth lesson provides an opportunity for the students to practice for the class test again.

The self-diagnosis sheets and the test were both collected and digitalized after the third lesson, so that they could be promptly returned to the students. The self-diagnosis sheets recorded the judgments at the beginning of the exercise period as well as the judgments before the test. Deliberately the test is not used before the third lesson, allowing the students to practice on the basis of their self-assessment and not to fully rely on their test results.

4.4 Results

Judgments and performance were modelled on a four-step scale from 0 to 1. For the judgments, the four steps relate to the rating scale (Fig. 4.2), with 0 for unconfident and 1 for very confident. For performance, the relation of right solutions in the test was modelled (nothing; less than 50%; at least 50%; everything). It is assumed for the judgments and for performance that the four steps have approximately the same distances between each other. Table 4.2 shows the average performance score of the competencies tested in the third lesson of the practice phases within classes and overall. The average performance of all students develops positively throughout the practice phases. The lowest average value of performance is in the second practice phase and shows a rising tendency in the following phases on average for all students.

The development of the values of absolute accuracy differ in all classes (see Table 4.3). In class 1 and class 3, the *AAI* values are nearly similar. Their values differ slightly at single measurement points and show a similar development in the course of the school year whereas they differ more in *Bias* score. The average value of class 1 shows a steady reduction in overestimation, whereas in class 3 the *Bias* shows more overestimation on average. The measured values of class 2 differ in *AAI* and the *Bias* in almost all practice phases as well as in their development throughout the year. Students in this class already show good accuracy at the beginning and little overconfidence on average. After that, the values of *AAI* rise till the third measurement point and the values of *Bias* fluctuate between the measurement points.

For the analysis of the relation between performance and self-assessment, the 48 learners were grouped by conducting a hierarchical cluster analysis (Table 4.4), with the aim to get more homogenous groups of students (Wendler and Gröttrup 2016).

Table 4.2 Average scores of performance (standard deviation)

Assessment	1	2	3	4
Class 1	0.525 (0.228)	0.500 (0.178)	0.700 (0.230)	0.600 (0.158)
Class 2	0.667 (0.201)	0.496 (0.214)	0.627 (0.193)	0.667 (0.168)
Class 3	0.596 (0.195)	0.497 (0.209)	0.690 (0.235)	0.729 (0.178)
All students	0.594 (0.216)	0.498 (0.201)	0.674 (0.223)	0.685 (0.175)

Table 4.3 Average scores of absolute accuracy (standard deviation)

Assessment		1	2	3	4
Class 1	*AAI*	0.176 (0.120)	0.186 (0.069)	0.174 (0.134)	0.125 (0.096)
	Bias	0.171 (0.187)	0.156 (0.201)	0.063 (0.206)	0.033 (0.168)
Class 2	*AAI*	0.111 (0.103)	0.139 (0.098)	0.175 (0.148)	0.157 (0.088)
	Bias	0.046 (0.205)	0.194 (0.161)	0.098 (0.256)	0.148 (0.200)
Class 3	*AAI*	0.197 (0.110)	0.219 (0.141)	0.197 (0.150)	0.139 (0.114)
	Bias	0.208 (0.169)	0.252 (0.223)	0.145 (0.217)	0.024 (0.199)
All students	*AAI*	0.166 (0.117)	0.184 (0.113)	0.182 (0.144)	0.138 (0.102)
	Bias	0.149 (0.198)	0.202 (0.203)	0.103 (0.229)	0.064 (0.197)

Table 4.4 Allocations of students by cluster

Cluster	Students	Female	Male
1	23	11	12
2	14	10	4
3	11	4	7

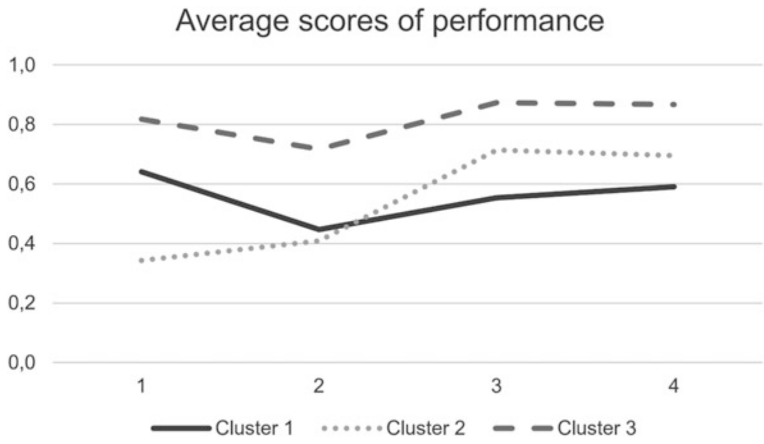


Fig. 4.4 Average scores of performance by cluster

Specifically, clusters were analysed with average linkage within groups, based on the squared Euclidean distance as the measurement scale interval (see Fig. 4.4). The learners' test performance at the four measurement points provided the variables for the grouping. Thus, the development between the four measuring points serves as the basis for the classification into different clusters. In each cluster, students are grouped with similar progress in the performance scores. It is assumed that groups of students will show different development in performance during the school year. For example, a group of top performing students with low values in *AAI* and *Bias* will represent an accurate self-assessment.

Table 4.4 presents three clusters. According to the dendrogram (see Fig. 4.5), visualizing the minimum distances in each step on the horizontal axis (Wendler and Gröttrup 2016), the largest increase of heterogeneity would arise in the last step, combining the last two clusters. In this case the cluster solution should contain two clusters referring to Wendler and Gröttrup (2016). Nonetheless the presented solution contains three clusters, because in the penultimate step combined clusters (cluster 1 and 2) differ in their average scores and describe two different developments of performance (Table 4.5). Especially, the development of cluster 2 is of interest, as discussed below.

It is assumed to be a good solution, because the average scores of performance in the clusters differ from each other and the standard deviation of each measurement point is lower in the cluster than for the whole sample (Fromm 2012). So, the aim to get more homogenous groups of students through the use of cluster analysis is achieved.

The analysis resulted in three different clusters that display different developments in test performance. Two of the three clusters (cluster 1 and cluster 3) show a perceptible drop in performance at the second practice phase, followed by an

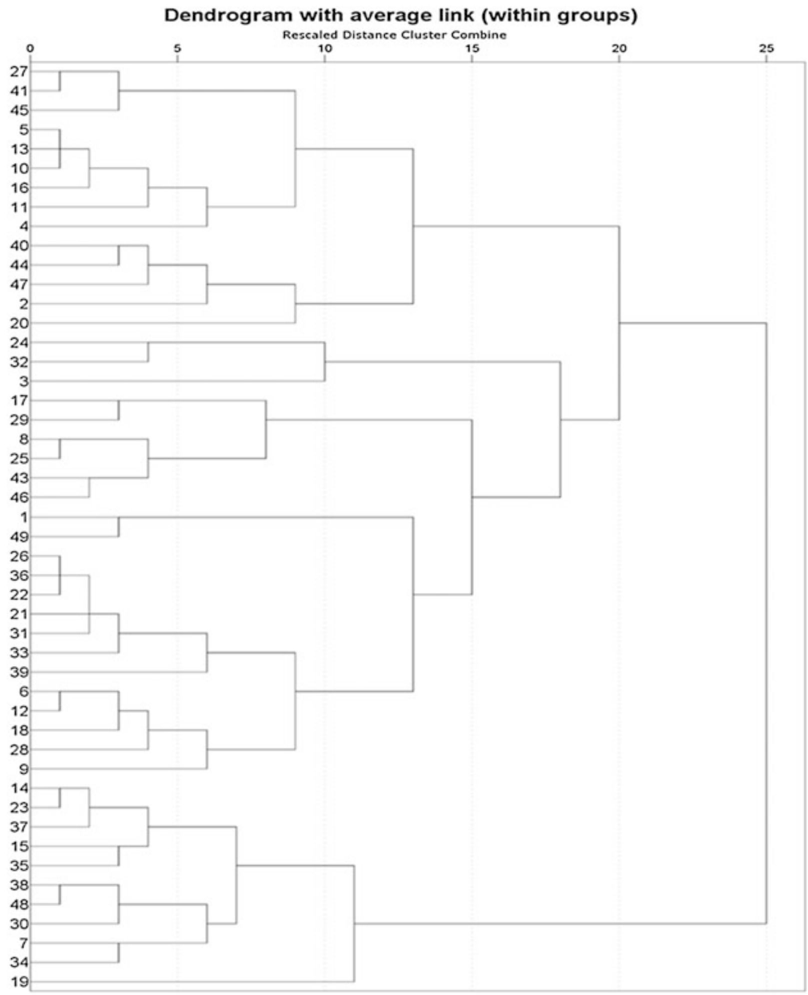


Fig. 4.5 Dendrogram of average link (within groups)

Table 4.5 Average scores of performance (standard deviation) grouped by cluster

Assessment	1	2	3	4
Cluster 1	0.641 (0.141)	0.447 (0.178)	0.554 (0.203)	0.591 (0.154)
Cluster 2	0.343 (0.097)	0.409 (0.158)	0.714 (0.160)	0.695 (0.135)
Cluster 3	0.818 (0.121)	0.717 (0.122)	0.873 (0.164)	0.867 (0.094)
All students	0.594 (0.216)	0.498 (0.201)	0.674 (0.223)	0.685 (0.175)

increase in test performance. Cluster 3 represents the efficient learner of the sample. This cluster comprises a quarter of the sample with four female and seven male students. Cluster 1 represents half of the students with a balanced proportion of gender and a test performance that is average. Cluster 2 shows the most interesting development. The learners in this group start with below-average test performance and show an improvement in the course of the evaluation. Furthermore, female students are overrepresented in this group (10 female and 4 male).

Looking at the *AAI* and the *Bias*, the three clusters also differ in their development (see Table 4.6). There is a notable decrease of the *AAI* score in clusters 2 and 3, which indicates improvement of the accuracy of self-assessment (see Fig. 4.6). For cluster 1 the *AAI* score is not decreasing as in the other two clusters. In contrast, it rises at the second and third practice phases and slightly decreases towards the end. Hence, the accuracy is not improving notably until the end of the school year. Analyses of *Bias* provide similar results (see Fig. 4.7). *Bias* of clusters 2 and 3 is approximating to zero, which implies a decrease of overestimation. This is strengthened through the development of *AAI* in the clusters, showing a decrease of differences between judgments and performance. The overestimation of cluster 1 increases after the first practice phase but also decreases towards the end.

Similar to development of performance, the *Bias* development of cluster 2 is most interesting. Cluster 2 best reflects the assumptions from the experience reports of the three groups. The accuracy is low at the beginning and the learners overestimate their performance, represented in the high value of the *Bias* at the first measurement point. But over time the overestimation decreases and the self-assessment becomes more accurate. In addition, cluster 2 supports the assumptions of a relationship between performance and the accuracy of judgments. With an improvement in test performance, accuracy of the judgments also improves, represented through the decrease of *AAI* and *Bias*. Furthermore, the other clusters support this assumption, too, because cluster 3 represents high performers that also have a higher accuracy in comparison to the other clusters, and show less overestimation in particular.

Table 4.6 Average scores of absolute accuracy (standard deviation) grouped by cluster

Assessment		1	2	3	4
Cluster 1	*AAI*	0.135 (0.094)	0.196 (0.103)	0.205 (0.158)	0.174 (0.113)
	Bias	0.108 (0.173)	0.205 (0.222)	0.191 (0.173)	0.152 (0.207)
Cluster 2	*AAI*	0.244 (0.127)	0.222 (0.116)	0.189 (0.107)	0.151 (0.078)
	Bias	0.286 (0.188)	0.254 (0.193)	0.024 (0.253)	−0.024 (0.172)
Cluster 3	*AAI*	0.123 (0.093)	0.111 (0.090)	0.127 (0.141)	0.057 (0.037)
	Bias	0.055 (0.160)	0.131 (0.147)	0.018 (0.233)	0.000 (0.117)

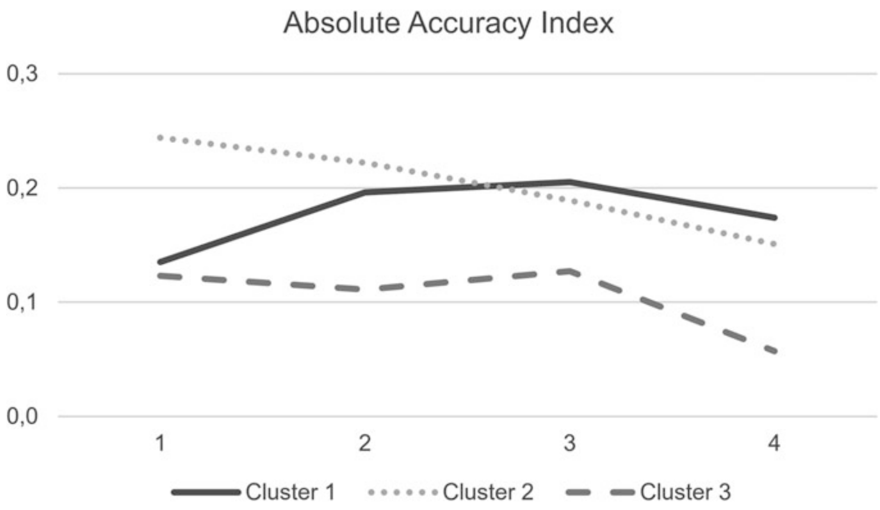


Fig. 4.6 Absolute accuracy index of cluster

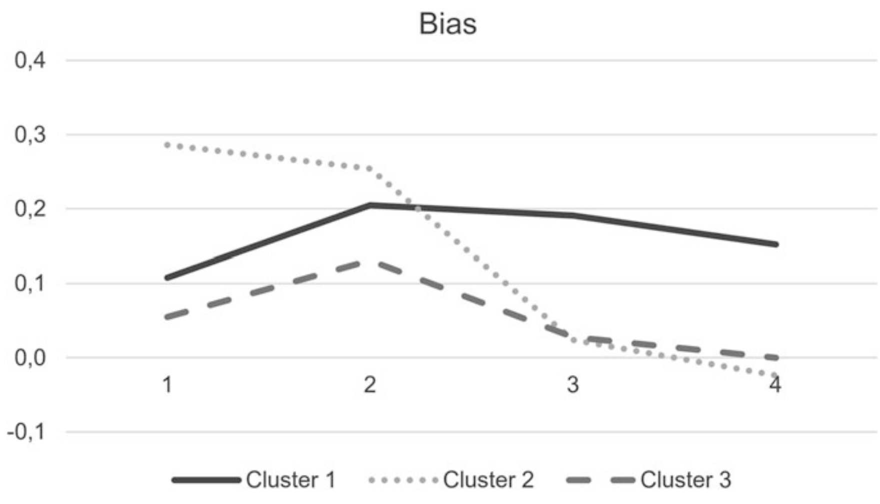


Fig. 4.7 Bias of cluster

4.5 Discussion

The data show an overall positive development of the accuracy of estimation. The slight deterioration after the first practice phase might be explained by the shift from primary to secondary school. At the beginning of the fifth grade, content from the primary school curriculum is repeated in order to harmonize the different levels of the students who come from various primary schools. Because of this, the content at the beginning of the fifth grade could be more familiar to the students than the content that follows, thereby allowing a facilitated assessment of the performance. The divergent development in class 2 could also be a consequence of the higher proportion of students who already possess a strong knowledge base from elementary school. This argumentation is strengthened by the values of performance (see Table 4.2). At the first practice phase, the performance in all three classes is higher than at the second practice phase. The difference in class 2 is the largest, because of the high performance at the first practice phase.

Apart from the low performance scores in the second practice phase, the development of the performance is positive. The performance score does not increase vastly but it shows a positive tendency. In addition, the collected data supports the findings from the experience reports on the accuracy of the judgments. At the beginning, self-diagnosis shows only a slight fit between students' self-assessment and their actual performance. However, it improves over time. The average accuracy of the judgments, measured by the *AAI*, improves towards the end of the school year and the overconfidence, measured by the *Bias*, decreases.

The findings from the cluster analysis show the accuracy is changing in connection to the development in performance. This is especially observable in the analyses of cluster 2. It comprises students who start with low performance and accuracy at the beginning and increase their performance and accuracy towards the end of the assessment. This positive development is also observable in their scores of absolute accuracy. The errors in self-assessment are decreasing and the direction of the errors changes from high overestimation in comparison to the rest of the sample to a slight underestimation. Thus, there seems to be an adjustment of self-assessment taking place when performance changes. Cluster 3 comprised of the top-performers shows the best rates in *AAI* and *Bias* throughout the assessment.

The presented results contribute to the existing findings on self-assessment as a part of classroom assessment as well as to metacognitive judgments in the field of pedagogical psychology. The accuracy seems to be partly influenced by individual performance, as the data of the second practice phase, in particular, show. In this phase, individual performance decreases in all classes as well as the accuracy, as the increase of the *AAI* and the *Bias* display. After the second practice phase, performance and accuracy both improve across all classes.

Despite the limitations of the small sample size, the presented results can be interpreted as a first confirmation of the experience reports. Recording longitudinal data of larger samples using self-diagnosis sheets would require a lot of organization. In addition to organization, there are many other aspects that have to be controlled to generate comparable groups of students for research. In the preliminary study, care was taken to have comparable conditions for implementation in the classes, in order to minimize disturbing effects. Using large samples, like classes from different schools, has the consequence that some requirements must be abandoned. It would be hard to find schools that are using the same book and school curriculum to generate groups with a comparable learning progression. A larger sample would make potential disturbing effects less strong.

Qualitative research might help obtain finer grained analyses of the effects of self-diagnosis sheets as part of classroom assessment. Another aspect, which results from the small sample size, is the difficulty of using a control group in the study design. In this study, a control group was not used, so as to avoid reducing the size of the already small sample.

Despite these first positive results, it is necessary to further investigate the use of self-diagnosis sheets. The presented findings need to be replicated in other samples to confirm the benefits of self-diagnosis sheets as part of classroom assessment. Another point of interest is whether similar clusters could also be found in other studies. It seems clear that a group of top performers could be identified in each sample but perhaps it would be more interesting to examine if there is a group of low performers who benefit from the use of self-diagnosis sheets. Researchers should also examine to what extent self-assessment develops over several school years and where factors contribute to its improvement or deterioration.

References

Achilles, H. (2011). Selbst organisierte Prüfungsvorbereitung mithilfe von Selbsteinschätzungsbogen unterstützen [Support self-organized examination preparation using self-assessment sheets]. *PM. Praxis der Mathematik in der Schule, 41,* 17–22.

Bol, L., & Hacker, D. J. (2012). Calibration research: Where do we go from here? *Frontiers in Psychology, 3,* 229.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367–393). Los Angeles, CA: SAGE.

Bruder, R. (2008). Üben mit Konzept [Practicing with concept]. *Mathematik Lehren, 147,* 4–11.

Danielson, C., & Marquez, E. (1998). *A collection of performance tasks and rubrics: High school mathematics*. Larchmont, NY: Eye on Education.

Depka, E. (2007). *Designing assessment for mathematics* (2nd ed.). Thousand Oaks, CA: Corwin Press.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*(1), 98–121.

Fernholz, J., & Prediger, S. (2007). "… weil meist nur ich weiß, was ich kann!": Selbstdiagnose als Beitrag zum eigenverantwortlichen Lernen ["… because usually only I know what I can!": Self-diagnosis as a contribution to self-responsible learning]. *PM. Praxis der Mathematik in der Schule, 49*(15), 14–18.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911.

Fromm, S. (2012). *Datenanalyse mit SPSS für Fortgeschrittene 2: Multivariate Verfahren für Querschnittsdaten* [Data analysis with SPSS for advanced 2: Multivariate procedures for cross-section data]. Wiesbaden Germany: VS Verlag für Sozialwissenschaften.

Garner, R., & Alexander, P. A. (1989). Metacognition: Answered and unanswered questions. *Educational Psychologist, 24*(2), 143–158.

Hacker, D. J., Bol, L., & Keener, M. C. (2008). Metacognition in education: A focus on calibration. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 429–455). New York, NY: Psychology Press.

Hafner, T. (2008). Digitales Testen, Diagnostizieren und Fördern [Digital testing, diagnostics and promoting]. *Mathematik Lehren, 150,* 66.

Horstkemper, M. (2006). Fördern heißt diagnostizieren: Pädagogische Diagnostik als wichtige Vorraussetzung für individuellen Lernerfolg [Promoting means diagnosing: Pedagogical diagnostics as an important prerequisite for individual success in learning]. *Friedrich Jahresheft, XXIV,* 4–7.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 11–21.

Maier, U. (2010). Formative Assessment – Ein erfolgversprechendes Konzept zur Reform von Unterricht und Leistungsmessung? [Formative assessment—A successful concept for the reform of education and performance measurement?]. *Zeitschrift für Erziehungswissenschaft, 13*(2), 293–308.

McGatha, M. B., & Bush, W. S. (2013). Classroom assessment in mathematics. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 449–460). Los Angeles, CA: SAGE.

Meier, S. L., Rich, B. S., & Cady, J. (2007). Teachers' use of rubrics to score non-traditional tasks: Factors related to discrepancies in scoring. *Assessment in Education: Principles, Policy & Practice, 13*(1), 69–95.

Reiff, R. (2006). Selbst- und Partnerdiagnose im Mathematikunterricht: Gezielte Förderung mit Diagnosebögen [Self- and partner diagnosis in mathematics teaching: Targeted support with diagnostic sheets]. *Friedrich Jahresheft, XXIV,* 68–72.

Reiff, R. (2008). Selbst- und Partnerkontrolle.: Ein effizientes Verfahren zur produktbezogenen Diagnostik [Self- and peercheck: An efficient procedure for product-specific diagnostics]. *Mathematik Lehren, 150,* 47–51.

Salle, A., vom Hofe, R., & Pallack, A. (2011). Fördermodule für jede Gelegenheit: SINUS. NRW-Projekt Diagnose & individuelle Förderung [Promoting modules for every occasion: SINUS. NRW—Project diagnosis & individual support]. *Mathematik Lehren, 166,* 20–24.

Schraw, G. (2009). Measuring metacognitive judgements. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 415–429). New York, NY: Routledge.

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology, 18*(4), 455–463.

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12*(4), 437–475.

Straumberger, W. (2014). Wirksamkeit von Selbstdiagnose [Effectiveness of self-diagnosis]. In J. Roth & J. Ames (Eds.), *Beiträge zum Mathematikunterricht 2014* (pp. 1183–1186). Münster, Germany: WTM-Verlag.

Wendler, T., & Gröttrup, S. (2016). Cluster analysis. In T. Wendler & S. Gröttrup (Eds.), *Data mining with SPSS modeler* (pp. 587–712). Cham, Switzerland: Springer International Publishing.

Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of the learning sciences* (pp. 63–87). Cambridge, England: Cambridge University Press.

Winter, F. (2006). Diagnosen im Dienst des Lernens: Diagnostizieren und Fördern gehören zum guten Unterrichten [Diagnoses at the service of learning: Diagnosis and promotion are part of the good teaching]. *Friedrich Jahresheft, XXIV,* 22–25.

Author Biography

Waldemar Straumberger is lecturer and doctoral researcher at the University of Bielefeld in Germany. His research focuses on classroom assessment, especially formative assessment and possibilities to use self-assessment as part of it. He is also interested in using technology for teaching mathematics.

Part III
Technology as a Tool for Classroom Assessment

Chapter 5
Using a Digital Flip Camera: A Useful Assessment Tool in Mathematics Lessons

Ann Downton

Abstract This paper describes how two early career teachers in Australia used digital Flip cameras as an assessment tool as part of their regular mathematics lessons to improve their assessment practices. Both teachers taught Grade 1/2 and identified formative assessment in mathematics as a challenge as they had trouble collecting data that captured students' thinking and reasoning during a mathematics lesson. The purpose of this study was to identify the ways in which early career teachers utilized technology as a tool to enhance their formative assessment practices within a mathematics lesson. The findings suggest that the use of Flip cameras enabled the teachers to capture authentic assessment data and improve their documentation and analysis of the students' learning, and better cater to the students' learning needs when planning.

Keywords Formative assessment · Digital flip camera · Primary school mathematics · Teachers' use of technology

5.1 Background

A *Flip camera* or *Flip camcorder* refers to a camera with a USB arm that can plug into a computer for charging and for ease of downloading the videos (Gao and Hargis 2010). Such a tool has been used across disciplines, including in secondary schools, to promote technology-assisted active learning and increase student engagement. Using a tool such as this enables teachers to gather pre/post assessment data on how students performed on a task that can provide real-time feedback to students and enable them to reflect on their learning (Gao and Hargis 2010). Capturing videos of students' learning assists teachers to gain a sense of how conceptual understanding develops and what students are gaining from their

A. Downton (✉)
Faculty of Education, Monash University, 14 Langford Street,
Surrey Hills, VIC 3127, Australia
e-mail: ann.downton@monash.edu

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_5

learning experiences. Playing the video back to the students provides an opportunity to delve deeper into their thinking processes and problem-solving strategies post lesson (Boardman 2007; Keeley and Tobey 2011). The use of Flip cameras is a simple way to engage students in the assessment process as they are working. However, there is no current Australian research that explores their use in primary school mathematics classrooms.

5.2 Literature Review

Within Australia over recent decades, higher expectations have been placed on the assessment of students' mathematical learning and the need to move beyond traditional assessment practices that are inadequate for assessing student thinking and disposition (Masters 2013). Aligned with these expectations has been growing interest in how highly effective teachers use assessment as part of their teaching to promote improved student learning (Black and Wiliam 1998; Wiliam and Thompson 2007). Part of this reform agenda has been to have a unified view of educational assessment, namely that the "fundamental purpose of assessment is to establish where learners are in their learning at the time of their assessment" (Masters 2013, p. 6). Such a view conceptualizes assessment as the process of establishing where students are in their long-term learning in terms of their developing knowledge, skills, and understandings (Masters 2013). This holistic view concurs with those of Clarke (1997) and Wiliam (2010) both of whom maintain that assessment (formative and summative) is central to learning. Furthermore, Newhouse suggests that assessment could be termed as "making learning visible" (2015, p. 215) as the judgements teachers make are a result of either their observations or inferred from what they observed. Within the mid-1990s, formative assessment or *assessment for learning* (Wiliam 2010) became more evident in mathematics lessons in primary schools within Australia as it informs teachers during their daily decision-making in the classroom, and provides direction for subsequent planning. Black and Wiliam defined formative assessment as "all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (1998, p. 8). Formative assessment provides both teacher and students with insights into the student's learning and feeds forward into the teacher's planning. This process of collecting, storing and analyzing data to assist in making informed decisions about learning is cyclical and commonly referred to as "learning analytics" (Newhouse 2015, p. 223). McDonough and Gervasoni (1997) represented this cyclic nature of teaching linked to assessment as a process of exploring, reflecting, and responding (Fig. 5.1). They argued that it is through reflecting on insights gained by observing and listening to students during a lesson that teachers better understand how children learn mathematics. These reflections inform teacher planning on a daily basis.

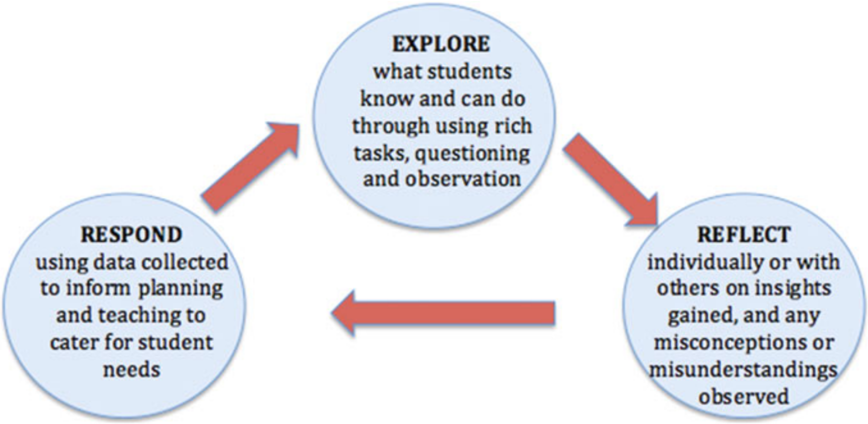


Fig. 5.1 Assessment informing teacher practice (adapted from McDonough and Gervasoni 1997)

In addition, observing and questioning are two assessment practices commonly used in the mathematics classroom that enable teachers to develop insights into students' conceptual understanding, reasoning, and disposition when considering how to adapt instruction for individuals (Clarke and Wilson 1994; Furtak and Ruiz-Primo 2008). The most effective teachers are likely to be those who approach assessment as an opportunity for students to show what they know and can do, who link assessment and teaching (making these links clear to students), and consider assessment as a precursor for action (Clarke 1997). However, research suggests that the demands on a teacher's time are such that there is often some disconnect between what is observed and noted in the enactment in the classroom, and how it is later interpreted (Della-Piana 2008). This was the issue that the two early career teachers in this study experienced prior to using the digital Flip camera in their mathematics lessons. The following research question guided the study:

> In what ways do early career teachers use technology to enhance their formative assessment practices?

5.3 Method

The two teachers reported here were part of a small study conducted in 2010 that examined Early Career Teachers' (teachers in their first three years of employment) use of Information and Communications Technology (ICT) as part of their regular mathematics lessons. Five teachers (four female, one male) who accepted the invitation to participate in the study were graduates of the university in which the researchers worked. Prior to the commencement of the study, the researchers conducted an after-school meeting at the university. In this meeting, participants explained how they intended to explore the use of ICT in their mathematics classrooms. The only technology they had used as part of their practice were calculators and the interactive white-board. Both schools had separate computer rooms.

5.3.1 Participants and Context

Julia and Sarah, both in their mid-twenties, had two years teaching experience with Grade 1/2 classes in different schools within Melbourne, Australia. Sarah's school was situated in a predominantly middle-class suburb, whereas Julia's school was situated in a less affluent area. Both cohorts of Grade 1/2 classes consisted of 24 students (aged 6 and 7).

The classroom environment of both teachers was inviting, stimulating, and well organized. The students worked in mixed-ability pairs during mathematics lessons, the duration of which was approximately 60 min per day. The lesson structure consisted of a short tuning-in task as a whole class (approximately 10 min), followed by exploration time in which students worked in pairs on an open-ended task for approximately 30 min, and then a whole class plenary session in which selected students shared their learning (approximately 15–20 min). Both classrooms were student-centered with the teacher observing, listening and interacting with the students as they worked, using probing questions such as, "Why did you choose to use that strategy?", "What is another way you might solve the problem?", "Can you prove to me that you have found all the possibilities?" The plenary time was an opportunity for the students to share their solution strategies with the class whilst the teachers guided the discussion using questioning that challenged student thinking such as, "If we change the numbers do you think that strategy would work?"

5.3.2 Teachers' Use of Flip Cameras

A key assessment focus of both teachers was students' articulation of their mathematical thinking. To elicit and gain insights into students' thinking, they were using observation and probing questions (a practice modelled during their preservice education), but they had difficulty in adequately recalling key insights that occurred in mathematics lessons and finding subsequent time to document them. Within this study, the teachers' goal was to see whether the use of a Flip camera would assist them to capture authentic assessment data relating to student thinking as it occurred in the course of a mathematics lesson. Both teachers decided to use the Flip cameras as part of the exploration time during the lesson when students worked in pairs on the given task without teacher assistance. Prior to the commencement of the study, each teacher showed the students the Flip camera and explained how it would help the teacher to learn more about students' learning. Sarah chose to use the Flip camera twice a week, whereas Julia decided to use it on a daily basis for short periods of time and as a reflective tool the following day to revisit previous learning with particular students. Replaying selected video snippets with individual students would allow Julia to challenge any misconceptions and provide students with an opportunity to consider what they might do differently and why. Sarah also planned to replay the video clips to the students post lesson for them to record their reflections on their learning in writing. Both intended to conduct 'mini interviews' with students while they were working, using probing questions to elicit their thinking.

5.3.3 Data Collection

Three sources of data were collected: the teachers' Flip videos; the researchers' observational notes of the four lessons; and the semi-structured interviews post lessons. The specific questions asked of teachers during the semi-structured interview included:

- What prompted you to use the Flip camera at that particular point of the lesson?
- Why did you choose those particular students to probe and video?
- In what way has the use of this tool assisted you to improve your formative assessment practices?
- What were some of the insights gained about the students' learning that the use of the Flip camera afforded?
- What were the challenges or limitations in using this particular tool?
- What modifications would you make for the future use of it to support your assessment practice?

Each post lesson semi-structured interview was audio-recorded and conducted immediately after the lesson for approximately 30 min in duration. Prior to the commencement of the study, the researchers visited each classroom to explain their role to the students when they visited each fortnight. So as not to distract the learning, the researchers sat on either side of the classroom and independently observed each of the eight lessons (four for each teacher) over the course of eight weeks (one lesson per fortnight in each class). They used time-point observations (Sarantakos 2013) of when the teachers used the Flip camera and the action that led to the use. Having both researchers observe the lessons increased the reliability and validity of the data. The lessons were not digitally recorded as the researchers felt an additional camera would be intrusive and unnecessary, as the purpose was to focus on the particular number of occasions that the teachers used the Flip camera.

5.3.4 Data Analysis

The audio recordings of the semi-structured interviews and the Flip videos were transcribed and analyzed using a grounded theory approach (Strauss and Corbin 1990), as were the researchers' field notes, to identify any themes that emerged over the course of the four lessons and to identify any similarities between the teachers' insights. The analysis of the semi-structured interview data is not presented in this paper due to limited space.

5.4 Results and Discussion

From the researchers' observations, it was evident that the teachers used the tool in quite contrasting ways. Sarah set the camera up on the table in front of the students whom she was observing, whereas Julia tended to hold it in her hand as a recorder while she interacted with individual students. Each student interaction in Sarah's observed lessons took approximately one minute, whereas those in Julia's class took approximately 3 min. Sarah went to particular pairs of students in each observed lesson and conducted five one-minute 'mini interviews' within each lesson as students worked on a task. In contrast, Julia roved and observed the students as they worked on a task and videoed six students in each of the four observed lessons. Both teachers gave the students approximately 10 min to engage in the task uninterrupted before videoing.

From the analysis of the post lesson interview data and video clips, it was evident that Julia's use of the Flip camera was triggered by particular observations of a student's representations, such as a mismatch between the drawing and the equation (that highlight misconceptions), or the use of sophisticated strategies. In two different lessons (observed lessons one and three), Julia noticed possible student misconceptions and captured these on video and used them as part of the plenary to gauge whether other students might have a similar misconception or if they could identify what was wrong.

In contrast, Sarah selected five students prior to each lesson with whom she conducted one-minute 'mini interviews' within each lesson, and used the same line of questioning with each student to assist with her post-lesson analysis. Her selection of students was informed by earlier analysis of student assessment data relating to their understanding of a concept or development of a skill, with the intention of having a snapshot of all her students' explanations of thinking and their learning by the end of the school term. Sarah's focus was on collecting assessment data for post-lesson analysis and student reflection, whereas Julia utilized some of the data during the plenary, when appropriate. Both used the video clips post lesson to enable students to reflect on their learning. Sarah's children recorded their learning in writing and their reflections were made into a PowerPoint presentation to add to their digital portfolios. Julia encouraged her students to think about the mathematics they were using and how they might solve the problem differently. In this way, Julia was using the video data to delve more deeply into the students' thinking processes and problem-solving strategies (Keeley and Tobey 2011).

Analysis of the video clips also revealed that questioning was a strong aspect of both teachers' assessment practices. Part of the questioning was to elicit student thinking, use of mathematical language, and being able to convince the teacher. Examples of these were evident in the following dialogue of a one-minute clip from Sarah's second observed lesson and dialogue from a three-minute clip from Julia's third and fourth observed lessons.

5.4.1 Dialogue Between Sarah and David

Sarah interviewed David (Year 1 student), the third child recorded during this lesson, the focus of which was subitising collections and building up number facts to ten. He was working with Oscar and playing a game of 'How many counters are hiding?' in which students take turns to arrange some counters (no more than 10) under a card, while their partner looks away. The counters are revealed to the partner for 10 seconds and then covered again, and he/she has to work out the quantity of hidden counters.

S: How many counters are under the card?
D: I think there are 10.
S: What did you see?
D: 2 on the top, 3 on the side, and 3 on the other side.
S: So how many is that so far?
D: 8 and then I sort of saw 2 on the bottom at the side.
S: Can we see what was there? Lift the lid please Oscar.
D: Two at the top, 3 at the side and 3 on the other side and only 1 at the bottom so there are 9.
S: How do you know it is 9 and not 10?
D: Because there is only 1 at the bottom not 2.
S: Can you prove it to me in another way?
D: Double 3 is 6, and another 3 is 9.

Sarah's assessment focus was on students' ability to subitise small collections, make reasonable estimates, and to find the quantity of a collection without using a 'count by ones' strategy. She later recorded that David's estimation was good, as was his visualization and explanation of the arrangement of the collection and use of doubles as a strategy to prove that he was correct.

5.4.2 Dialogue Between Julia and Kevin

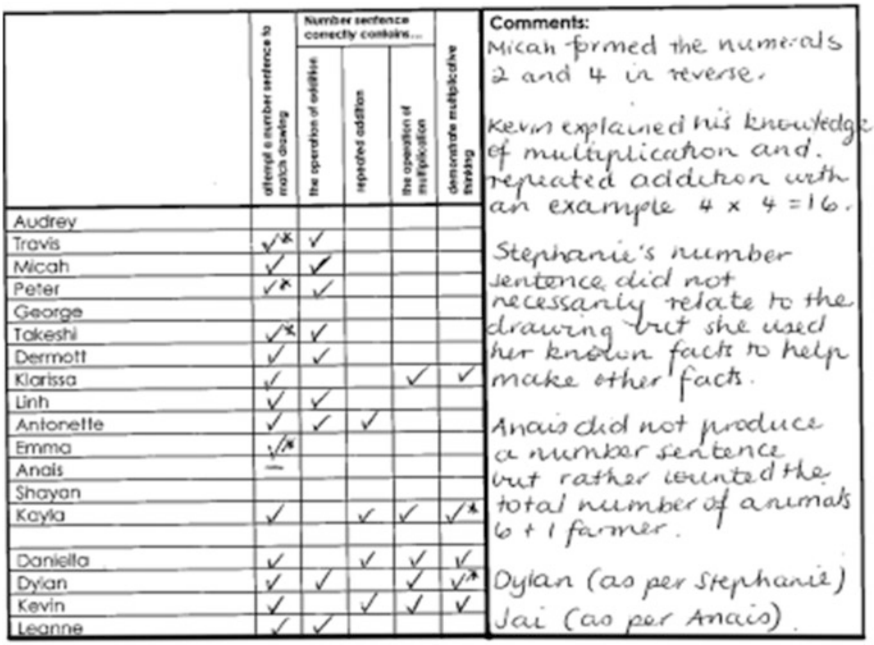
Lesson 3 focus: Generating and interpreting a multiplication situation. Students were given a template with 20 legs from the Animal legs task (Downton et al. 2006), then drew a possible number of animals required, and expressed this as a number sentence (see Fig. 5.2). Julia chose to interview Kevin as she was perplexed by his recording of two contrasting number sentences, one reflecting multiplicative thinking, and the other, additive thinking.

J: Tell me about your task? What animals do you have in your farm?
K: A person, a cow, donkey, goat, pig, and a chicken.
J: How did you write this number sentence ($4 \times 4 + 2 \times 2 = 20$) to match your picture?
K: 2 animals that have 2 legs and 4 animals have 4 legs.
J: Which number sentence did you write first?

Fig. 5.2 Kevin's work sample

K: 4 times 4 plus 2 times 2.
J: How did you work it out?
K: I was thinking of something like that before when we did the Teddy Bear's picnic.
J: So how did you get this number sentence from there?
K: The long way of 4 times 4 is 4 plus 4 plus 4 plus 4 and the long way of 2 times 2 is 2 plus 2.
J: How did you know that this sort of addition is related to multiplication?
K: Times is the same as groups of, 4 times 4 equals 16, and when you do 4 groups of 4 it equals 16. (He drew dots in each of the circles.)
J: Now try and figure out whether there is a different solution.

Julia asked Kevin to explain the thinking he used, and she recorded insights on a checklist (see Fig. 5.3), which she elaborated on post class when she replayed the video clips of each student. She noted that Kevin demonstrated his understanding of multiplication using the notation $4 \times 4 + 2 \times 2 = 20$, and clearly explained the relationship between repeated addition and multiplication.



	attempt a number sentence to match drawing	Number sentence correctly contains... the operation of addition	repeated addition	the operation of multiplication	demonstrate multiplicative thinking
Audrey					
Travis	✓*	✓			
Micah	✓	✓			
Peter	✓*	✓			
George					
Takeshi	✓*	✓			
Dermott	✓	✓			
Klarissa	✓			✓	✓
Linh	✓	✓			
Antonette	✓	✓	✓		
Emma	✓*				
Anais	—				
Shayan					
Kayla	✓		✓	✓	✓*
Daniella	✓		✓	✓	✓
Dylan	✓	✓		✓	✓*
Kevin	✓		✓	✓	✓
Leanne	✓	✓			

Comments:
Micah formed the numerals 2 and 4 in reverse.

Kevin explained his knowledge of multiplication and repeated addition with an example 4 x 4 = 16.

Stephanie's number sentence did not necessarily relate to the drawing but she used her known facts to help make other facts.

Anais did not produce a number sentence but rather counted the total number of animals 6 + 1 farmer.

Dylan (as per Stephanie)
Jai (as per Anais)

Fig. 5.3 Excerpt of Julia's assessment checklist related to the Animal legs task

5.4.3 Dialogue Between Julia and Miranda

Lesson 4 focus: Exploring the perimeter of 2D shapes using a range of materials. In particular, Julia was interested in observing whether students used the material correctly (i.e., put materials end to end, leaving no gaps), recorded the quantity and the unit, and if they could generalize that the larger the unit, the less quantity required. Her assessment focus was on students' reasoning and their demonstrated understanding of the measuring process. While roving, Julia noticed that Miranda was perplexed at her answers, which she knew were incorrect but was not sure why.

J: Miranda, what's wrong?
M: When I counted the big blocks there were 24 and when I counted the little blocks there were 23.
J: Why do you think that's a problem?
M: Because these are littler there should be more, and the big ones should be less, because they are bigger.
J: Ok, so what are you going to do to solve this problem?
M: Redo it and do the same thing I did with the big blocks.
J: So where do you think you went a bit wrong? Where do you reckon the problem is?
M: I think I didn't fill up the little gaps and points here, but I did with the big blocks.
J: Ok, so check it and see.

Julia noted Miranda's problem-solving strategies, her explanation of thinking and her awareness of the relationship between the size of the unit, and quantity required. Julia also used the video of Miranda's dilemma as part of the plenary discussion to draw out the generalization and considerations when measuring.

5.4.4 Summary and Insights from Post-lesson Reflections

Over the course of the study, both teachers reported that their lesson planning was more focused and their assessment within their daily practice more purposeful and targeted. They both documented insights from the video recordings on an Excel spreadsheet (see Fig. 5.4). The spreadsheet was designed so that it could be used for any lesson. The full spreadsheet has ten foci: nature and quality of written work, use of problem-solving strategies, reasoning and mathematical skills, use of materials/diagrams, engagement, explanation of thinking, organisation, persistence, use of generalisation, and confidence. Under each heading are descriptors of what the teachers might observe. Additional descriptors were added if needed. For example, if a student used doubling and halving it would be added under reasoning and mathematical skills. The excerpt provided (Fig. 5.4) includes two distinct lessons, one relating to perimeter in which Miranda was interviewed, and the other to the multiplication in which Kevin was interviewed. For the purpose of this paper, both insights were shown on the one spreadsheet and the lesson focus and student names were merged into one column. The teachers had separate spreadsheets for each topic, and they could extract specific information about particular students. Doing so assisted them to keep more accurate records of students' learning, provided an efficient way to develop a profile of each student's mathematical understanding over the course of the year, and informed future planning related to students' needs (Boardman 2007; Masters 2013).

The following excerpts from the final post-lesson interview are in response to the question, "In what way has the use of this tool assisted you to improve your assessment practices?"

> Sarah: It's given me authentic assessment data that I can use to inform my planning and it has given me the opportunity to get the children to reflect on their learning. Capturing the children's learning in this way is really powerful and informative because it gives me a visual and audio snapshot of each child's learning and their written self-assessment. Importing the videos into the children's digital portfolios with their journal comments is further evidence of their learning and developing understandings. It is a fantastic way for parents to see authentic assessment of their child's learning. The use of the Flip camera has improved my questioning and documentation of my assessment data.

> Julia: The use of the Flip camera has given me greater opportunities to use students' thinking in the plenary and to gain deeper insights into students' developing understanding of particular mathematics concepts. The digital recordings have enabled me to collect authentic assessment data and keep accurate records of the students' learning. I would not have thought about using an Excel spreadsheet before this. I feel my planning is much more focused on the children's needs as a result, and I am much more relaxed because I know I don't have to remember things that happen in the moment. Using the videos to reflect on my own practice was a bonus as it made me realise that I needed to use deeper questions at times and give students more 'wait time'.

Students' names and video clip lesson focus	Nature and quality of written work	Use of problem solving strategies	Reasoning and mathematical skills	Explanation of thinking	Organisation
Lesson Focus	systematic thinking	trial and error	recalling facts	clear and precise	systematic
	little evidence of thinking	guess and check	counting	mostly clear but has misconceptions	efficient
1. Perimeter of shapes	how (steps only)	drawing a table	listing	lacks language to express thinking	limited skills
	evidence of making connections	make it simpler	comparing	uses mathematical language	no idea
2. Multiplication (Animal legs)	how, why and checking process	working backwards	explaining	limited understanding of concept	
	an evaluation or comparison about	using materials	justifying	high level and convincing	
	efficiency		generalising	logical	
Miranda	how and why and checking process	using materials	explaining	clear and precise	systematic
			justifying	logical	
			generalising	high level and convincing	
Kevin	evidence of making connections	drew a picture	recalling facts	uses mathematical language	efficient
	how and why and checking process		explaining	clear and precise	
			explaining	uses mathematical language	

Fig. 5.4 Excerpt of Julia's assessment by observation spreadsheet

The teachers also reported that the students were more articulate and confident in explaining and justifying their thinking during class discussions as a result of the use of the Flip cameras. These findings reflect those of Boardman (2007) who found that the use of audio recorders and digital cameras facilitated kindergarten children's reflective thinking processes.

Some of the challenges Julia mentioned related to being disciplined not to over use the Flip camera in the plenary. Both teachers indicated that they were devoting more time to reflecting on the insights gained (McDonough and Gervasoni 1997), which made their planning and teaching more focused but did limit the amount of time devoted to other areas of the curriculum.

5.5 Concluding Remarks

These findings highlight the value of using Flip cameras in the mathematics classroom as an assessment tool combined with teacher questioning to make students' reasoning visible. The teachers identified a need—how to adequately capture authentic learning as it is enacted in the classroom, and chose to use the technology that best suited their classroom practice, which is linked to the notion of 'interpretative flexibility' of technology use in the classroom (Ruthven et al. 2008). Another example of this is the variability in which the teachers used the Flip camera in their classroom. While their needs were similar, their use of the Flip camera was quite different. Sarah used it in a focused and predetermined way to collect formative assessment data that she could subsequently use to inform her practice and with the students as a reflective tool in their written self-assessment, the artefact of which became part of the student's digital learning portfolio. In contrast, Julia implemented it to capture insights into students' thinking as they engaged with a mathematical task, and used these insights in the plenary. She was particularly interested in the strategies students were using and any errors or misconceptions that were evident from observing their representations. Unlike Sarah, Julia had not pre-determined which children she would select; she used the Flip camera on a daily basis, whereas Sarah used it twice a week.

These findings also indicate the teachers' exploration of the use of technology beyond the collection of data to analysis and design of spreadsheets, digital storage, and in Sarah's case digital portfolios. This reflects Masters' (2013) suggestion that the use of technology can transform assessment processes. Furthermore, Clarke-Midura and Dede argue that digital technologies offer "exciting opportunities to design assessments that are active and situative, and that measure complex student knowledge and provide rich observations for student learning" (2010, p. 311). Both teachers indicated that they wanted to capture students' thinking, and in so doing make their learning visible (Newhouse 2015), which was achieved using the Flip camera. This resolved the issue of the disconnect between what is observed and noted in the enactment in the classroom, and how it is later interpreted (Della-Piana 2008). From these findings, it is reasonable to suggest that these teachers have a greater understanding of assessment as a cyclical process of establishing where students are in their learning at a given time in relation to their developing knowledge, skills, and understandings (Masters 2013), and reflecting on these insights to assist them in making informed decisions (McDonough and Gervasoni 1997; Newhouse 2015). Furthermore, the digital technologies made the range of approaches to assessment more accessible to and more manageable for the teachers (Newhouse 2015).

This small study only examined two teachers' flexible use of technology in early years' classrooms for a limited period of time, so one must be circumspect in relation to the findings. However, it does provide an avenue for further research regarding teachers' flexible use of technology for formative assessment practices. A larger study could examine the teachers' flexible use of technology across a range of year levels for a longer duration to gain a sense of whether such practice enables teachers to have greater insights into their students' current learning, where they are going, and how to support them to achieve that goal (Wiliam 2010). Within such a study there would be opportunities to compare and contrast the different uses, building on the work of Ruthven et al. (2008) and any evidence of improvement in teachers' formative assessment practices, subsequent classroom practice, and student learning. A further study might explore teachers' use of the data captured in the lesson as a reflective tool to assist students to reflect on their learning and set subsequent learning goals.

Acknowledgements My thanks go to my colleagues Dr. Anne Scott and Dr. Donna Gronn from the Australian Catholic University for encouraging me to further analyze the data.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Boardman, M. (2007). I know how much this child has learned. I have proof: Employing digital technologies for documentation processes in kindergarten. *Australian Journal of Early Childhood, 32*(3), 59–66.

Clarke, D., & Wilson, L. (1994). Implementing the assessment standards for school mathematics: Valuing what we see. *Mathematics Teacher, 8*(7), 542–545.

Clarke, D. J. (1997). *Constructive assessment in mathematics: Practical steps for classroom teachers*. Berkeley, CA: Key Curriculum Press.

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education, 42*(3), 309–328.

Della-Piana, G. M. (2008). Enduring issues in educational assessment. *Phi Delta Kappan, 89*(8), 590–592.

Downton, A., Knight, R., Clarke, D., & Lewis, G. (2006). *Mathematics assessment for learning: Rich tasks & work samples*. Melbourne, Australia: Mathematics Teaching and Learning Centre, Australian Catholic University.

Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education, 92,* 799–824.

Gao, J., & Hargis, J. (2010). Promoting technology-assisted active learning in computer science education. *The Journal of Effective Teaching, 10*(2), 81–93.

Keeley, P., & Tobey, C. R. (2011). *Mathematics formative assessment: 75 Practical strategies for linking assessment, instruction, and learning*. Thousand Oaks, CA: Corwin Press.

Masters, G. N. (2013). Reforming educational assessment: Imperatives, principles and challenges. In S. Mellor (Ed.), *Australian Education Review* (pp. 1–68). Melbourne, VIC, Australia: ACER Press.

McDonough, A., & Gervasoni, A. (1997). Gaining insights into children's learning of mathematics: Opportunities, strategies and responses for the classroom. In D. M. Clarke, P. Clarkson, D. Gronn, M. Horne, L. Lowe, M. Mackinlay, & A. McDonough (Eds.), *Mathematics: Imagine the possibilities* (pp. 145–151). Brunswick, VIC, Australia: Mathematical Association of Victoria.

Newhouse, C. P. (2015). Making learning visible through digital forms of assessment. In M. Henderson & G. Romeo (Eds.), *Teaching and digital technologies: Big issues and critical questions* (pp. 214–228). Melbourne, VIC: Cambridge University Press.

Ruthven, K., Hennessey, S., & Deaney, R. (2008). Constructions of dynamic geometry: A study of the interpretative flexibility of educational software in classroom practice. *Computers & Education, 51,* 297–317.

Sarantakos, S. (2013). *Social research* (4th ed.). New York, NY: Palgrave Macmillan.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications.

Wiliam, D. (2010). The role of formative assessment in effective learning environments. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 135–159). Paris, France: OECD Publishing.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Author Biography

Ann Downton is a lecturer in Mathematics Education at Monash University, Clayton Campus in undergraduate and post-graduate mathematics education and curriculum units. Her main research interests relate to how students construct their mathematical knowledge, in particular, their development of multiplicative thinking which was the focus of her doctoral studies; how teachers challenge children's mathematical thinking; catering for low-attaining students in the classroom; mathematics planning and assessment practices. Other research interests include problem posing in the primary years and links to mathematical modelling; and teachers' professional learning and their pedagogical practices.

Chapter 6
The Use of Digital Technologies to Enhance Formative Assessment Processes

Annalisa Cusi, Francesca Morselli and Cristina Sabena

Abstract We focus on formative assessment processes carried out, by the teacher and the students, through the use of digital technologies. The research is situated within the European Project FaSMEd, in which a new model connecting the role of technology to classical views on formative assessment is proposed. Through data analysis from teaching experiments in Italian schools using connected classroom technology, we highlight how the specific choices concerning the use of technology and the design of the activities can enable the enactment of formative assessment strategies at the teacher's, the students', and the peers' levels.

Keywords Formative assessment strategies · Connected classroom technologies Making thinking visible · Argumentation · Teaching-learning processes

6.1 Introduction

Digital technology can provide great support both to students and teachers in getting information about students' achievement in real-time. Relying on this hypothesis, the FaSMEd Project, "*Improving progress for lower achievers through Formative Assessment in Science and Mathematics Education*," investigates, by means of design-based research methodology (Cobb et al. 2003), the role of technologically enhanced formative assessment (FA) methods in raising students' attainment levels, with special attention to low achievers (for a detailed presentation of the project, see also Wright et al., in this volume).

A. Cusi (✉) · C. Sabena
Department of Philosophy and Education, University of Turin,
Via Gaudenzio Ferrari 9, 10124 Turin, Italy
e-mail: annalo@tin.it

F. Morselli
Department of Mathematics, University of Genova, Via Dodecaneso 35,
16146 Genoa, Italy

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_6

Several studies provide evidence about how new technology can be used as an effective tool in supporting FA processes (Quellmalz et al. 2012). In our design experiments in Italy, we focus in particular on connected classroom technologies that are networked systems of personal computers or handheld devices specifically designed to be used in a classroom for interactive teaching and learning (Irving 2006). Research has shown that connected classroom technology may provide several opportunities for classroom practice: create immersive learning environments that give powerful clues to what students are doing, thinking, and understanding (Roschelle et al. 2004); make students take a more active role in the discussions (Roschelle and Pea 2002); encourage students, through immediate private feedback, to reflect and monitor their own progress (Roschelle et al. 2007); and enable teachers to monitor students' progress and provide appropriate remediation to address student needs (Irving 2006).

Within the FaSMEd Project, we designed mathematical activities and developed a methodology aimed at fostering formative assessment processes by means of connected classroom technology. The overall design relies on two fundamental assumptions: (i) in order to raise students' achievement, FA has to focus not only on cognitive, but also on metacognitive and affective factors; (ii) argumentation practices can be developed as crucial FA tools in the mathematics classroom. The teacher's role is, in particular, envisaged according to a cognitive apprenticeship perspective (Collins et al. 1989; see the theoretical framework section).

In this paper, we highlight how the specific choices concerning the use of technology, the design of the activities and the adopted methodology can foster the enactment of formative assessment strategies at the teacher's, the students', and the peers' levels. We ground our discussion on the analysis of a teaching-learning episode in an Italian grade 5 school. The analysis will be framed by a three-dimensional framework developed within the FaSMEd project. The framework highlights how digital technologies may support formative assessment processes carried out by teachers and students. In particular, we show how students are supported in "making their thinking visible" (Collins et al. 1989) through efficient formative assessment strategies.

6.2 Theoretical Framework

Within the FaSMEd Project, FA is conceived as a method of teaching where

> […] evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited. (Black and Wiliam 2009, p. 7)

In this method, three crucial processes are identified: establishing where learners are in their learning; establishing where learners are going; and establishing how to get there (Wiliam and Thompson 2007). Moreover, Wiliam and Thompson (2007) provide a model for FA in a classroom context as consisting of five key strategies: (A) Clarifying and sharing learning intentions and criteria for success; (B) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding; (C) Providing feedback that moves learners forward; (D) Activating students as instructional resources for one another; (E) Activating students as the owners of their own learning. Three main agents are involved in these processes: the teacher, the learners, and their peers.

Within FaSMEd, the model by Wiliam and Thompson (2007) has been extended to include the use of technology in FA processes. The result is a three-dimensional model taking into account three main dimensions (Aldon et al. 2017; Cusi et al. 2016): (1) the five FA key-strategies introduced by Wiliam and Thompson (2007); (2) the three main agents that intervene (the teacher, the student, the peers); and (3) the functionalities through which technology can support the three agents in developing the FA strategies (see the chart in Wright et al., in this volume).

The third dimension (functionalities of technology) is the new one integrated to the previous model. More specifically, we subdivide the functionalities of technology in three main categories, according to the different uses of technology for FA within our project:

(1) *Sending and displaying*, when technology is used for sending questions and answers, messages, files, or displaying and sharing students' worksheets or screens to the whole class.
(2) *Processing and analyzing*, when technology supports the processing and the analysis of the data collected during the lessons, such as the statistics of students' answers to polls or questionnaires, the feedback given directly by the technology to the students, the tracking of students' learning paths.
(3) *Providing an interactive environment*, when technology enables the creation of an interactive environment within which students can work individually or collaboratively on a task or a learning environment where mathematical/scientific content can be explored.

In FaSMEd, a main goal was studying the exploitation of new technologies in order to raise students' achievement. As stressed by the European Commission Report on low achievement in mathematics and science, "low achievement is associated with a range of factors that are not only cognitive in nature" (European Commission n.d, p. 27). In particular, research has highlighted the role played by metacognition (Schoenfeld 1992) in fostering students' achievement in mathematics. The importance of focusing on the metacognitive dimension, in particular when working with low-achievers, is stressed by Gersten et al. (2009, quoted in Scherer et al. 2016), who identify the selection and sequencing of instructional examples and the students' verbalization of their own strategies as fruitful components that support students who face difficulties in mathematics. However, research has also shown the impact of affective factors (McLeod 1992) in students' achievement, as documented, for example, by Fadlelmula et al., who declare that "if mathematics educators want to enhance students' mathematics achievement, they may need to consider motivational factors along with learning strategies, rather than considering each factor in isolation" (2015, p. 1372).

In our study, we planned and developed classroom activities with the aim of fostering students' development of ongoing reflections on the teaching-learning processes at both a metacognitive and affective level. More specifically, we planned activities with a strong focus on students' sharing of the thinking processes with the teacher and their classmates, supporting them in making thinking visible (Collins et al. 1989).

Cognitive apprenticeship theory proposes a model of instruction that incorporates some key aspects of the apprenticeship of practical professions. Collins et al. (1989), in particular, identified specific teaching methods that should be designed to give students the opportunity to observe, engage in, and invent or discover expert strategies in context:

(1) *modeling*, which requires an expert (the teacher or a classmate) to carry out a task externalizing his/her internal processes;
(2) *coaching*, that is the expert's observation of students while they are facing a task, in order to give them stimuli, supports, feedbacks;
(3) *scaffolding*, which refers to the support the expert gives to students in carrying out a task;
(4) *fading*, which refers to the gradual removal of the support to enable students to autonomously complete the task;
(5) *articulation*, which involves those methods aimed at getting students to articulate their knowledge, reasoning, or problem-solving processes to become able to consciously control their own strategies;
(6) *reflection*, aimed at making students compare their own problem-solving processes with those of an expert (the teacher or another student).

These categories can be subdivided into two groups: the first group (modelling, coaching, scaffolding, and fading) refers to the methods mainly aimed at fostering students' development of specific competencies through processes of observation and guided practice; the second group (articulation and reflection) includes methods aimed at making students reflect on experts' approaches and learn how to consciously control their own strategies.

Another crucial assumption concerns the central role of argumentation. Mathematical argumentation is conceived as "the discourse or rhetorical means (not necessarily mathematical) used by an individual or a group to convince others that a statement is true or false" (Stylianides et al. 2016, p. 316). The argumentation process encompasses ascertaining (when the individual removes his or her own doubts about the truth or falsity of a statement) and persuading (when the individual or the group remove the doubts of others about the truth or falsity of a statement) (Harel and Sowder 2007). The concept of argumentation was widely studied by scholars in the last decades, with an emphasis that can be traced back to two different kinds of sources. The first one relies on the possible links between argumentation and proving processes (Durand-Guerrier et al. 2012). The second source of interest for argumentation relies on the growing consensus on the importance of classroom discourse as a source of mathematical learning (Sfard 2001; Yackel 2004). Argumentation as a social activity was studied by many scholars, starting from Krummheuer (1995). More recently, some researchers have addressed the role of the teacher in managing argumentation in the classroom, discussing the double role of the teacher in giving arguments and in dealing with the arguments provided by the students, towards the creation of a "collective" argumentation (Conner et al. 2014). Although research did not explicitly address the possible connection between argumentation and formative assessment, in our research we argue that argumentation practices may support formative assessment processes.

6.3 Methodology and Research Question

In line with the perspective introduced in the previous section, we planned activities with a strong argumentative component and we searched for a technological tool that allows the students to share their productions, and the teacher to easily collect the students' opinions and reflections during or at the end of an activity. The activities were carried out in a connected classroom environment, in which students were using tablets connected with the teacher's laptop. Specifically, we chose a software (IDM-TClass) which enables the teacher to monitor students' work, to show (to one or more students) the teacher's screen and also the students' screens, to distribute documents to students and to collect documents from the students' tablets, and to create instant polls and immediately show their results to the teacher's laptop, which can be connected to a wider screen by means of a data projector or an interactive whiteboard.

The experimental phase involved 19 teachers from three different clusters of schools (primary and lower secondary schools) located in the North-West of Italy. In order to foster collaborative work and argumentation, students are asked to work in pairs or small groups (of three) on the same tablet.

The use of connected classroom technologies has been integrated within a set of activities on relations and functions, and their different representations (symbolic representation, tables, graphs). We adapted activities from the ArAl project (Cusi et al. 2011) and from The Mathematics Assessment Program, designed and developed by the MARS Shell Center team at the University of Nottingham (http://map.mathshell.org/materials/lessons.php). Specifically, we prepared a set of worksheets aimed at supporting the students in the verbalization and representation of the relations introduced within the lesson, enabling them to compare and discuss their answers, and making them reflect at both the cognitive and metacognitive levels. The worksheets, which have been designed to be sent to the students' tablets or to be displayed on the interactive whiteboard (IWB) or through the data projector, can be divided into four main categories:

(1) *Problem worksheets*, introducing a problem and asking one or more questions;
(2) *Helping worksheets*, aimed at supporting students, who have difficulties with the Problem worksheets;
(3) *Poll worksheets*, prompting a poll between proposed options;
(4) *Discussion worksheets*, prompting a focused discussion.

In this paper, we show how our specific choices concerning the use of technology and the design of the activities promoted the development of FA processes. More specifically, we address the following question:

> In what ways could focusing on "making thinking visible", through the support provided by connected classroom technologies, enable the activation of FA strategies?

During the design experiments, all lessons were video-recorded. Aligned with design-based research methodology (Cobb et al. 2003), one researcher (one of the authors) was always in the class as both an observer and a participant; she supported the teacher in the use of the technology and in the management of the class discussion. When we use this term *class discussion*, we refer to the idea of mathematical discussion developed by Bartolini Bussi: "Mathematical Discussion is a polyphony of articulated voices on a mathematical object (e.g., a concept, a problem, a procedure, a structure, an idea or a belief about mathematics), that is one of the motives of the teaching-learning activity … A form of mathematical discussion is the scientific debate that is introduced and orchestrated by the teacher on a common mathematical object in order to achieve a shared conclusion about the object that is debated upon (e.g., a solution of a problem)" (1996, pp. 16–17).

The video-recordings of the lessons have been examined to identify meaningful episodes to be transcribed, which were chosen as "selected aspects of the envisioned learning and of the means of supporting it as paradigm cases of a broader class of phenomena" (Cobb et al. 2003, p. 10). Other collected data were field notes from the observers, teachers' interviews after sets of lessons, questionnaires posed to students at the end of the activities, and interviews with groups of students.

In the following, we analyze an episode from a class discussion developed in primary school, referring both to the FaSMEd three-dimensional framework and to the cognitive apprenticeship methods.

6.4 Analysis of an Episode from a Classroom Discussion

The episode is taken from a classroom discussion carried out in grade 5. The discussion is focused on a problem worksheet called *Match the story* (Fig. 6.1), from a sequence of lessons on time-distance graphs, which we called *Tommaso's walk*. The students have previously worked on two sequences of lessons (about 16 hours) on functions and graphs, set within the context of early algebra. The lessons involve interpreting, comparing and discussing different representations (verbal, symbolic, graphical) of relations between variables and are adapted from

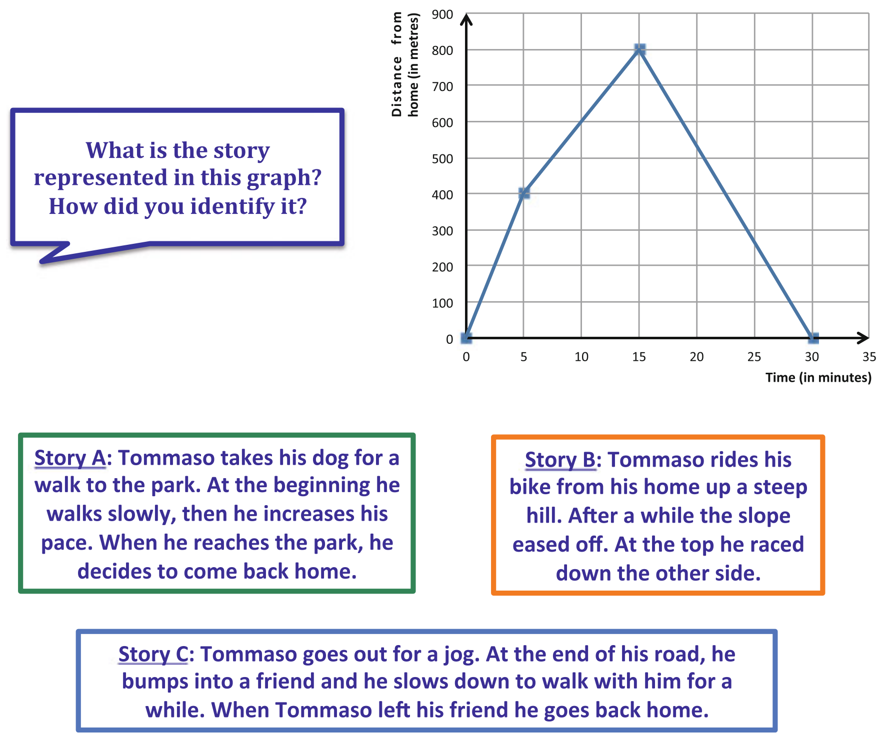


Fig. 6.1 The problem worksheet *Match the story*

two activities from the ArAl Project: "The archaeologist Giancarlo" and "La festa di Primavera." (All sequences of lessons can be found at https://microsites.ncl.ac.uk/fasmedtoolkit/category/partner/university-of-turin/.)

Tommaso's walk is our adaptation of the activity ***Interpreting distance-time graphs*** developed within the Mathematics Assessment Program. This activity was chosen to be adapted and implemented by all the partners involved in FaSMEd. In tune with the results of the TIMSS seven-nation comparative study (Hiebert et al. 2003), the common aim was to adopt approaches which preserve the complexity of concepts and methods, rather than simplifying them. Accordingly, the activities are designed and implemented with the aim of fostering the students' construction of meaning through formative assessment. Before facing the Tommaso's walk sequence, students are introduced to time-distance graphs by an experience with a motion sensor, which produced the distance-time graph of their movement along a straight line. Subsequently, they face the interpretation of a given time-distance graph according to a given story (referring to the walk of a boy, Tommaso) and focus, in particular on the meaning of ascending/descending lines and horizontal lines within a time-distance graph, and on the distinction between the concepts of "distance from home" and "distance that was walked through."

Match the story (Fig. 6.1) is the 6th worksheet of the sequence and is aimed at making students: (a) consolidate their competencies in the interpretation of a time-distance graph; (b) interpret the slope of the graph as an indication of the speed; (c) consolidate their competencies in recognizing complete justifications of given answers. The sequence then develops through matching between different graphs and the corresponding stories and finishes with the construction of graphs associated with specific stories. (The complete sequences of the designed worksheets can be found at https://microsites.ncl.ac.uk/fasmedtoolkit/2016/11/16/time-distance-graphs-idm-tclass/.)

After working on the interpretation of time-distance graphs for two lessons (about 4 hours), students are given the *Match the story* worksheet. The worksheet requires students to identify which story corresponds to a given graph among three proposed stories. To solve the task and identify the correct story (C), students have to interpret the slope of a line, within a time-distance graph, as an indicator of the speed. Stories B and A were designed as containing typical mistakes with this kind of task. Story B presents the typical mistake of interpreting time-distance graphs as drawings (in this case, the drawing of a hill). An interesting aspect is related to the main reason why this choice is not correct: story B implies that the distance from home should increase, while the last section of the graph represents a "return to home." Story A and story C are very similar. Identifying the correct story requires the student to observe that the graph represents, through the changing of the slope from the first to the second section, a decreasing of the speed.

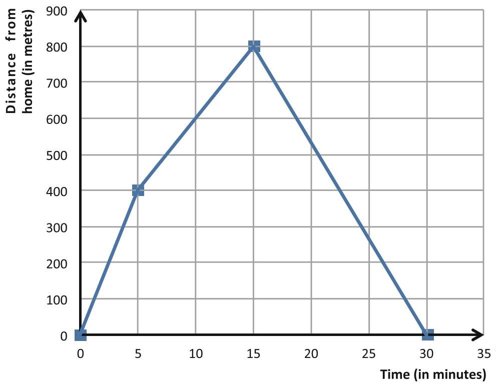
At the beginning of the lesson, the worksheet is sent from the teacher's laptop to the students' tablets. Students work in pairs or small groups of three to solve it. To support the students who face difficulties, two helping worksheets were designed. The first one (Fig. 6.2) provides students with a given table to collect the distances from home next to the corresponding times (0, 5, 15 min). This helping worksheet also proposes guiding questions to make students observe that the same distance (400 m) was walked in different periods of times, highlighting when Tommaso was quicker.

During the working group activity, the teacher can send the *helping worksheets* to the pairs/groups of students who face some difficulties. Receiving the *helping worksheets* represents therefore feedback for the students, because they become aware that their answer should be improved, and at the same time, they receive support to face the task. So, sending helping worksheets is a means to foster the activation of FA strategy C (*Providing feedback that moves learners forward*).

After facing the task and answering the questions, the pairs/groups send back their written productions to the teacher. When all groups send back their answers, the teacher groups the written answers and shows some of them on the IWB to set up a class discussion, activating FA strategy B (*Engineering effective classroom discussions that elicit evidence of student understanding*). Specifically, four answers are projected on the IWB.

The discussion starts focusing on Carlo and Elsa's answer, which is projected on the interactive whiteboard:

What is the story represented in this graph? How did you identify it?



Story A: Tommaso takes his dog for a walk to the park. At the beginning he walks slowly, then he increases his pace. When he reaches the park, he decides to come back home.

Story B: Tommaso rides his bike from his home up a steep hill. After a while the slope eased off. At the top he raced down the other side.

Story C: Tommaso goes out for a jog. At the end of his road, he bumps into a friend and he slows down to walk with him for a while. When Tommaso left his friend he goes back home.

Help to identify the story:

Collect, in the following table, the information provided by the graph:

Time	Distance from home
0 minutes	
5 minutes	
15 minutes	

Answer the following questions:

- What is the distance that Tommaso has walked through during the first 5 minutes?
- What is the distance that Tommaso has walked through in the period of time from 5 minutes to 15 minutes?
- Which is the period of time (among those analysed in the previous answers) during which Tommaso walks quickly?
- What is, therefore, the story represented in the graph?

Fig. 6.2 The helping worksheet for the *Match the story* task

In our opinion, (the story) B is not right because a sensor cannot measure the height. (The story) C is not correct because the graph tells that Tommaso initially walks slowly, then more rapidly; however, the story tells the contrary. The story A tells something that, probably, is possible.

After the teacher reads the text, Carlo and Elsa immediately declare that they realized they made a mistake. Carlo says that, however, the justification they gave to discard story B is right. The teacher and the researcher help the students notice that the argument they propose ("a sensor cannot measure the height") is not the correct one because a motion sensor could also be used to study the motion of a person walking on a hill, even if in the classroom this was not experienced. The discussion then focuses on the reasons why story B could not be accepted, and Sabrina asks to speak:

(347) Sabrina: Because [the story] B, practically, … I see a sort of drawing that looks like a hill…so I describe it as I see it and not…
(348) Researcher: So you are saying: "The story B…the graph resembles a hill… this fact could lead me to make a mistake".

Sabrina highlights that the reference to the hill in story B could make the students think that the graph represents the same hill that Tommaso is climbing, which is a typical mistake in this kind of activity. The researcher's strategy can be referred to the *articulation* and *reflection* categories of cognitive apprenticeship methods. In fact, by revoicing Sabrina's observation to make it more explicit for the other students, the researcher focuses the discussion on the possible mistake, thus providing students with tools to monitor their future work. In this way, Sabrina's contribution to the discussion is exploited as a resource for her classmates, and *FA strategy D* is activated.

The episode continues with the researcher challenging the students with the aim of making them focus on a fundamental part of story B, which assures them that the story cannot be associated with the graph:

(349) Researcher: There is also another reason why B is not right… Let's look at the graph for a while. Let's see if you can find it looking at the graph. Why is B not right?

Many pupils raise their hands.

(350) Researcher: A lot of hands have been raised. Who can start? Giacomo…
(351) Giacomo: The story C: "Tommaso went … When Tommaso left his friend, he walked back home". And you cannot find it over there (*pointing to story B written at the whiteboard*)…

Voices.

(352) Researcher: Wait (*speaking to the other students*). Maybe I understood what Giacomo wants to say. He says: here we can read "he walked home" (*pointing to this sentence in story C*). Here you can read "he goes back" (*indicating the sentence*

in story A). Here (*indicating story B*) you cannot find it. …Why is it not correct that "he goes back home" is not written in this story? (*speaking to Giacomo*)

Giacomo remains silent.

(353) Researcher (*speaking to Giacomo*): Why do you say that it is not correct that here we cannot find the sentence "he goes back home"?
(354) Giacomo: Because, over there, we can find that it [the line], then, goes down (*he indicates the graph on the IWB*).
(355) Researcher: You say: here, the graph is going down (*moving her finger along the descending part of the graph, from the point (15,800) to the point (30,0)*), it goes down toward the horizontal axis. What does it tell us?

Giacomo remains silent.

(356) Researcher: What is Tommaso doing?
(357) Giacomo: He is going back…
(358) Teacher: Good!

In this excerpt, it is possible to observe the activation of several categories of cognitive apprenticeship methods by the researcher. On one side, her interventions can be interpreted within the *articulation* and *reflection* categories, because she revoices two interventions by Giacomo to make them more explicit for the other students (lines 353 and 355):

line 351, when he stresses that, differently from the stories A and C, the story B does not include the fact that, at the end, Tommaso goes back home, and

line 354, when, asked to explain why it is so important that the story includes the fact that Tommaso goes back home (line 353), he focuses on the descending line of the graph.

At the same time, the researcher poses to Giacomo specific questions to support him not only in making explicit his reasoning, but also in refining and consolidating it, carrying out a *scaffolding* process through the questions: "Why is it not correct that "he goes back home" is not written in this story?" (line 352); "Why do you say that it is not correct that here we cannot find the sentence "he goes back home"?" (line 353); "What does it tell us?" (line 355); and "What is Tommaso doing"? (line 356).

The combination of *articulation* and *reflection* interventions and of *scaffolding* strategies also enables Giacomo to carry out the task posed by the researcher (lines 355–357). This observation puts the approach of the researcher in the *modeling* category of cognitive apprenticeship methods, that is, the researcher proposes Giacomo answer those questions that guide an expert in the solution of the task. In this way Giacomo becomes a model of reasoning for his classmates. For these reasons, this excerpt can be considered an example of an effective activation of *FA strategies D and E*. Giacomo, in fact, thanks to the support provided by the researcher, is activated as the owner of his learning (FA *strategy E*). At the same time, the interventions aimed at making Giacomo's ideas more explicit enable him to become an instructional resource for the other students (FA *strategy D*).

The discussion continues then as follows:

(359) Researcher: Let's listen to other observations.
(360) Teacher: Did you listen to what Giacomo said? …I don't know. Someone, in my opinion, was lost.
(361) Carlo: Can I explain it?
(362) Researcher: Carlo is going to explain what Giacomo said.
(363) Carlo (*speaking to his classmates*): Because Giacomo said that, in the answers A and C, these two stories explain that, at the end, … A tells that he goes back, C tells that he goes home … while C doesn't tell this thing. And, if we look at the graph, … the line …it goes down …it goes down at a certain moment. It approaches the horizontal axis, which is the home, it is right…but B doesn't specify it.
(364) Teacher: Instead of "It doesn't specify"…
(365) Researcher: Is it only that B doesn't specify this? It tells something that contradicts…

Livio, Adriana, Ambra raise their hands. Ambra is asked to speak.

(366) Ambra: It tells that …that it goes down to the other side. It [the graph] seems a hill, so it goes down to the other side. But …
(367) Noé: It is a graph, not a hill!
(368) Researcher: Noé says: "it is a graph, not a hill".
(369) Noé: Because…
(370) Researcher: Then, if Tommaso went down to the other side, …?
(371) Ambra: He wouldn't come…
(372) Arturo: He wouldn't be at home.
(373) Valeria: Yes! … and, in C, you can read "he goes back home".
(374) Researcher: He (*indicating Arturo*) says: "he wouldn't be at home".

In the subsequent part of the discussion the pupils are guided to observe that, if story B were the correct one, the last part of the graph should be an ascending line.

In this third excerpt, the teacher and the researcher act together in order to foster a sharing of the ideas expressed by Giacomo in the previous excerpt. The technique they adopt is to ask other students to revoice Giacomo's interventions (lines 360 and 362). This approach could be, again, located within the *articulation* and *reflection* categories of cognitive apprenticeship methods. The effectiveness of this approach is evident when Carlo asks to explain his classmate's observation (line 361), activating himself both as *owner of his learning* (strategy E) and as *an instructional resource for the other students* (strategy D).

The subsequent interventions by the teacher and the researcher aim at supporting Carlo and the other students in the correct interpretation of the graph and in the identification of the reasons why story B should be discarded. The interventions in lines 364, 365, 370 can be therefore referred to as the *modeling* and *scaffolding* categories of cognitive apprenticeship methods. Thanks to these interventions, other students take the responsibility of their own learning (*strategy E*), as Ambra (line 366), Noé (lines 367), Arturo's (line 372) and Valeria's (line 373) interventions testify.

6.5 Conclusions

In this paper, we argued that the *sending* and *displaying* functionality of connected classroom technologies and the focus on making students' thinking visible could foster the activation of FA strategies. As a first remark, we stress the role played by technology in supporting the different phases of this lesson and the subsequent activation of *FA strategies*: the worksheets are *sent* by the teacher to the students and vice versa (fostering, in case of helping worksheets, the activation of *FA strategy C*); then the students' written answers are *displayed* on the IWB, enabling the teacher to carry out a class discussion during which different feedback is provided (*FA strategy B and C*); and the students read carefully, discuss and compare their classmates' answers. In this way, students are activated both as owners of their learning (*FA strategy E*) and as resources for their peers (*FA strategy D*).

As a second remark, our analysis shows relevant interrelations between:

(a) the activation of *articulation* and *reflection* categories of cognitive apprenticeship methods and the activation of *FA strategies C* (*Providing feedback that moves learners forward*), *D* (*Activating students as instructional resources for one another*), *and E* (*Activating students as the owners of their own learning*);
(b) the activation of *modeling* and *scaffolding* categories of cognitive apprenticeship methods and the activation of *FA strategy E* (*Activating students as the owners of their own learning*), because students are guided to "act as experts" in facing the tasks. As a result, students' ideas are made more explicit, enabling them to become instructional resources for their classmates (*strategy D*).

Moreover, the combination of the different teacher's interventions delineates specific ways of managing class discussions, so they support an effective activation of *FA strategy B* (*Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding*).

The analysis we developed highlights also how argumentation, besides being part of the mathematical task at issue (students are required to justify their answers), turns into a means to enhance formative assessment. It is possible to distinguish two main argumentative moments that characterize the structure of the lessons we carry out during our design experiments: first when pairs/groups of students are asked to accompany their answers with an explanation of their plausibility; and second, a collective moment, when the class, under the guidance of the teacher, examines some selected group answers. When, during the collective argumentative moment, students explicitly state the reasons behind their answers, they are led to become owner of their learning (*FA strategy E*); when the classmates intervene and explain why the answer at issue doesn't hold and should be modified, they become resources for their classmates (*FA strategy D*); moreover, the teacher and classmates give feedback on the proposed argumentation and, in this way, clarify what are the learning objectives (*FA strategy A*), that is to say what are the relevant features an argumentation should have. The role of argumentation in formative assessment activities and the connection between argumentation and FA strategies is an issue that we are planning to study in depth.

We are also developing an analysis of teacher's interventions in relation to the goal of providing specific feedback to students. In particular, we are going to integrate the analysis presented in this paper with the analysis on the strategies of feedback to focus on the intentionality behind the teacher's interventions.

Acknowledgements The research leading to these results has received funding from the European Community's Seventh Framework Programme fp7/2007–2013 under grant agreement No [612337].

References

Aldon, G., Cusi, A., Morselli, F., Panero, M., & Sabena, C. (2017). Formative assessment and technology: Reflections developed through the collaboration between teachers and researchers. In G. Aldon, F. Hitt, L. Bazzini, & U. Gellert (Eds.), *Advances in mathematics education. Mathematics and technology: A CIEAEM source book* (pp. 551–578). Cham, Switzerland: Springer.

Bartolini Bussi, M. G. (1996). Mathematical discussion and perspective drawing in primary school. *Educational Studies in Mathematics, 31,* 11–41.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiment in educational research. *Educational Researcher, 32*(1), 9–13.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing and mathematics! In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conner, A., Singletary, L. M., Smith, R. C., Wagner, P. A., & Francisco, R. T. (2014). Teacher support for collective argumentation: A framework for examining how teachers support students' engagement in mathematical activities. *Educational Studies in Mathematics, 86*(3), 401–429.

Cusi, A., Malara, N. A., & Navarra, G. (2011). Early algebra: Theoretical issues and educational strategies for bringing the teachers to promote a linguistic and metacognitive approach to it. In J. Cai & E. J. Knuth (Eds.), *Early algebraization: Cognitive, curricular, and instructional perspectives* (pp. 483–510). Berlin, Germany: Springer.

Cusi, A., Morselli, F., & Sabena, C. (2016). Enhancing formative assessment strategies in mathematics through classroom connected technology. In C. Csíkos, A. Rausch, & J. Szitányi (Eds.), *Proceedings of PME 40* (Vol. 2, pp. 195–202). Szeged, Hungary: Psychology of Mathematics Education.

Durand-Guerrier, V., Boero, P., Douek, N., Epp, S. S., & Tanguay, D. (2012). Argumentation and proof in the mathematics classroom. In G. Hanna & M. de Villiers (Eds.), *Proof and proving in mathematics education* (New ICMI Study Series 15, pp. 349–367). New York, NY: Springer.

European Commission. (n. d.). *Addressing low achievement in mathematics and science.* Final report of the thematic working group on mathematics, science and technology (2010–2013). Retrieved from http://ec.europa.eu/education/policy/strategic-framework/archive/documents/wg-mst-finalreport_en.pdf.

Fadlelmula, F. K., Cakiroglu, E., & Sungur, S. (2015). Developing a structural model on the relationship among motivational beliefs, self-regulated learning strategies and achievement in mathematics. *International Journal of Science and Mathematics Education, 13,* 1355–1375.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, O., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242.

Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 805–842). Greenwich, CT: Information Age Publishing.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. Washington, DC: U.S. Department of Education National Center for Education Statistics.

Irving, K. I. (2006). The impact of educational technology on student achievement: Assessment of and for learning. *Science Educator, 15*(1), 13–20.

Krummheuer, G. (1995). The ethnography of argumentation. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 229–269). Hillsdale, NJ: Lawrence Erlbaum.

McLeod, D. (1992). Research on affect in mathematics education: A reconceptualization. In D. Grouws (Ed.), *Handbook of research on mathematics learning and teaching* (pp. 575–596). New York, NY: Macmillan.

Quellmalz, E. S., Timms, M. J., Buckley, B. C., Davenport, J., Loveland, M., & Silberglitt, M. D. (2012). 21st century dynamic assessment. In J. Clarke-Midura, M. Mayrath, & C. Dede (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 55–89). Charlotte, NC: Information Age Publishing.

Roschelle, J., & Pea, R. (2002). A walk on the WILD side. How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology, 1*(1), 145–168.

Roschelle, J., Penuel, W. R., & Abrahamson, L. (2004). The networked classroom. *Educational Leadership, 61*(5), 50–54.

Roschelle, J., Tatar, D., Chaudhury, S. R., Dimitriadis, Y., & Patton, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. *Computer, 40*(9), 42–48 (Published by the IEEE Computer Society).

Scherer, P., Beswick, K., DeBlois, L., Healy, L., & Moser Opitz, E. (2016). Assistance of students with mathematical learning difficulties: How can research support practice? *ZDM: International Journal on Mathematics Education, 48,* 633–649.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York, NY: Macmillan.

Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational Studies in Mathematics, 46,* 13–57.

Stylianides, A. J., Bieda, K. N., & Morselli, F. (2016). Proof and argumentation in mathematics education research. In A. Gutiérrez, G. C. Leder, & P. Boero (Eds.), *The second handbook of research on the psychology of mathematics education* (pp. 315–351). Rotterdam, The Netherlands: Sense Publishers.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.

Wright, D., Clark, J., & Tiplady, L. (this volume). Designing for formative assessment: A toolkit for teachers. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe* (ICME-13 Monographs) (pp. 207-228). Cham, Switzerland: Springer International Publishing AG.

Yackel, E. (2004). Theoretical perspectives for analysing explanation, justification and argumentation in mathematics classrooms. *Journal of the Korea Society of Mathematical Education Series D: Research in Mathematical Education, 8*(1), 1–18.

Author Biographies

Annalisa Cusi graduated in Mathematics at Modena and Reggio Emilia University in 2001, where she obtained a Ph.D. in Mathematics in 2009. She's been teaching mathematics and physics in upper secondary school since 2001. She worked as a research fellow at the University of Turin from 2014 to 2016 within the European Project FaSMEd. Her main research interests are innovation in the didactics of algebra; the analysis of teaching/learning processes, with a focus on the role played by the teacher; methods to promote early algebraic thinking in young students; teacher professional development; and formative assessment processes in mathematics.

Francesca Morselli graduated in Mathematics at the University of Genoa (2002) and obtained her Ph.D. in Mathematics at the University of Turin (2007). Since 2015, she is associate professor of mathematics education in the Department of Mathematics of the University of Genova (Italy), where she works in preservice and inservice teacher education programs. Her research focuses on argumentation and proof in mathematics; formative assessment in mathematics classrooms; and the interaction between affective and cognitive factors in the teaching and learning of mathematics.

Cristina Sabena is Associate Professor in the Department of Philosophy and Science Education of the University of Turin (Italy), where she teaches mathematics education to future primary teachers. She participated in the European Project FaSMEd (*Improving progress for lower achievers through Formative Assessment in Science and Mathematics Education*) on the theme of supporting formative assessment in mathematics through new technologies. Her research focuses mainly on the use of semiotics for studying gestures and multimodality in the development of theoretical thinking in mathematics classrooms, and the networking of different theoretical approaches in mathematics education.

Chapter 7
Supporting Online E-Assessment of Problem Solving: Resources and Constraints

Galit Nagari-Haddif and Michal Yerushalmy

Abstract Our research focuses on the e-assessment of challenging 'construction' e-tasks designed to function as a dynamic interactive environment of multiple linked representations (MLR); we explore the effect of constraints on the variation in the students' response space. Students are asked to determine whether an existential statement is correct. If they answer "yes," they construct an example in a MLR environment to support their answer; otherwise, they provide an explanation. The submitted example may be a sketch or an algebraic expression that can be checked automatically. Using a design-based research methodology, we describe a two-cycle study, focusing on one e-task on the topic of tangency to a function. Findings suggest that adding constraints to a logical mathematical statement enriches the variation of the response space and helps reveal different characteristics of students' thinking.

Keywords Automatic assessment · Examples · Response-space
Construction tasks · Calculus

7.1 Introduction

First-generation e-assessments were limited to multiple-choice questions, subsequently enhanced by short verbal or numeric answers (Scalise and Gifford 2006). Studies show that assessments based exclusively on questions of this type lead to limited learning and incorrect inferences about the purpose of the assessment, such as "there is only one right answer," "the right answer resides in the head of the teacher or test maker," and "the role of the student is to get the answer by guessing" (Bennett 1993). Black and Wiliam defined formative assessment as encompassing

G. Nagari-Haddif (✉)
Faculty of Education, University of Haifa, Haifa 31905, Israel
e-mail: gnagarih@campus.haifa.ac.il

M. Yerushalmy
University of Haifa, Haifa 31905, Israel

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_7

"all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning in which they are engaged" (1998, p. 7). Based on socio-constructivist philosophy, Mislevy (1994) described formative assessment as the gathering of evidence to support inferences about student learning.

In the present paper, our focus is on task design principles that help reveal different characteristics of students' thinking automatically. We explore ways to design tasks for formative e-assessment that expand students' response space and support our understanding of their knowledge. We focus on analysis of freehand sketches and symbolic expressions automatically, using the STEP (Seeing the Entire Picture) platform described in Olsher et al. (2016). The results of the analysis are used to characterize the response of the entire sample and to offer ways to characterize individual student's responses, to provide individual feedback to the student, and to improve teaching that responds to students' needs. We briefly discuss implications of our findings, related to teachers' instruction and students' learning. A detailed description of the way the automatic checking was performed is beyond the scope of this paper.

7.2 Theoretical Background

Multiple linked representations (MLRs) are useful with tasks that involve decision-making and other problem-solving skills, such as estimation, selecting a representation, and mapping changes across representations (e.g., Yerushalmy 2006). E-tasks involving MLRs also provide feedback to students, which reflect a process of inquiry while being assessed (reflecting feedback), and therefore have the potential to support the formative assessment of problem-solving processes, to catalyze ideas, and to provide an indication of students' perceptions and concept images, as described by Tall and Vinner (1981). Examples generated by participants require higher-level skills and may mirror students' conceptions of mathematical objects, their pedagogical repertoire, their difficulties, and possible inadequacies in their perceptions (Hazzan and Zazkis 1999; Sangwin 2003; Zazkis and Leikin 2007). Another use of examples is for determining the validity of mathematical statements. The framework by Buchbinder and Zaslavsky (2009) proved useful in constructing tasks that elicit logical connections between examples and statements, as well as in assessing this type of understanding.

In the present study, we focus on the construction of e-tasks in which the students are asked to determine whether an existential statement is correct. If they answer "yes," they construct an example in the MLR environment to support their answer. Otherwise, they provide an explanation. The example may be a sketch or an algebraic expression that generates a graph. The existential statement in this case concerns derivatives and specifically aims to assess the understanding of tangency to a function. The e-task assesses knowledge grounded in different areas of the high-school curriculum. Definitions of tangency appear in several topics throughout secondary-school math; they are related to Euclidean geometry (a tangent to a circle), analytic geometry, and calculus (Winicki-Landman and Leikin 2000). In geometry, students learn that the tangent line touches a circle in one point, often resulting in a limited concept image of a tangent in relation to a more general curve. Later, students are introduced to a formal or semi-formal definition of the tangent to the graph of a function. The global relations between figures that characterize tangency in Euclidean geometry are different from the local aspects that characterize tangency in analysis.

The global aspects of tangency to a function were identified by Biza et al. (2008), as one of three basic perspectives that students hold on tangency, and may be demonstrated by a concept image in which the tangent line has only one common point with the curve, and it leaves the curve in the same semi-plane. Other perspectives found in that study reflect the local aspects of tangency to a function, such as a tangent that intersects the curve at another point, or tangency at an inflection point. In these cases, the tangent does not leave the curve in the same semi-plane.

For the present study, we analyze the emerging response space: to distinguish between those who are influenced by the definition of the tangent to a circle, and therefore think that a tangent to a function in two points is not possible; those who think that a tangent to a function in two points is possible in certain limited cases and functions; and those who believe that tangency in two points may occur in various cases and functions. By asking students to construct and submit examples, we aimed to determine the richness of the students' concept image based on the attributes of the submitted functions (e.g., without extrema and discontinuities) and of the tangents (e.g., non-horizontal tangent lines). We can also determine the students' notions about the attributes of the mutual relations between functions and tangents (e.g., tangents that leave the curve in the same semi-plane, as identified by Biza and colleagues).

We explored the domain of response spaces constructed as answers to e-assessment problems in a recent work, and the present study is the outgrowth of our previous results (Nagari-Haddif and Yerushalmy 2015; Yerushalmy et al. 2017). In this study, we use a design-based research methodology that relies on iterative cycles of design. We describe a two-cycle study, focusing on one e-task. In the first cycle, we implemented the guiding design principles for a rich assessment task and categorized the submissions of the response space we obtained. Based on the results we refined the design of the original e-task, and conducted the second cycle. We report the data of the submissions of high school students (ages 16–17), all of whom learned the same curriculum with different teachers in different schools, without special emphasis on technology, although they were all conversant with the basic tools and technologies. Before the experiment in each cycle, students underwent a preparatory activity to become familiar with the STEP platform. The first cycle was conducted with 21 participants; the second one, 12 months later, with 86 participants, none of whom were part of the first cycle. During the experiments, each student solved 10 e-tasks in calculus, spending an average time of 10 min on a task.

7.3 First Cycle: Version A

As described in Yerushalmy et al. (2017), this cycle was designed to study what we can learn (in real time) about the students' knowledge by asking them to construct a supporting example in an MLR environment in answer to the existential claim: "There are functions that have one tangent line to their graph in two different points. Is this statement correct? If you answer 'yes,' submit a confirming example using a symbolic or graphic representation (a sketch). Otherwise, submit an explanation." The submissions created a response space that appears non-distinctive. Analyzing the 21 submitted answers, we found that all students answered that the statement was correct, and provided an example of a continuous function with at least two extrema. Eighteen students chose a function with extrema of equal *y*-coordinates, and constructed a horizontal tangent line through the extrema. Only three students submitted a non-horizontal tangent line, with the tangency points located near the extreme points. Seventeen students submitted a sketch, and four entered a symbolic expression. All symbolic expressions were of continuous trigonometric functions, with a horizontal tangent line. The homogenous response space did not provide in-depth information about the students' knowledge, and the task was not sufficiently interesting to serve as a formative assessment item. We did not expect this homogeneous response space, and found that the naïve examples represent functions with a "w" shape, or periodic ones with horizontal tangent lines at the extreme points, because in each case a tangent in one extreme point is tangent in the other extreme points (with the same *y* value). To obtain a less homogeneous response space, we conducted a second cycle of experiments, different from the first one in the design aspects of the task. In Yerushalmy et al. (2017) we report on the effect of one aspect, asking for three different examples rather than one, as a design principle that was found to help generate a heterogeneous response space. Here we analyze and report on another aspect of change: adding constraints to a logical mathematical statement. We created a sequence of three tasks based on the same existential statement; the original task (A) and two additional tasks (B1, B2), in which new mathematical constraints appear.

7.4 Cycle 2: Adding Constraints to the Logical Mathematical Statement (Versions B1, B2)

Although constraints are not always regarded as positive because they appear to limit freedom, creativity is released as a result of meeting constraints: increasing constraints extends awareness of what is possible (Watson and Mason 2005). Response space can be explored or extended when further constraints are included in the existential statement task to focus on particular characteristics of the examples (Zaslavsky and Zodik 2014). As constraints are added, learners may be directed toward slightly unfamiliar territory, helping them discover that there are

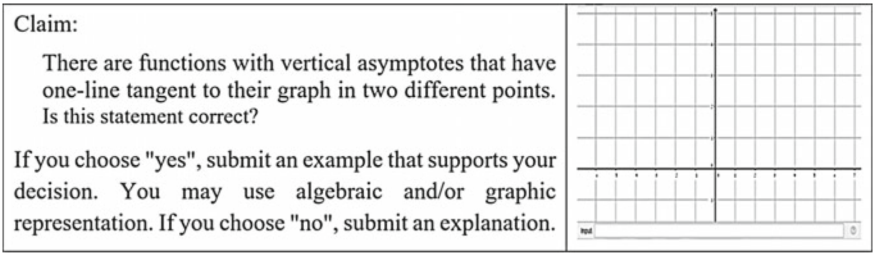


Fig. 7.1 Multiple points of tangency task: Version B1 (the function has a vertical asymptote)

possibilities beyond those that had come immediately to mind. Constraints are designed to support learners in thinking beyond their “habitual boundaries” (Watson and Mason 2005). Constraints that are manifest in the design of the tool affect students’ engagement with the guiding interactive diagram and contribute to making the task an interesting challenge (Naftaliev and Yerushalmy 2017). In light of these studies, we designed constraints to challenge the homogeneous response space found in the first cycle. We attempt to explore whether students’ concept image would include considerations of tangency to functions that have vertical asymptotes (Version B1, Fig. 7.1) and to those that have no extreme points (Version B2, Fig. 7.3).

7.4.1 Version B1: Tangent in Two Points to a Function with a Vertical Asymptote

We conducted the same type of analysis as described in Yerushalmy et al. (2017), therefore we report on the same categories of mistakes (Table 7.1) and characteristics of the submissions (Table 7.2).

Constraining the function to have a vertical asymptote (Fig. 7.1) resulted in 30.2% of mistaken submissions (Table 7.1): 60 students (69.8%) answered “Yes” and submitted a correct example that supports the claim; 10 students answered “No” and added an explanation, such as “A tangent line cannot intersect an asymptote,” “I couldn’t construct such an example,” “The tangent is not defined where the function is not defined,” “Because if for one value of x there are two values of y, this is not a function,” etc. These answers imply that some students’ image of tangency to a function in two points refers to continuous functions only.

The characteristics of the submissions were varied (Table 7.2): in 15 cases, tangency points were located near the extreme points and the tangent line was not horizontal. These cases appeared also in the first cycle, and raise an interpretation challenge: Was the sketch meant to represent a tangent at the extreme points, and therefore the tangent had to be parallel to the x-axis, or was it intended to be only approximately at the extreme points, making the submission correct? Twenty five

Table 7.1 Types of mistakes in the second cycle (Versions B1, B2)

Type of mistake		Number of students N = 86 (100%)	
		Version B1 (asymptote)	Version B2 (no extreme point)
Mistakes	Function that does not satisfy the constraints	4 (4.7%)	5 (6%)
	Incorrect symbolic expression	2 (2.3%)	2 (2.3%)
	Not submitted	4 (4.7%)	3 (3.5%)
	Not a function	3 (3.5%)	1 (1.2%)
	The line is not a tangent	3 (3.5%)	2 (2.3%)
	No tangent line added	0 (0%)	1 (1.2%)
	Incorrect answer ("no")	10 (11.6%)	19 (22.1%)
Correct answer		60 (69.8%)	53 (61.6%)
Total		86 (100%)	86 (100%)

students submitted a non-horizontal tangent line (e.g., Fig. 7.2c); 14 students submitted a function without extreme points (e.g., Fig. 7.2c, d); 23 students submitted a function constructed of a duplicated (mirrored) period with tangency points at the extrema (e.g., Fig. 7.2a); 27 students submitted a variety of relationships at different points: intersection (e.g., Fig. 7.2c), tangency that does not leave the curve in the same semi-plane (e.g., Fig. 7.2b–d). Three students submitted tangency at an inflection point (e.g., Fig. 7.2c, a); 16 students submitted a tangent line at the edges of the definition domain (e.g., Fig. 7.2d). Four students submitted symbolic expressions that belong to different families of functions, such as $f(x) = \frac{1}{\sin(x)}$ (Fig. 7.2b). We can assume that the student created this function, which has a vertical asymptote, based on the fact that $y = \sin(x)$ is a continuous function that has a tangent line in more than one point. Similarly, function $f(x) = \frac{x^2+2x+4}{\sqrt{x^2-4}}$ was constructed based on the notion that in some functions the vertical asymptotes reset the denominator to zero.

7.4.2 *Version B2: Tangent in Two Points to a Function Without Extreme Points*

When we added to the function the constraint of the functions not having extreme points (Fig. 7.3), we received 38.4% mistaken submissions (Table 7.1): 53 students (61.6%) answered "Yes" and submitted a correct example that supports their claim; 19 students answered "No" and added an explanation, such as "For two points of tangency we need two points with the same value, and it is possible only if the function has extreme points," "A function will never have the same *y* value unless there is an extreme point to return it back to the same *y* value," etc. The answers suggest that some students' image of a tangency to a function in two points refers to

Table 7.2 Characteristics of objects submitted for all versions (A, B1, B2)

Object	Attributes		Version A (first cycle)	Version B1 (asymptote)	Version B2 (no extreme)
Function	Continuous, periodic function with infinite number of tangency points		✓	✓	✓
	Function constructed of a duplicated (mirrored) period with tangency points		✓	✓	✓
	Symbolic representation		✓	✓	✓
	Discontinuous function (with a vertical asymptote)		×	✓	✓
	A function without extreme points		×	✓	✓
	Linear function (coincides with tangent)		×	×	✓
	Possibly incorrect representations of functions	Function sketch includes non-univalent regions	×	✓	✓
Tangent	Variety of relationships at different points: intersection, tangency (does not leave the curve in the same semi-plane)		✓	✓	✓
	Non-horizontal tangent line		✓	✓	✓
	Possibly incorrect representations of tangents	Non-horizontal tangent close to extreme points	✓	✓	✓
		Added line is not tangent to the function in 2 points	✓	✓	✓
	Tangent and function intersect at tangency point (inflection point of tangency)		×	✓	✓
	Tangent line at the edges of the definition domain		×	✓	✓
	Total number of relevant attributes		7	12	13

(a) **(b)** **(c)** **(d)**

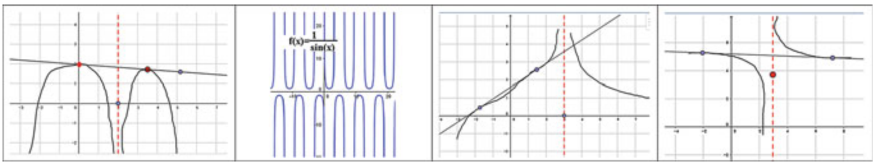


Fig. 7.2 Sample of correct submissions (second cycle, Version B1: The function has a vertical asymptote)

Claim:

There are functions without extremum points that have one line tangent to their graph in two different points. Is this statement correct?

If you choose "yes", submit an example that supports your decision. You may use algebraic and/or graphic representation. If you choose "no", submit an explanation.

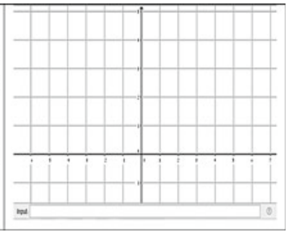


Fig. 7.3 Multiple points of tangency task: Version B2 (Cycle 2: Function without extrema)

horizontal tangents in extreme points only. The characteristics of the submissions varied (Table 7.2): 52 students submitted a non-horizontal tangent line (e.g., Fig. 7.4a–c); 10 students submitted a function with a vertical asymptote (e.g., Fig. 7.4d).

Eight students submitted a straight line as the function (e.g., Fig. 7.4c): this is a case in which the tangent and the function converge, and each point of the function is a tangency point (one of the students, however, answered “No” and explained: “A function without extreme points is a straight line and therefore the tangent line cannot intersect it in two different points,” suggesting that some students’ image of a function without extreme points is a linear function); 5 students submitted a tangent at an inflection point (e.g., Fig. 7.4d).

Four students submitted a symbolic expression that belongs to different families of functions, such as $f(x) = -\cos(x) + 2x$ (Fig. 7.4a) and $f(x) = x^5 + \sqrt[3]{x}$; 9 students submitted a tangent line at the edges of the definition domain (e.g., Fig. 7.4b); 20 students submitted a variety of relationships at different points: a tangent that also intersects the function, and tangency that does not leave the curve in the same semi-plane (e.g., Fig. 7.4d).

The refined design resulting from the constraints imposed in Versions B1 and B2 produced a heterogeneous response space that shed light on the characteristics of the different ways of students’ thinking (Table 7.2). The identified mistakes, (Table 7.1), were confirmed by the submitted verbal explanations. Our first conjecture was confirmed: adding constraints to a logical mathematical statement enriches the response space. Each of the two constraints on the type of the required function (Versions B1 and B2) produced a heterogeneous response space that became a means for distinguishing among the students’ conceptions. Table 7.2 shows that for Versions B1 and B2, the response space of the submissions consisted of 12 and 13 characteristics (out of 13 possible) respectively, whereas for Version A, it consisted of only seven characteristics (out of 13 possible).

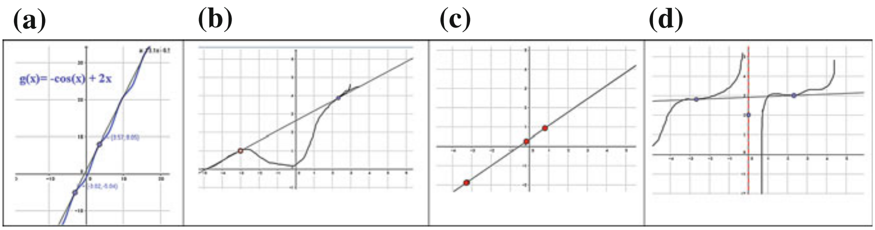


Fig. 7.4 Sample of correct submissions (Cycle 2, Version B2: Function without extreme points)

7.4.3 Analyzing Personal Response Spaces and Concept Images

When invited to construct their own examples, learners both extend and enrich their personal response space, but also reveal something regarding the sophistication of their awareness of the concept or technique (Liz et al. 2006). Watson and Mason (2005) regarded the notion of a personal response space as a tool for helping learners and teachers become more aware of the potential and limitations of experience with examples. If experiencing extensions of one's response space is sensitively guided, it contributes to flexibility in thinking and makes possible the appreciation and adoption of new concepts (Zaslavsky and Zodik 2014).

In the second cycle, we examined two versions of the existential statement, B1 and B2, concerning the same topic (tangency to a function at two points), automatically creating for each student a personal example that consists of a pair of submissions. This enabled us to distinguish between students' conceptions and diagnose more convincingly the aspects of their personal knowledge. For example, we were able to characterize students s70 (Fig. 7.5) and s84 (Fig. 7.6), whose pair of examples differ in the number of characteristics and mistakes (Table 7.3), and diagnose each one separately. For Version B1, student s70 (Fig. 7.5) submitted a function with a horizontal tangent at extreme points, with the same *y* value.

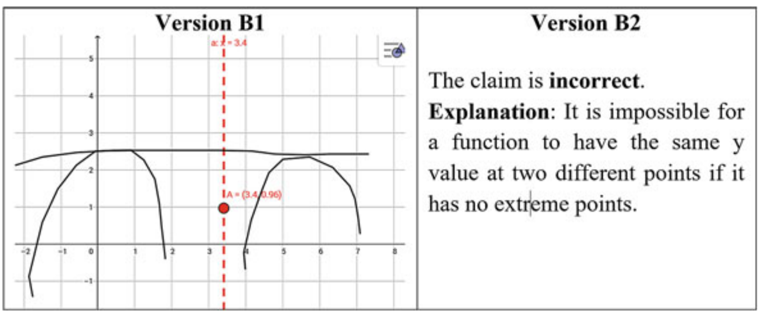


Fig. 7.5 Personal response space of student s70

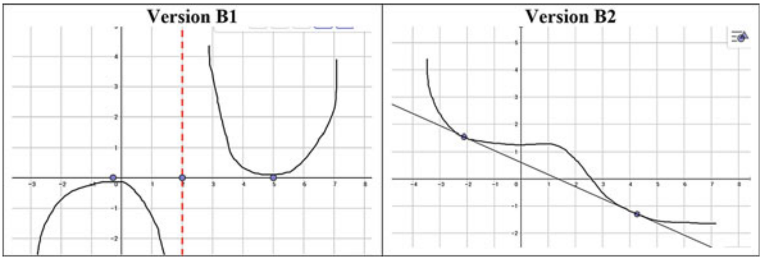


Fig. 7.6 Personal response space of student s84

Table 7.3 Characteristics of answers submitted by students s70 and s84, integrating Versions B1, B2

Object	Attributes	s70	s84
Function	Function constructed of a duplicated (mirrored) period with tangency points at extrema	✓	×
	A function without extreme points	×	✓
Tangent	Variety of relationships at different points: intersection, tangency (does not leave the curve in the same semi-plane)	×	✓
	Non-horizontal tangent line	×	✓

We assume that student s70 thinks that tangency at two points is possible only at extreme points with the same *y* value.

The negative response of student s70 to the e-task in Version B2 (regarding a function without extreme points) verifies this assumption: "It is impossible for a function to have the same *y* value at two different points if it has no extreme points." Student s70's answer is incorrect for several reasons: (a) the student marked as "incorrect" the originally correct claim: it is possible for a function without extreme points to have a tangent line at two different points; (b) it is possible for a function without extreme points to have the same *y* value at two different points, for example, $f(x) = \tan(x)$; and (c) tangency at two points is possible when the *y* values of the points are not equal (as demonstrated in Fig. 7.4a, b).

For Version B1, student s84 (Fig. 7.6) submitted a function with a horizontal tangent at extreme points, with the same *y* value. Based on this result, we infer that the student assumes that the tangent to a function at two points must be a horizontal line, and that tangency points must be extreme points. But the student's correct submission for Version B2 (Fig. 7.6), in which the function has no extreme points and the tangent line is non-horizontal, contradicts this assumption. The two functions that the student submitted (Fig. 7.6) suggest that the student's image of tangency to a function in two different points consists of functions that need not have extreme points, may have asymptotes, and the tangent line may be non-horizontal.

7.5 Summary and Discussion

Construction tasks appear to be an appropriate type of rich e-task for the validation of correct existential statements, because they have multiple correct solutions that can be automatically checked and analyzed. But we saw that the tasks submitted in the first cycle formed a narrow and homogenous response space, which may not provide adequate information about the students' knowledge beyond immediate perceptions. These tasks did not make it possible for us to characterize the variation between answers. In other words, asking for an example that supports a mathematical statement does not guarantee that the resulting response space will be rich, even if there are an infinite number of possible examples.

Constraining the type of function forced students to move out of their comfort zone. Given a choice of multiple representations when submitting their answers (Figs. 7.1 and 7.3), most students chose to submit a sketch, in both cycles. The small number of symbolic submissions (most of them were of trigonometric functions) indicates that students find sketching to be more intuitive, as described in Yerushalmy et al. (2017) and an alternative to uncontrolled trial-and-error behavior (Nagari-Haddif and Yerushalmy 2015).

The results show the delicate balance between giving students rich resources to freely construct examples, without any constraints, and asking them to construct strictly constrained examples. Both options may fail to encourage the development of a rich response space. The former might lead to a narrow response space, despite an infinite number of possible examples, as we found in the first cycle; the latter option leaves no room for imagination, personal creativity, and expressiveness because there is only one correct answer. Although both options could be problematic, we demonstrate a successful attempt to design online assessment interactive tasks, appropriately constrained, based on previous students' response space, which on one hand does not limit students to a single possible answer and on the other hand helps assess the students' knowledge of the concept at hand.

Automatic analysis of student submissions can provide feedback to teachers and students, helping them understand the students' conceptions of mathematical objects and of the mutual relationships between them. This analysis can affect future instruction, both in the classroom and online: teachers may receive online feedback about the submission characteristics of the entire class (e.g., Table 7.2) or of individuals (e.g., Table 7.3). According to the feedback they receive, teachers may conclude that they need to strengthen some properties of particular concepts or ideas, either in the classroom or with respect to an individual student. Similarly, students may receive detailed individual feedback, based on the characteristics of their submission. This, together with the given MLR, may encourage them to reflect on their actions during problem solving and following submission.

Acknowledgements This research was supported by the Israel Science Foundation (grant 522/13).

References

Bennett, R. (1993). On the meaning of construct. In R. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1–27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Biza, I., Christou, C., & Zachariades, T. (2008). Student perspectives on the relationship between a curve and its tangent in the transition from Euclidean Geometry to Analysis. *Research in Mathematics Education, 10*(1), 53–70.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Buchbinder, O., & Zaslavsky, O. (2009). A framework for understanding the status of examples in establishing the validity of mathematical statements. In *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 225–232). Thessaloniki, Greece: Aristotle University of Thessaloniki.

Hazzan, O., & Zazkis, R. (1999). A perspective on "give an example" tasks as opportunities to construct links among mathematical concepts. *Focus on Learning Problems in Mathematics, 21*(4), 1–14.

Liz, B., Dreyfus, T., Mason, J., Tsamir, P., Watson, A., & Zaslavsky, O. (2006, July). Exemplification in mathematics education. In *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 126–154). Prague, Czech Republic: Charles University, Faculty of Education.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*(4), 439–483.

Naftaliev, E., & Yerushalmy, M. (2017). Engagement with interactive diagrams: The role played by resources and constraints. In A. Leung & A. Baccaglini-Frank (Eds.), *Digital technologies in designing mathematics education tasks: Potential and pitfalls* (pp. 153–173). Switzerland: Springer International Publishing.

Nagari-Haddif, G., & Yerushalmy, M. (2015). Digital interactive assessment in mathematics: The case of construction E-tasks. In *Proceedings of the 9th Congress of the European Society for Research in Mathematics Education* (pp. 2501–2508). Prague, Czech Republic: Charles University, Faculty of Education and ERME.

Olsher, S., Yerushalmy, M., & Chazan, D. (2016). How might the use of technology in formative assessment support changes in mathematics teaching? *For the Learning of Mathematics, 36*(3), 11–18.

Sangwin, C. J. (2003). New opportunities for encouraging higher level mathematical learning by creative use of emerging computer aided assessment. *International Journal of Mathematical Education in Science and Technology, 34*(6), 813–829.

Scalise, K., & Gifford, B. (2006). A framework for constructing "Intermediate Constraint" questions and tasks for technology platforms computer-based assessment in E-learning. *The Journal of Technology, Learning, and Assessment*, *4*(6). https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1653. Accessed 20 June 2017.

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics, 12*(2), 151–169.

Watson, A., & Mason, J. (2005). *Mathematics as a constructive activity. Learners generating examples*. Mahwah, NJ: Lawrence Erlbaum Associates.

Winicki-Landman, G., & Leikin, R. (2000). On equivalent and non-equivalent definitions: Part 1. *For the Learning of Mathematics, 20*(1), 17–21.

Yerushalmy, M. (2006). Slower algebra students meet faster tools: Solving algebra word problems with graphing software. *Journal for Research in Mathematics Education, 37*(5), 356–387.

Yerushalmy, M., Nagari-Haddif, G., & Olsher, S. (2017). Design of tasks for online assessment that supports understanding of students' conceptions. *ZDM: International Journal on Mathematics Education, 49* (5), 701–716. https://doi.org/10.1007/s11858-017-0871-7

Zaslavsky, O., & Zodik, I. (2014). Example-generation as indicator and catalyst of mathematical and pedagogical understandings. In Y. Li, A. E. Silver, & S. Li (Eds.), *Transforming mathematics instruction: Multiple approaches and practices* (pp. 525–546). Cham, Switzerland: Springer.

Zazkis, R., & Leikin, R. (2007). Generating examples: From pedagogical tool to a research tool. *For the Learning of Mathematics, 27*(2), 15–21.

Author Biographies

Galit Nagari-Haddif is a Ph.D. student under the supervision of Prof. Michal Yerushalmy in the Mathematics Education department at the University of Haifa, Israel. Her study is supported by the ISF and Trump foundation and is about digital interactive assessment in calculus for high school students.

Michal Yerushalmy is a professor in the department of Mathematics Education at the University of Haifa, Israel. She studies mathematical learning and teaching, focused on the design and large-scale implementation of reformed curricula and on processes involved in learning with multiple external representations, bodily interactions, and modeling. She has authored and designed numerous software packages and interactive textbooks (The Visual-Math project) and has designed ways to make technology accessible for mathematical inquiry learning everywhere using mobile phones (The Math4Mobile project); she has also worked to support the teaching of guided inquiry curriculum (The STEP project).

Chapter 8
Suggestion of an E-proof Environment in Mathematics Education

Melanie Platz, Miriam Krieger, Engelbert Niehaus and Kathrin Winter

Abstract This paper deals with electronic proofs (e-proofs) implemented with existing open-source web applications (i.e., IMathAS) to be used in school and academic teacher education in mathematics. In these e-proofs, the learner will be supported and guided to prove a given theorem. An educational e-proof environment is a preliminary step towards a complete proof with paper and pencil in order to support students in building arguments and overcoming difficulties related to a traditional approach. IMathAS is a web-based mathematics assessment tool for delivery and semi-automatic grading of mathematical homework. In this paper, a theoretical framework to identify preferences of learners for logical reasoning is presented. This framework determines the structure of the IT-environment. Therefore, a brief introduction to the system is given; benefits, requirements, and constraints as well as further perspectives are discussed.

Keywords E-proofs · IMathAS · Proof validation · Distractors

M. Platz (✉)
Institute for Mathematics Education, University of Siegen,
Herrengarten 3, 57072 Siegen, Germany
e-mail: platz@uni-landau.de

E. Niehaus
Institute of Mathematics, University of Koblenz-Landau, Fortstr. 7,
76829 Landau, Germany
e-mail: niehaus@uni-landau.de

M. Krieger · K. Winter
Institute for Mathematical, Natural Sciences and Technology Education,
University of Flensburg, Auf dem Campus 1b, 24943 Flensburg, Germany
e-mail: miriamkrieger@wwu.de

K. Winter
e-mail: kathrin.winter@uni-flensburg.de

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_8

8.1 Introduction

> Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two facilities, which we may call intuition and ingenuity. The exercise of ingenuity in mathematics consists in aiding the intuition through suitable arrangements of propositions, and perhaps geometrical figures or drawings. (Alan Turing, 1912–1954) (cf. Ladewig 2016, p. 329)

In the process of proving, an "assertion will be deducted step by step in a formal deduction from theorems and definitions which are deemed to be known" (Meyer 2007, p. 21). In mathematics education, the teaching of reasoning skills plays a major role. These skills form the basis for proofs of mathematical statements and contexts dealt with in higher school grades in different contexts. It is especially logical reasoning and argumentation that support the formation of understanding. This is particularly valid in an everyday social context, where argumentation assists in the imparting of orientation competency, decision finding, and decision justification, peaceful conflict solution and the basic principles of democracy. In our case —science—the most important aim is to produce truths based on argumentations (cf. Budke and Meyer 2015). Furthermore, most research on mathematical education emphasizes the importance of mathematical reasoning as an integral part of doing mathematics (cf. Kramarski and Mevarech 2003).

Although digital instructional media are discussed controversially, they more than ever have a growing impact on learners. In 1998, Pavlik mentioned that "Today's children and youth are the heaviest users of new media technology" (Pavlik 1998, p. 394). This citation is still valid as verified by several studies like the KIM-studies in Germany (MFS 2000–2015). Cataloglu (2007) developed an internet-mediated assessment portal as a pedagogical learning tool and discovered that the majority of students were excited to use the computers and the computer supported instructions seemed to have a positive influence on their motivation. In conclusion, the students spent much more time solving problems in and out of class. Additionally, the added value of the use of multimedia educational material has been empirically proven repeatedly (cf. Handke and Schäfer 2012). In Alcock and Wilkinson (2011), e-proofs were designed to address the problem of teaching of proof comprehension in large, teacher-centered undergraduate mathematics lectures. Theoretical proof comprehension issues were supposed to be addressed within the practical context of traditional lectures by making the structure and reasoning used in a proof more explicit without cluttering its presentation.

> Each e-proof consists of a sequence of screens. Each screen shows the theorem and the whole proof, with much of the latter "greyed out" to focus attention on particular lines. Relationships are highlighted using boxes and arrows, and each screen is accompanied by an audio file which students can listen to as many times as they wish. (Alcock and Wilkinson 2011, p. 9)

The e-proof used in Alcock and Wilkinson (2011) differs from the e-proof used in the approach described in this paper. Those presented in Alcock and Wilkinson “are interactive, but only in a weak sense that the student controls the pace and sequence of the content” (2011, p. 13). As this method did not lead to an improvement of proof-comprehension compared to a proof presented in a lecture, they will explore “the possibility of allowing the students to construct their own e-proofs for submission as a part of an assignment in a course on communicating mathematics” (Alcock and Wilkinson 2011, p. 14) in future research. This kind of e-proof is closer to our approach of an e-proof-system. However, IMathAS (http://imathas.com), the system described in this article, may also include keeping both distractors to reduce the guessing probability (cf. Winter 2011), proof puzzles (cf. Ensley and Crawley 2006), and direct graphs of connected arguments (cf. Niehaus and Faas 2013) allowing for flexibility in conceptual argumentation. The research philosophical stance in the study is a mixture of pragmatism and realism. The research approach is inductive and mixed methods are used in a longitudinal time horizon.

8.2 Objectives

The overall objective is the development of an adaptive learning, test and diagnostic system, including the substantial and methodological potential of fuzzy logic and artificial neural networks to include a tutorial component for learning, on the one hand and, on the other hand, to improve automatic correction and to enable automatic valuation with possibly low manual correction efforts. Within the proposed E-proof system, the solutions of an E-proof task given by students are checked against the correct solution which can include different pathways from the preconditions to the conclusions. Wrong justifications that the learner chose or justifications the learner forgot will be used for analyzing the proof competencies of the learner and for the evaluation of distractors for the E-proof system. Further distractors, which are based on typical errors in argumentation, are determined through a didactical analysis of the proof competencies of a learner on a blank piece of paper (cf. Platz et al. 2017). Distractors serve as an analysis tool of argumentation competencies and for the future development of a support-system for the improvement of argumentation competencies. The correction effort shall be kept as low as possible while preferably having high degrees of freedom, which means that the learner is not forced to present one certain solution determined by the teacher, but different pathways to prove a theorem are esteemed and are supposed to be automatically evaluated through the system. In this paper, a theoretical framework to determine the preferences of learners for logical reasoning is presented. The theoretical framework determines the structure of the IT environment. IMathAS is presented and the benefits, requirements and constraints of using it in the form of an E-Proof-Environment are discussed.

8.3 The IMathAS E-proof Environment

IMathAS is an open-source web-based Internet Mathematics Assessment System which can be used within a web browser. A gradebook included into IMathAS allows automatic grading of mathematical homework, tests and electronic assessments. The questions are algorithmically generated, and numerical and mathematical expressions can be generated by the computer. Furthermore, free text and essay environments can be included with manual grading by the teachers. IMathAS allows accurate display of mathematics and graphs. A randomizer-function enables the creation of individual questions for all students. Questions created with the randomizer-function are structurally equivalent to each other. One advantage is that the results cannot be plagiarized, as the students have to solve the tasks on their own. The biggest benefit for the teachers and authors is a shared joint repository of questions and tasks. The questions can be included in a public library accessible for all teachers/tutors in the community. It is possible to modify and improve questions from other authors and integrate them into one's own private library. In this vein, the philosophy of collaboration and sharing among teachers and lectures of multiple educational institutions shall be supported. Alternatively, it is also possible to keep the questions in private libraries.

Most educational facilities have financial constraints. Therefore, the requirement was derived that a web-based learning environment should be provided as open-source in order to enhance free access to the learning of mathematics. Moreover, the cost of development for an e-proof system is minimized if just the e-proof system is realized as a kind of plug-in in an existing open-source solution. The aim to provide an e-proof learning environment within an existing open-source solution has additional advantages, because new releases of the underlying mathematical assessment system will be available to the e-proof plug-in as well. IMathAS was selected because it is open source and it provides a shared joint repository to a potentially large target group (universities, schools) in Rhineland-Palatinate (Germany), but beyond that it can be used on a transnational scale as well.

8.4 E-proofs in IMathAS

8.4.1 The Structure of an E-proof

Meeting the different requirements and functions of proofs, such as verification or discovery (cf. Bell 1976; de Villiers 1990), we classify the prototype for an e-proof environment as an interpolation aid between understanding of a given proof of a certain theorem (e.g., in a textbook) and the creation of one's own proof for the same given theorem on a blank piece of paper. In a first step, the proof can be comprehended gradually on a PC (cf. Alcock and Wilkinson 2011); following this, proof fragments shall be ordered (proof puzzle, cf. Ensley and Crawley 2006) and finally false proof-fragments can be added that can be determined via the diagnosis of typical mistakes of students (cf. among others Winter 2011). In this way, the degree of freedom will be increased, but also the correction effort. The aim is to keep the correction effort as low as possible while having preferably high degrees of freedom (cf. Platz and Niehaus 2015).

Consequently, an e-proof is not as flexible as a proof in a paper-and-pencil approach. Within an e-proof system, the solutions of an e-proof task given by students are checked against the correct solution, which can include different pathways from the preconditions to the conclusions. The considered proofs for the e-proof environment can be decomposed into single fragments. Each fragment consists of three components (see Fig. 8.1), namely

1. the connection with the previous fragment (resp. proof step)
2. the description of the proof step itself and
3. one or more justifications on the proof step.

In Fig. 8.2, the structure of an e-proof is illustrated by three possible solution graphs.

Additionally, it is possible to give the students the opportunity to define their own proof steps with an integrated editor in IMathAS and to use these proof steps in the e-proof environment. This method is closer to creating a proof on a blank piece of paper. Furthermore, pre-defined and semi-defined proof steps can be combined in the e-proof environment. (See Fig. 8.3 Students' view of the e-proof-environment in IMathAS.) The consequence of allowing self-defined proof steps is that, at present, the teacher needs to correct the student's electronic solution manually.

8.4.2 *Valuation of an E-proof*

One problem of the automatic valuation of e-proofs is that a certain proof fragment is expected in a certain position within a proof. To improve this situation, the correct solution is represented as a graph and, therefore, shortest path algorithms existing in graph theory (cf. e.g., Goldberg and Harrelson 2005) are helpful for the assessment of proofs, although longer paths are not necessarily wrong. The problem can be understood as a pure task of graph theoretical arrangement of proof fragments and justifications as attributes of edges in the solution graph in which a subset of proof-fragments and links between them are selected to form a logical sequence for the theorem. This sequence starts from an initial state with the prevailing conditions, a proof gap with a goal state is defined by the conclusion in the theorem (interpolation proof). The edge in the graph defines links between pairs of fragments.

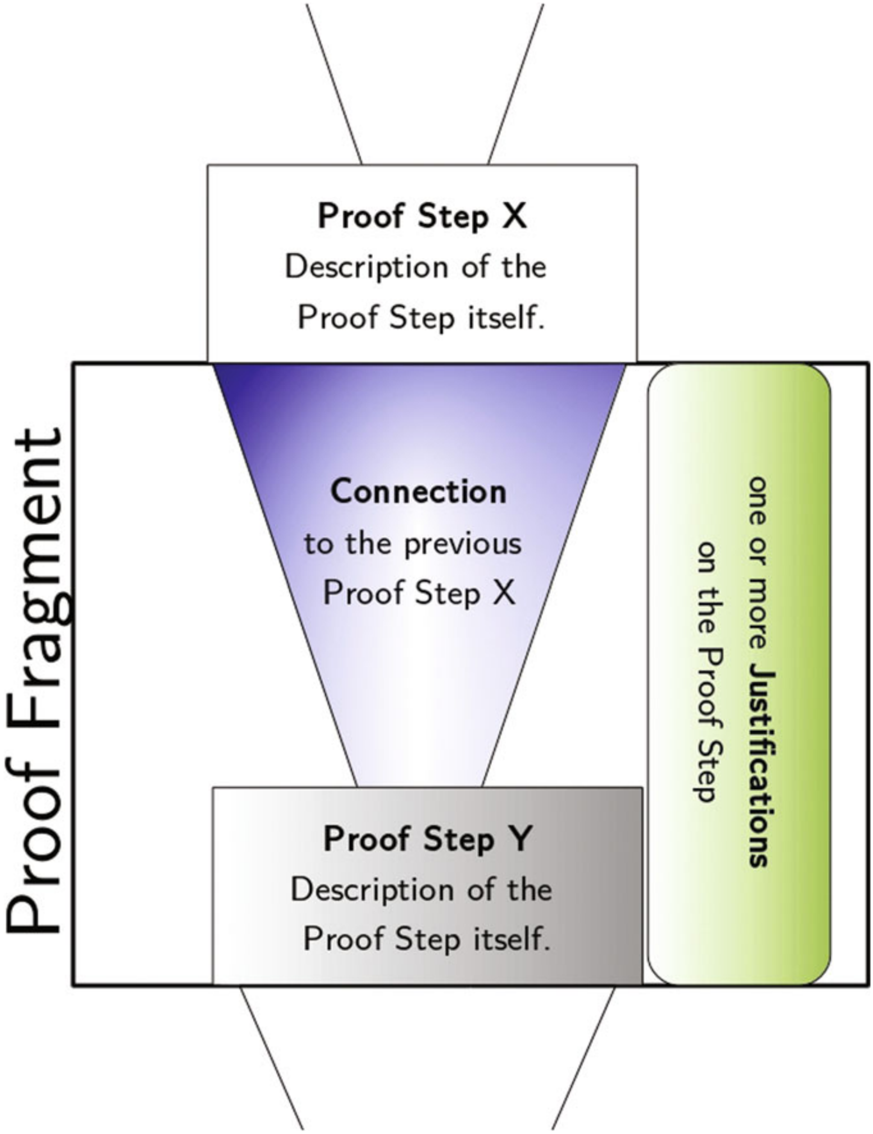


Fig. 8.1 Visualization of a proof fragment

8.4.3 Formative Assessment

Beside the summative assessment that can be performed with the e-proof system, formative assessment can be implemented into teaching with the system. The progress of a learner during the learning process can be determined with the aim to

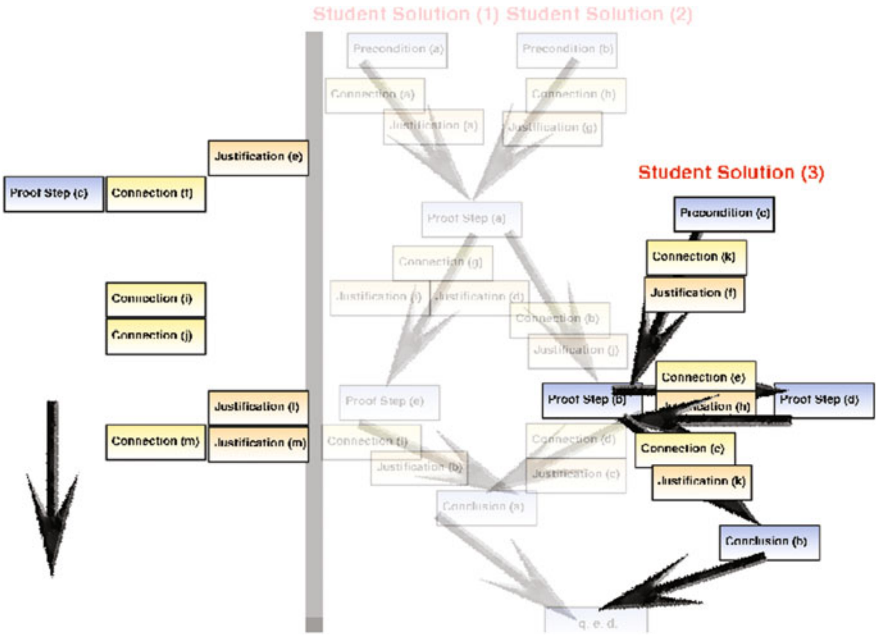


Fig. 8.2 The structure of an e-proof illustrated by three possible solution graphs. The associate applet is available at http://e-proof.weebly.com/applet-the-structure-of-an-e-proof.html

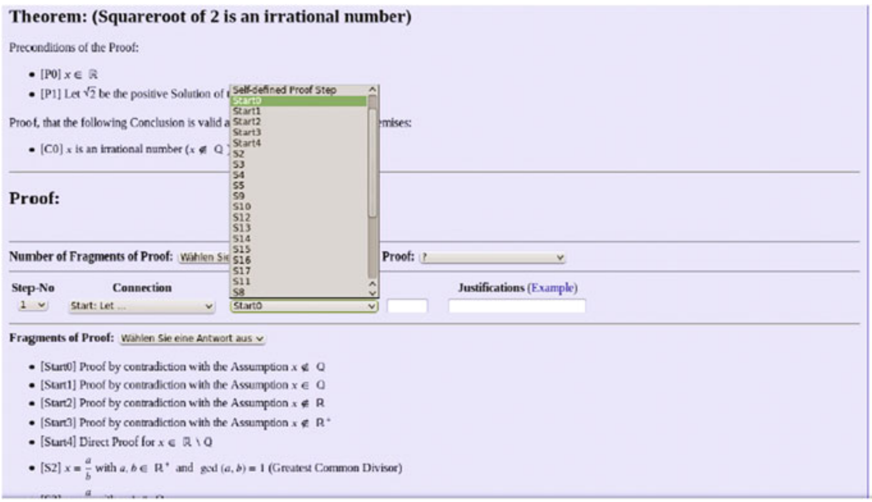


Fig. 8.3 Students' view of the e-proof-environment in IMathAS

16 students (25%) have not started this assessment. They are not included in the numbers below.

#	Question	Grade	Average Score Attempted	Average Attempts (Regens)	% Incomplete	Time per student (per attempt)	Clicked on Help	Preview
1	E-PROOF: Konstante Funktionen - Ableitung	Grade	65%	14.35 (0)	22.9%	22.74 min (1.55 min)	N/A	Preview
2	Funktionentheorie: Wegintegrale berechnen	Grade	66%	4.39 (0)	62.5%	21.59 min (3.5 min)	N/A	Preview
3	E-PROOF: Abelsches Lemma für Potenzreihen	Grade	84%	8.63 (0)	27.1%	16.66 min (1.73 min)	N/A	Preview
4	E-PROOF: Holomorphie von Potenzreihen by Engelbert Niehaus	Grade	53%	11.63 (0)	50%	18.59 min (1.46 min)	N/A	Preview
5	E-PROOF: Umordnungssatz fuer absolut konvergente Reihen	Grade	71%	9.62 (0)	45.8%	17.28 min (1.63 min)	N/A	Preview
6	E-PROOF: Majorantenkriterium fuer Reihen in C	Grade	91%	6.58 (0)	35.4%	9.07 min (1.21 min)	N/A	Preview
7	E-PROOF: Konvergenzradius von Potenzreihen - Ableitung by Engelbert Niehaus	Grade	68%	9.74 (0)	43.8%	18.01 min (1.63 min)	N/A	Preview
8	e-PROOF: Differenzierbarkeit Umkehrfunktion	Grade	79%	8.81 (0)	33.3%	13.12 min (1.23 min)	N/A	Preview

Fig. 8.4 Exemplary statistics on tasks in IMathAS

collect information to derive measures for supporting the learner in learning (cf. Handke and Schäfer 2012), in our case proving. Currently, this can be done via an author-function of the system, which enables the author to see the inputs of each learner in real time, as well as the reaction and summative assessment of the system. (See Fig. 8.4.) Furthermore, it is possible to display a summary of the assessment results and to export the students' answer details.

The teacher can then intervene into the learning progress by addressing the students personally. In the future, the system is supposed to act as an intelligent tutorial system, which supports the learner with automated feedback based on repetitive formative assessment. In addition, research on typical student errors in elementary mathematical proofs is currently being conducted to identify suitable individual didactical help to support the students in their learning process and to derive appropriate distractors (see next section) to enhance the e-proof-system.

8.4.4 *Included Distractors for a Diagnostic Evaluation*

An ideal type of correct and false puzzle fragments allows for diagnostic evaluation transcending the decision of "wrong" or "correct" by providing false proof steps that are similar to the correct ones, so that learners have to read carefully. Standard student errors are considered and appropriate fragments for the prospective standard errors are provided. By applying this method, scores can be given in a more differentiated way and students receive diagnostic feedback referring to their (individual) deficits. Those assessments ought to be considered as a basis for well-targeted individual support of students relating to personal argumentation and proof competence. In addition, those results could be used to adapt teaching in relation to the needs of students. Above that, learners do also receive feedback on their current skills in order to encourage them in their learning process and document their progress in knowledge.

8.5 Research Methodology

The basis of our approach forms the so-called research onion from Saunders et al. (2012) that describes several layers of a process that ought to be passed through. In so doing, the research philosophical stance is a mixture of pragmatism and realism. On the one hand, the stance involves realism because researchers are independent of the test group and so will not create biased results; data are collected via the system, among others, and the learner is not influenced by any researcher. New methods of research should be taken into account, if they could improve the system for a more reliable outcome. On the other hand, pragmatism is used because the system develops itself based on the action of the learners and the teachers in order to create a practical approach used to find solutions for problems.

The research approach is inductive, because so far, few results exist on the development of an e-proof system; thus, research is done to create theory. While doing so, grounded theory is utilized by using inductive methods to predict and explain behavior as well as by building theory and improving the system. Thus, big data are collected from observations via the system, theory and predictions are generated based on this and then those predictions are tested.

Using mixed methods, quantitative and qualitative research methods are combined in the process of the study, data collection, and analysis. The combination of both types of research allows for an offset of limitations of each method and for prevention of gaps of data. As the e-proof system consists of several components and has several layers, all the components have to be regarded separately: the research for appropriate distractors, the student's solutions to enable the provision of decision support to the teacher; and the GUI (Graphical User Interface) to improve the comprehension of the user and to exclude that wrong answers are based on a bad GUI. Furthermore, the system itself has to be tested and improved continually.

Because events and behaviors are analyzed using concentrated samples over a longer period, the time horizon is longitudinal. Data collection will be done via the system by collecting the answer of the learners and the inputs of the teachers, but also by comparing paper-and-pencil based solutions to the inputs into the system to refine the functionality of the e-proof system. Data collection will be realized via questionnaires and interviews/observations based on the learners' and teachers' expectations and impressions with the system. By implementing diagnostic distractors into the system, careless mistakes or guessing will be considered to increase the diagnostic result obtained by the system.

As an initial step, 144 written examinations on a lecture at the University of Koblenz-Landau which was supposed to impart basic proving skills for teacher trainees for primary and secondary education were analyzed, as well as 3 written examinations by mathematically gifted pupils who participated in the same lecture (cf. Platz et al. 2017). This study was based on a pilot study for distractor determination of Platz and Niehaus (2015). Typical student mistakes were derived from the students' solutions to enable the implementation of the proving tasks as e-proofs into the system, including analytical distractors as exercise support for students participating in the lecture in further semesters. Methodologically, the study is based on the concept of Winter (2011) about distractor determination for electronic assessments. Initially, the proofs prompted in the examination were implemented into the e-proof system by decomposing them into proof fragments (without having knowledge about the students' errors). Furthermore, expected students' errors were implemented as distractors.

Based on this study, student mistakes could be determined from students' paper-and-pencil based solutions and they could be formulated as distractors for implementation into the e-proof system. The determined distractors were compared with the initial e-proof decomposition (without knowledge about the students' errors), which was in very small steps and linear. The actual students' errors differed from the expected students' errors. It became obvious that this kind of decomposition might not be helpful for learners, if rote memorization of the single proof steps should be avoided. The focus should rather be shifted to finding and understanding the proof idea.

Consequently, the determined distractors will be implemented into the e-proof system to enable better support of the learners. This concept will be tested in the same lecture at the University of Koblenz-Landau in order to detect if this concept is fruitful and how this concept can be optimized.

8.6 Benefits and Challenges

Using a free of charge web-based Internet Mathematics Assessment System appears to be the most obvious advantage of the application, although there may be further benefits for both learners and teachers as well as challenges. First, the system includes a forthcoming tutorial component, a tool that supports learning as an adaptive system. This tutorial component contains the substantial and methodological potential of fuzzy logic and artificial neural networks to include a tutorial component for learning, to improve automatic correction, and to enable an automatic valuation with possibly low manual correction efforts. Fuzzy logic can be used to represent linguistic values and arguments used within mathematical proving by the learners, which tend to be fuzzy. Artificial Neural Networks are used to enable the system to learn and to act as an intelligent tutorial system. Moreover, an immediate personal, but at the same time anonymous, feedback to single tasks can be given. This fosters real-time content-specific learning and encourages shyer students to work on different problems, even though they may not be likely to solve tasks in a more personal atmosphere or face-to-face interaction. In addition, working with new media can be exciting and motivating for learners and contributes to greater engagement in learning. Furthermore, gamification, that is the use of video game elements in non-gaming systems to improve user experience and user engagement (Deterding et al. 2011), could be implemented into the system in the future. As a first step towards the realization of gamification, an earn-as-you-learn reward-based badge system (cf. Herselman and Botha 2014) will be incorporated.

Second, new technology methods adjusted to the needs of learners can be applied and struggles with writing mathematical formula language may be avoided due to implemented fragments. Nevertheless, these fragments still have to be retraced in mind and set into a logical order, so that the comprehension of mathematical language is still required. For teachers, a targeted pool of shared tasks may reduce time for preparation as well as correction if they can fall back on well-established tasks and evaluations. Above that, personal experiences can be shared and discussed by building up networks.

Both teachers and learners have to face various challenges concerning the use of a web-based assessment system. First and foremost, difficulties could arise if learners may not be very adept at working with similar programs, because for a novice the graphical user interface (GUI) is not easy to comprehend. Therefore, the GUI should be modified or adapted to facilitate input, editing, and correction. So far, students' own proof steps still have to be corrected manually by the teacher. An improvement of the correction effort at this point would be desirable.

This kind of working on a given task is not as free as with paper-and-pencil proof. There can be a large number of proofs for one theorem and some of these proofs might not be considered yet; even if a proof step does not contribute to the proof, it is not logically wrong and the proof is still correct. Also, because there are limited data space and time, it is not possible to include each conceivable proof for a theorem into a system. This provides a challenge to software development concerning the implementation of the requirements of the system to support learners in gaining proof competencies. The other very difficult question to answer is "What is a good proof?". In order to claim interdepartmental validity, such a derivation must be documented by other true statements, that is, the procedure is coded so that the chain of arguments or the complete traceability to axioms can be examined by the members of the professional community, which is in our case the community of teachers. The so-called "community" is thus responsible for an evaluation and validation function, as it tests the claim of validity of a proof. A proof thus establishes a reasoning, which, if found to be valid, is generally accepted (Brunner 2014).

So far, no big community developing e-proofs has been formed and no large number of proofs is included into the public library of IMathAS, which still needs to be created. For this reason, the collaboration of and sharing between different institutions should be cultivated and the awareness of open-source Internet Assessment systems has to be increased. Therefore, the system should be enhanced and optimized continuously to minimize challenges and derive maximum benefits. In so doing, a Community of Practice (cf. among others Hoadley 2012) as a collaborating group of people sharing the same aims can be established.

Table 8.1 Example of a chain of decisions in evidence based medicine analogous to mathematical logic

Connection	Proof step decision	Justification (reason)
…	Diagnosis X	Anamnesis of patient X
Implies	Treatment Y of patient X	Scientific results for disease X

8.7 Further Perspectives

IMathAS may not only be limited to a mathematical context, but the generative method could also be transferred to various other fields, such as evidence-based medicine. As decisions of treatment cannot clearly be divided into "wrong" or "right" (in analogy to mathematical logic), there are similar structures of decision making found by using the best medical evidence available (cf. Sackett et al. 1996). In Table 8.1, an example of such a chain of decisions is given.

On that account, IMathAS could also contribute to initial medical studies as well as support prospective doctors during the practical year with virtual case studies without severe consequences for any patient. However, contents in this field still need to be implemented.

8.8 Summary and Conclusion

In this paper, a theoretical framework to identify preferences of learners for logical reasoning was presented. This framework determines the structure of the IT environment. Therefore, a brief introduction into an open source environment, namely IMathAS, was given. As the system was established in all universities and educational facilities in Rhineland-Palatinate in Germany, collaboration on a joint repository was encouraged and will be implemented in order to develop tasks and solutions being adapted to the individual needs of learners and educational facilities. Based on the former results, a generic system should be developed in the future, in order that no limits are set on the content design. The aim is the development of an adaptive learning, test and diagnostic system, including the substantial and methodological potential of fuzzy logic and artificial neural networks to include a tutorial component for learning and to improve automatic correction and to enable automatic valuation with possibly low manual correction effort.

Based on a first study described in Platz et al. (2017), a focus shift of the system seems reasonable: the detection and understanding of the proof idea is essential and should be supported. As a first step towards this goal, the determined distractors derived from paper-and-pencil based students' solutions will be implemented into the e-proof system to enable better support for the learners. This concept will be tested in lectures focused on imparting basic proof competencies to detect if this concept is fruitful and how this concept can be optimized. Furthermore, the design of the system will be reconsidered and the potentials and limitations of software development and realization in this context will be sounded.

References

Alcock, L., & Wilkinson, N. (2011). E-proofs: Design of a resource to support proof comprehension in mathematics. *Educational Designer, 1*(4), 1–19.

Bell, A. (1976). A study of pupil's proof explanation in mathematical situations. *Educational Studies in Mathematics, 7*(1/2), 23–40.

Brunner, E. (2014). Was ist ein Beweis? In Mathematisches Argumentieren (Ed.), *Begründen und Beweisen* (pp. 7–25). Berlin/Heidelberg, Germany: Springer.

Budke, A., & Meyer, M. (2015). Fachlich argumentieren lernen—Didaktische Forschung zur. *Argumentation in den Unterrichtsfächern*, 9–30.

Cataloglu, E. (2007). Internet-mediated assessment portal as a pedagogical learning tool: A case study on understanding kinematics graphs. *European Journal of Physics, 28*(4), 767–776.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: Defining gamification. In A. Lugmayr, H. Franssila, C. Safran, & I. Hammouda (Eds.), *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). New York, NY: ACM.

de Villiers, M. (1990). The role and function of proofs in mathematics. *Pythagoras, 24,* 17–24.

Ensley, D. E., & Crawley, J. W. (2006). *Discrete mathematics: Mathematical reasoning and proof with puzzles, patterns, and games*. New York, NY: Wiley.

Goldberg, A. V., & Harrelson, C. (2005). Computing the shortest path: A search meets graph theory. In *Proceedings of the sixteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 156–165). Philadelphia, PA: Society for Industrial and Applied Mathematics.

Handke, J., & Schäfer, A. M. (2012). *E-learning, E-teaching und E-assessment in der Hochschullehre*. Eine Anleitung: Oldenbourg Verlag.

Herselman, M., & Botha, A. (2014). *Designing and implementing an information communication technology for rural education development (ICT4RED) initiative in a resource constrained environment: Cofimvaba school district, Eastern Cape, South Africa*. Pretoria, South Africa: CSIR Meraka.

Hoadley, C. (2012). What is a community of practice and how can we support it? In D. Jonassen & S. Land (Eds.), *Theoretical foundations of learning environments* (pp. 287–300). Mahwah, NJ: Lawrence Erlbaum Associates.

Kramarski, B., & Mevarech, Z. R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40*(1), 281–310.

Ladewig, R. (2016). *Schwindel: Eine Epistemologie der Orientierung* (Vol. 6). Mohr Siebeck.

Medienpädagogischer Forschungsverbund Südwest (MFS) (2000–2015). KIM-Studie 1999–2014. *Kinder und Medien, Computer und Internet*.

Meyer, M. (2007). *Entdecken und Begründen im Mathematikunterricht: Von der Abduktion zum Argument*. Hildesheim: Franzbecker.

Niehaus, E., & Faas, D. (2013). Mathematische Beweise in elektronischen Klausuren in der Lehramtsausbildung. In G. Greefrath, F. Käpnick, & M. Stein (Eds.), *Beiträge zum Mathematikunterricht 2013* (pp. 704–707). Münster: WTM-Verlag.

Pavlik, J. V. (1998). *New media technology. Cultural and commercial perspectives* (2nd rev. ed.). Boston, MA: Allyn and Bacon.

Platz, M., Krieger, M., Niehaus, E., & Winter, K. (2017). Distraktorenermittlung für elektronische Beweise in der Lehre. In *Beiträge zum Mathematikunterricht 2017*. Münster: WTM-Verlag.

Platz, M., & Niehaus, E. (2015). To “E” or not to “E”? That is the question. Chance & Grenzen eines E-Proof-Systems zur Förderung von Beweiskompetenzen. In F. Caluori, H. Linneweber-Lammerskitten & C. Streit (Hrsg.), *Beiträge zum Mathematikunterricht 2015* (pp. 704–707). Münster, Germany: WTM-Verlag.

Sackett, D. L., Rosenberg, W. M., Gray, J. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine. *British Medical Journal, 313*(7050), 170–171.

Saunders, M., Lewis, P., & Thronhill, A. (2012). *Research methods for business students* (4th ed.). Harlow, England: Pearson Education Ltd.

Winter, K. (2011). Entwicklung von Item-Distraktoren mit diagnostischem Potential zur individuellen Defizit- und Fehleranalyse. In *Didaktische Überlegungen, empirische Untersuchungen und konzeptionelle Entwicklung für ein internetbasiertes Mathematik-Self-Assessment*. Münster, Germany: WTM-Verlag.

Author Biographies

Dr. Melanie Platz currently works as deputy professor at the University of Siegen (Germany) in the field of mathematics education. In her research, she is focused on the development of an electronic proof system and the determination of analytical distractors in this context. Furthermore, she applies mathematical modelling for the mathematical optimization of spatial public health using the One Health approach in the ICT4D context. In didactics, her focus is on problem-oriented mathematics teaching by being embedded in concrete applications and the use of digital media in education.

Miriam Krieger is a prospective Ph.D. student from the Europe-University of Flensburg. She holds a B.Sc. and a M.Ed. in mathematics and biology from the University of Münster and was awarded a Deutschlandstipedium scholarship in 2011. Currently, she also teaches mathematics and biology at a German comprehensive school. Her research focuses mainly on electronic proofs, distractors and language diversity in classrooms.

Prof. Dr. Engelbert Niehaus is a full Professor of the Department of Environmental and Natural Sciences, University of Koblenz-Landau, Germany. In this capacity, he lecturers in Pure and Applied Mathematics and Mathematics Education. His research interests focus on adaptive modelling of learning processes and the application on neural networks in e-learning environments. He obtained his Ph.D. at the University of Münster in 1995 in Mathematics and Computer Science. His Ph.D. studies focused on topological algebras. After his Ph.D., he worked as a teacher in Mathematics and Computer Science. He finished his Habilitation in 2004 on adaptive modules in computer based learning environments for Mathematics.

Prof. Dr. Kathrin Winter is a full Professor of the Department of Mathematics and its Didactics at the Europe-University of Flensburg. In this capacity, she lectures in Applied Mathematics and Mathematics Education. Her research focuses on individual diagnostics, encouragement and consulting programs especially for pupils, students and vocational trainees. Therefore, she researches and develops adaptive online assessments including an individual diagnostic feedback —for example in the context of an e-proof environment.

Part IV
Statistical Models for Formative Assessment

Chapter 9
Cognitive Diagnostic Assessment: An Alternative Mode of Assessment for Learning

Carolyn Jia Ling Sia and Chap Sam Lim

Abstract This paper discusses how cognitive diagnostic assessment can be used as an alternative assessment for learning the topic of "time" in primary mathematics. Cognitive diagnostic assessment is an assessment that can provide meaningful feedback on student's cognitive strengths and weaknesses. The one described here was initially developed by the researchers and validated by a panel of seven expert primary mathematics teachers. It was then administrated to 127 Grade Six (12 years old) students. In this paper, several examples are given to illustrate how cognitive diagnostic assessment could be used as an assessment for learning. The limitations of using this approach as well as future research directions are also discussed.

Keywords Cognitive diagnostic assessment · Assessment for learning
Mathematics assessment · Attribute hierarchy method

9.1 Introduction

Assessment is a process of making judgment on students' learning progress. An assessment can be summative or formative depending on its purpose. Summative assessment is used to summarize or to capture the overall performance of students' learning whereas formative assessment is used with a smaller scope of the curriculum and mainly to monitor the learning progress of a student. Based on the definition of these terms, Black and William (1998) have equated assessment *of* learning with summative assessment and assessment *for* learning with formative assessment.

C. J. L. Sia (✉) · C. S. Lim
School of Educational Studies, Universiti Sains Malaysia,
11800 Gelugor, Penang, Malaysia
e-mail: c.sjling@gmail.com

C. S. Lim
e-mail: cslim@usm.my

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_9

According to the Assessment Reform Group, assessment *for* learning is defined as "the process of seeking and interpreting evidence for use by students and their teachers to decide where the students are in their learning, where they need to go and how best to get there" (2002, p. 1). Some examples of assessment for learning are journal writing, learning logs, written assessment, and item analysis of summative assessment. Among these, *cognitive diagnostic assessment* is an instrument that can help make formative inferences on students' cognitive strengths and weaknesses in a specific topic (Wu et al. 2012). Thus, we propose to use cognitive diagnostic assessment as an alternative mode of assessment for learning because its characteristics fulfill the requirement of assessment for learning.

Cognitive diagnostic assessment presented in this paper was the product of a larger project. Although the research objective of the project was to develop and validate this technique as an assessment framework in the learning of "time," this paper focuses on the use of cognitive diagnostic assessment as assessment for learning.

Time is an important concept in our daily life, for instance, in the duration of an event or the relationship between different units of time. Time is a basic concept of mathematics that is related to representations and symbol systems, such as clocks, calendars, lunar cycles and time-tables (Kelly et al. 1999). However, some researchers (e.g., Burny et al. 2009) have concluded that "time" is a complex concept that is not easy to teach to children. Rather, time is an abstract concept, and thus, it is difficult to link with other topics of measurement. For example, we cannot go to a shop and buy a quantity of time (McGuire 2007, as cited in Harris 2008). Thus, the cognitive diagnostic assessment described in this paper was developed as an instrument to identify learner's difficulties in learning the topic of time. Perhaps by using this assessment, the way a learner learns the topic of "time" can be revealed, and thus assist teachers' instructional planning.

9.2 Literature Review

Since the 1980s, researchers (e.g., Linn 1986; Messick 1984; Snow and Lohman 1989) have argued for the fusion of cognitive science and psychometric models. The argument was raised due to dissatisfaction with conventional assessments that are unable to provide in-depth information regarding students' learning progress (Glaser 1981). Nichols (1994) called this diagnostic assessment, which combines cognitive science and psychometrics, as *cognitive diagnostic assessment*.

9.2.1 Cognitive Diagnostic Assessment

According to Jang (2009), cognitive diagnostic assessment is an alternative assessment that aims to provide fine-grained analysis of students' skill mastery profiles and their cognitive knowledge state. It can also be used to measure students' cognitive structures and processing skills (Leighton and Gierl 2007). This fine-grained result is crucial as two students who score the same marks on a test might have different levels of skill mastery (or cognitive level). It helps the students to take essential actions to close the gap between their current competency levels and their desired learning goals (Black and William 1998).

Because cognitive diagnostic assessment consists of the psychology of learning, it explicates the assumptions about the knowledge, skills, or processes that are possibly possessed by the students. Assessment tasks are designed based on these assumptions. Then, scores are assigned based on students' responses. Thus, it is believed that cognitive diagnostic assessment is able to make inferences about students' specific knowledge structures and processing skills (Alves 2012). These inferences provide adequate knowledge about the teaching and learning process. In addition, these inferences should also provide sufficient information about students' learning progress to plan appropriate remedial steps as well as to carry out any instructional intervention. As stated by Ketterlin-Giller and Yovanoff (2009), the power of this model in identifying students' underlying cognitive processes, which could not be found in skill analysis and error analysis, can be used to enhance the remedial process.

Nichols (1994) proposed a five-step framework to develop cognitive diagnostic assessment: (1) substantive theory construction; (2) design selection; (3) test administration; (4) response testing; and (5) design revision. To begin the development, a substantive base constructed from past studies or theories of learning is essential. This substantive base includes assumptions regarding how an individual learns. Consequently, this base serves to guide item design to make inferences on a student's knowledge, skill, or process structure in solving the item.

After constructing the substantive theory or base, the second step of development is design selection. There are two choices of designs: observation design and measurement design. *Observation design* refers to the properties of assessment activity, for instance, the content of the assessment items and how these items are organized. Observation design aims to arrange the structure of the items to ensure that the students' ways of thinking can be observed based on their responses. Meanwhile, *measurement design* describes the procedures that were used to analyze and classify the responses. It aims to gather and combine students' responses to identify their ways of solving items.

The third step of the development involves administration of the designed assessment. This is to collect students' responses. These responses will be used to evaluate the designed items and arrangement of assessment activities. The next step of development is response scoring, which is implementation of the test theory. This step involves managing students' scoring. From this step, evidence about students' knowledge and process-structure, as well as differences among the structures, are identified.

These inferences are used to revise the observation and measurement designs used in the development process. There are two sources of evidence that can support the designs used: the extent of fit between the expected responses and the students' responses; and related past studies. If the evidence gathered is insufficient to support the selected designs, the selected designs may require review.

9.2.2 Cognitive Diagnostic Assessment in Education

Cognitive diagnostic assessment has been applied to various subjects, such as mathematics (Alves 2012; Broaddus 2011; Ye 2005) and English (Jang 2009; Lee and Sawaki 2009; Wang and Gierl 2011). For mathematics learning, the model has been applied to several topics, such as algebra, fractions, mixed number subtraction (Mislevy 1996; Sinharay and Almond 2007), multiplication and division with exponents (Birenbaum and Tatsuoka 1993), pre-algebra (Ye 2005), and counting numbers forward (Alves 2012).

In addition to developing cognitive diagnostic assessment for learning, researchers have investigated the usefulness of this model in education. For instance, Wang and Gierl (2011) found that, based on the results obtained in their study, the number of questions a student answered correctly was not proportional to the number of attributes that the student had mastered. In other words, a student who received a higher score did not necessarily master more attributes than a student who received a lower score. This finding revealed that results obtained from cognitive diagnostic assessment could reflect a student's learning progress better than a single score did. Thus, teachers can plan and design the lesson better according to students' learning situations.

Several researchers, such as Ketterlin-Giller and Yovanoff (2009), Russell et al. (2009) and Wu et al. (2012), have realized the effectiveness of using diagnostic results to identify students' misconceptions. Identification of students' misconceptions is vital as it serves as the basis of effective remedial materials preparation and design (Ketterlin-Giller and Yovanoff 2009). If cognitive diagnosis and instructional design fail to connect, students might not be able to receive the instructional support needed to overcome their misconceptions. This statement was supported by Russell et al.'s (2009) study whereby diagnostic information obtained from cognitive diagnostic assessment was used to help students conceptualize. The results showed that students' algebra ability had improved after the remediation. These studies show that this model has great potential to assist teachers to assess students' actual learning progress. Thus, they can make appropriate instructional decisions based on students' needs.

9.2.3 Theoretical Framework

According to Alves (2012), cognitive diagnostic assessment has four main characteristics that seem to fulfill the principles of assessment for learning (Assessment Reform Group 2002). Table 9.1 compares similarities between the principles of assessment for learning and the characteristics of cognitive diagnostic assessment.

Table 9.1 Similarities between assessment for learning and cognitive diagnostic assessment

Principles of assessment *for* learning (AfL)	Characteristics of cognitive diagnostic assessment (CDA)
Assessment for learning should be a part of effective instructional planning. Students should engage actively and conduct self-assessment throughout the process of learning.	Formative inferences on student's learning can be made, and thus, help in instructional decision making. The formative approach motivates students to engage more in learning as this encourages them to use assessment as a tool in the learning process (Jang 2009).
AfL emphasizes constructive feedback from both teacher and assessment, and how the feedback can be used to help students to improve themselves.	CDA provides fine-grained test scores, which yield information on whether students have mastered the defined set of attributes (knowledge, skills, or processes). This meaningful feedback provides constructive feedback as students are given opportunity to enhance their learning, as they know their strengths and weaknesses precisely.
AfL is designed based on students' learning process.	CDA is designed based on students' cognitive processes when solving the problems.
AfL is a tool used by teachers to make judgments based on students' responses (feedback from students) and make decisions about how to help the students and encourage the process of learning to occur.	CDA has the possibility to provide a better quality of diagnostic feedback to students and educational stakeholders. This diagnostic feedback will assist teachers to identify the needs of their students.

As shown in Table 9.1, it is clear that cognitive diagnostic assessment can be a good instructional tool to be used as assessment for learning. It provides meaningful feedback to the group of educational stakeholders on what the students have learned and how they are doing in their class. The report of assessment is no longer a piece of paper with all the grades and scores of different subjects.

The cognitive model that is created helps experts discover how students learn cognitively. When teachers or experts know how their students learn, they can easily identify the misconceptions in students' learning. Besides, experts can make a comparison between students' expected way of learning by the experts and the students' actual way of learning. Therefore, this comparison contributes to effective instructional planning.

Furthermore, cognitive diagnostic assessment helps students to engage actively in the teaching and learning process. It provides them precise information of themselves. They now have better insight on their strengths and weaknesses. This encourages students to take part in learning and work on their weak parts rather than just working blindly on everything. This, in a way, should motivate students to do self-learning and monitor their own learning progress, which is also the purpose of assessment for learning. In addition, this should build students' confidence towards mathematics, especially students who struggle with mathematics or have phobias related to mathematics. As mentioned by Gierl and Alves (2010), instructional interventions can then focus on reinforcing strengths while overcoming weaknesses.

9.3 Methodology

Because the main focus of this paper is to introduce cognitive diagnostic assessment as an alternative mode of assessment for learning, this section describes the development of the specific assessment used in the study reported here.

9.3.1 Identification of Attributes

In this study, we began the development of a cognitive diagnostic assessment by defining a set of 35 attributes on the topic of "Time." These attributes described the knowledge, skills, or processes that were required to solve a mathematical task correctly. Attributes can be identified by task analysis, protocol analysis, or expert review on curriculum specifications. In this study, expert review on primary mathematics curriculum specifications was first carried out, followed by the task analysis of past years' public examination questions. The curriculum specifications were reviewed to understand the scope of *time* intended to be taught throughout the six years of primary mathematics education. Then these syllabi were used to identify the knowledge or skills that were measured by the items in the Malaysian Primary School Achievement Test.

All the test items that were related to Time from the past five years (2009–2013) Malaysian Primary School Achievement Test were compiled. This set of selected items was then given to a panel of experts, which consisted of seven experienced primary mathematics teachers. Based on their teaching experiences and knowledge, they identified the questions that might be problematic for their students. This step was taken to ensure that the instrument developed later would be helpful for the students to overcome their difficulties in learning the topic of Time.

Based on the experts' responses, the 11 most problematic questions were shortlisted and piloted with 30 Grade Six (age 12) students. This step aimed to confirm if these items represented the problematic areas in learning the topic of Time from the students' perspective. Analysis of the students' responses confirmed that these students were weak in calculating the duration of an event and converting units of time. Later, based on these findings, the panel of experts was asked to solve each of the problematic tasks and state the required attributes needed to solve the problematic tasks correctly. Consequently, a list of attributes was identified. The list of attributes was then revised several times based on the students' responses until the final cycle of development.

9.3.2 Formation of the Cognitive Model

This set of attributes was linked to form a cognitive model. The panel of experts then arranged these relevant attributes hierarchically from the most basic to the most complex. Cognitive modelling is crucial as it yields the cognitive structure of students in solving the problems and a framework for item design (Roberts et al. 2009). It serves as fundamental to the assessment so that useful and detailed diagnostic feedback can be provided. Each cognitive model aimed to measure a specific learning outcome. In the developed assessment, there were 11 cognitive models that related to time. Models 1–8 focused on clock time while models 9–11 focused on calendar dates. Table 9.2 displays examples of cognitive models.

In this study, each cognitive model consisted of at least five attributes. Each attribute was measured by three parallel items. The parallel items serve two purposes: (a) to ensure the reliability of inferences (Alves 2012); and (b) to avoid the probability of guessing and slip. For students who had mastered the attributes, they were expected to answer all three items correctly. Students who were not so familiar with the attribute might answer one or two items correctly. Students who did not master the attribute would not be able to answer any of the items correctly.

9.3.3 Construction of Items

Three parallel items were needed to measure each attribute because the result from only one item for each attribute might not be sufficient to conclude the student's mastery level on that attribute. If two items were used and the student only answered one of them correctly, there was no valid conclusion. Therefore, at least

Table 9.2 Examples of cognitive models in the designed cognitive diagnostic assessment

Cognitive model	Description
1	Finding ending time of an event in hours and minutes
2	Finding duration of an event in hours and minutes
3	Finding starting time of an event in hours and minutes
4	Finding ending time involving both 12-hour and 24-hour time systems
5	Finding duration involving both 12-hour and 24-hour time systems
6	Finding starting time involving both 12-hour and 24-hour time systems
7	Finding the total duration of two events in minutes and seconds
8	Finding the difference between the duration of two events in minutes and seconds
9	Finding the ending date when involving concept of "After"
10	Finding the starting date when involving concept of "Before"
11	Finding the duration of two inclusive dates

three items were required to identify a student's performance on a specific attribute (Alves 2012). In addition, a substantial number of items (i.e., three) for each attribute were also required to achieve diagnostic purposes of the assessment.

9.3.4 Administration of Assessment and Procedure of Data Analysis

As the chosen psychometric model for this cognitive diagnostic assessment was Attribute Hierarchical Method, the attributes were arranged hierarchically (i.e., from the most basic attribute to the most complex attribute) in constructing the cognitive models. The panel of experts constructed 75 items based on the 11 cognitive models. The resulting set of items was then administered to 127 Grade Six students (age 12) and the students' responses were analyzed using the Attribute Hierarchical Method. This statistical procedure was used to analyze the observed responses in terms of mastery level for each attribute (Hubner 2010). Finally, based on the results generated from the psychometric procedure, a cognitive profile for each learner was generated. This cognitive profile presented a defined list of attributes that measured the learner's mastery level on each specific attribute. Hence, the cognitive profile helped the teachers to identify the cognitive strengths and weaknesses of the learner.

Note that each student's response was marked dichotomously with 1 indicating a correct response and 0 an incorrect response. The marking was based on each student's working steps in solving each item. If a student showed that he/she had mastered the attribute that was intended to be measured by the tasks, then he/she would be scored as *correct* or 1 for the item. Conversely, if a student got a correct final answer by doing an incorrect/inappropriate working step, then he/she would be scored as *incorrect* or 0 for the item. Hence, the accuracy of the final answer was not the sole consideration in the marking procedure; instead, the student's working steps to get the final answer were given more priority. Therefore, during the administration of the assessment each student was encouraged to write down his/her working steps clearly while solving the task as the student's working steps would help the researchers to better understand their students' cognitive processes.

As an illustration, Fig. 9.1 shows an example of a student whose response was scored as correct even though his final answer was incorrect. This student knew that to find starting time, he would need to subtract the duration from the ending time. He also knew that he needed to regroup 1 hour to 60 min, so that he could have enough minutes (60 + 5) to subtract 45 min. However, during the process of calculation, he made a mistake: 65 − 45 = 25 (should be 20). Hence, he ended up with the final answer as 6:25 p.m. instead of 6:20 p.m. (the correct answer). Nevertheless, this student still scored 1 for this response as he had shown that he had mastered the attribute (as discussed above) that was intended to be measured by this item.

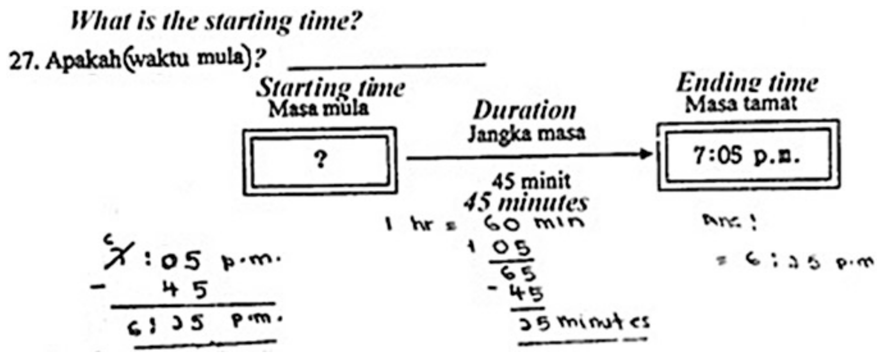


Fig. 9.1 An example of a student's response of which the final answer was incorrect but was still given a score of 1 (*Note* The Bold Italic words show English translation of the Malay words)

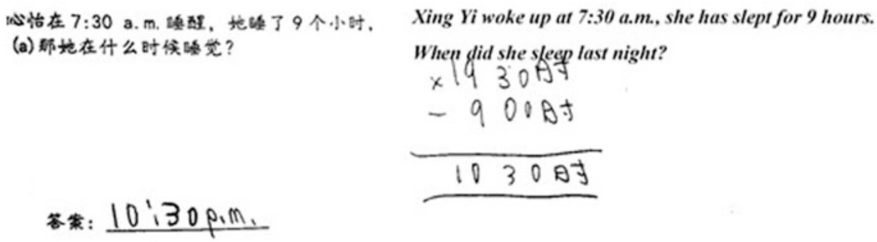


Fig. 9.2 An example of a student's response with a correct final answer but which was scored 0

Conversely, in Fig. 9.2, the student converted 7:30 a.m. to 1930 hours, then subtracted nine hours to get the answer. Although the final answer was correct (10:30 p.m.), his working step was not appropriate. He converted incorrectly between the 12-hour and 24-hour time systems; 7:30 a.m. is equivalent to 0730 hours, not 1930 hours. This shows that this student might have been confused between 12-hour and 24-hour time systems.

9.4 Findings and Discussion

Only data from one cognitive model are discussed here. The final assessment was administered to 127 Grade Six students, but only data of 10 students are reported to illustrate how cognitive diagnostic assessment can be used as assessment for learning. Table 9.3 displays the list of attributes for the first model which measures problem solving involving addition of start time and duration to find end time (in hours and minutes).

Table 9.3 List of attributes for cognitive model 1

	Description of attribute
A1	Knowing 1 hour = 60 min
A2	Addition involving time/duration without regrouping (hour and minutes)
A3	Addition involving time/duration with regrouping (from minutes to hour)
A4	Finding end time as involving addition of start time and duration
A5a	Transform word context into mathematical operation to find end time (in hour and minutes)
A5b	Reasonability of final answer as end time for word problems

Table 9.4 shows the observed response patterns and mastery levels of ten students on 16 items in this first model. All the responses were converted to a binary pattern with 1 indicating a correct response and 0 a wrong response. Similarly, 1 in the mastery level means the student mastered the attribute, whereas 0 means non-mastery.

There were 16 items in Cognitive Model 1 which measured six attributes. All attributes except Attribute A1 were measured by three items. A1 (knowing 1 hour = 60 min) was a very basic knowledge which could not been questioned in any other way; thus, it contained only one item (1 hour = how many minutes).

As shown in Table 9.4, although two candidates, C and I have the same score (11/16), their patterns of mastery level are different. Candidate C has mastered A1, A2, A4 and A5a whereas candidate I has mastered A1, A2, A3 and A4. Similar results were also found in Wang and Gierl's study (2011). They found that some students obtained the same score but did not necessarily possess the same pattern of attribute mastery. These students might have different knowledge states at the time of testing. This shows that the feedback provided by the cognitive diagnostic assessment is much more meaningful as it provides detailed information on a student's learning progress instead of just the total score (Ye 2005). In other words, this assessment can effectively be used to differentiate students who obtain the same score, subsequently identifying special issues that students have. This will certainly help teachers in grouping their students based on their students' strengths and weaknesses, allowing more effective differentiated teaching (Wu et al. 2012).

Next, although four candidates (A, E, H and J) appear to have mastered all six attributes, only candidates A and E have full scores (16/16). Meanwhile, candidates H and J scored 14/16 and 15/16, respectively. This result was unexpected as students were expected to answer the items correctly if they had mastered the attribute(s) measured by the items. Thus, candidates H and J might have committed some careless mistakes in one or two items. Based on the result, students will know which parts of the topic they should work harder or focus on. Cognitive diagnostic assessment provides a clearer goal for the student to achieve, which should certainly encourage the student to become more self-reflective and engage in learning (Alves 2012).

Table 9.4 Observed response pattern and mastery level of ten grade six students on six attributes

Candidate	Total score	Observed response pattern						Mastery level					
		I1	I2I3I4	I5I6I7	I8I9I10	I11I12I13	I14I15I16	A1	A2	A3	A4	A5a	A5b
A	16/16	1	111	111	111	111	111	1	1	1	1	1	1
B	4/16	1	111	000	000	000	000	1	1	0	0	0	0
C	11/16	1	111	100	111	110	010	1	1	0	1	1	0
D	12/16	1	111	101	111	110	100	1	1	1	1	1	0
E	16/16	1	111	111	111	111	111	1	1	1	1	1	1
F	4/16	1	111	000	000	000	000	1	1	0	0	0	0
G	13/16	1	111	111	111	111	000	1	1	1	1	1	0
H	14/16	1	111	101	111	110	110	1	1	1	1	1	1
I	11/16	1	111	111	110	010	000	1	1	1	1	0	0
J	15/16	1	111	101	111	111	111	1	1	1	1	1	1

Because the tasks of the assessment were designed based on how a student learns, we observed that all candidates were following the expected way of learning except candidate C. For instance, candidate G showed mastery of A1 to A5a, while candidate F only mastered A1 and A2. However, candidate C showed mastery of A1, A2, A4 and A5a. This pattern indicates that candidate C did not need to master A3 before A4 as we expected. An alternative explanation might be the candidate committed systematic errors for the tasks involving A3. Thus, based on this feedback, the teachers can find out more about how their students learn, and find ways to improve their instructional strategies (Ketterlin-Giller and Yovanoff 2009).

Finally, as each attribute was measured by three parallel items, this allows the teacher to gather sufficient evidence on their students' mastery level on a defined set of attributes or to identify persistent errors that might have been committed by their students. Subsequently, they could make formative inferences based on the evidence and provide their students appropriate suggestions and advice. As students' misconceptions in learning can be diagnosed in-depth (Buck and Tatsuoka 1998; Hartz et al. 2002), teachers can tailor their teaching approaches to individual student needs and stages of mental development, and direct students to achieve teaching objectives.

9.5 Conclusion

Cognitive Diagnostic Assessment is an assessment that can provide fine grain feedback to both teachers and students, which enhances the learning process, and thus achieves the purpose of assessment for learning. By using the statistical procedure, a score report can be generated. The score report displays attributes being measured and the student's mastery level on each attribute. It provides feedback on a student's strengths and weaknesses to educational stakeholders, particularly teachers and parents, and thus enhances effectiveness in preparing remedial work (Ketterlin-Giller and Yovanoff 2009; Wu et al. 2012).

Because the main objective of the larger project was to develop the cognitive diagnostic assessment, the analysis of students' responses was not shared with the students. However, in the future this can be done to help students obtain an idea on their learning progress as well as understand the difficulties that they have had in learning. This would then help them decide the actions to be taken in order to achieve the learning goal. Thus, we would have achieved the purpose of using assessment as a tool for learning.

However, in the process of developing the cognitive diagnostic assessment, we identified some limitations. It is time consuming to develop such as assessment as it is fine grained and developed based on students' ways of learning. In addition, lacking human resources is one limitation in developing such an assessment. To develop a cognitive diagnostic assessment, project leaders will need a panel of experts (experienced teachers) to identify the attributes and design the tasks. The assessment will need to be administered to a substantial number of students to increase the validity of the instrument. To make this process happen, a team of committed experts and supportive schools are needed throughout the entire process. Thus, teachers should be providing or guiding development of this kind of assessment through professional development programs, such as a two-day workshop. Then, experienced teachers can contribute based on their teaching experiences in the process of development. This will certainly encourage collaboration between novice teachers and experienced teachers.

Because the analysis of data involves psychometric procedures, teachers often cannot do it by themselves without the relevant psychometric software and the assistance of experts. Nevertheless, the latter issue might be addressed by developing a computerized system for cognitive diagnostic assessment.

The cognitive model constructed is very specific, and the items are constructed based on the cognitive model. These items can only be reused in the cognitive diagnostic assessment for topics related to that particular concept, such as time as prior knowledge. Meanwhile, if a teacher wants to develop such an assessment in the topic of algebra or fractions, he or she might need to go through the whole process of developing a cognitive diagnostic assessment, starting from the identification of attributes. We acknowledge this constraint in constructing such assessments. However, to reduce the workload, the teacher might use existing items (e.g., from other large-scale assessments or teaching materials) that can measure the attributes that he or she intends to measure.

From the perspective of future study, it is recommended to focus on the development of modules of remedial materials. Cognitive diagnostic assessment is used to diagnose skills that students have mastered and skills which are yet to be mastered at the end of a lesson. After the diagnostic process, remedial alternatives will be needed to help students cope with learning difficulties related to the interested domain. In other words, remedial work is an important process for both teachers and students as it helps teachers to apply suitable modules in order to accommodate students of different background knowledge or learning progress to achieve expected learning outcomes. In conclusion, cognitive diagnostic assessment is still a potential tool to achieve the purpose of assessment for learning despite all the mentioned limitations and constraints which are resolvable. Future studies may focus on how teachers can use this approach to develop formative assessment in the classroom as well as how students can use the results of such assessments.

Acknowledgements The study reported in this paper was made possible by the generous support from the Fundamental Research Grant Scheme (FRGS) of the Malaysian Ministry of Education and Universiti Sains Malaysia (Account No.: 203/PGURU/6711344).

References

Alves, C. B. (2012). *Making diagnostic inferences about student performance on the Alberta Education Diagnostic Mathematics Project: An application of the attribute hierarchy method* (Unpublished doctoral dissertation). University of Alberta, Edmonton, Alberta, Canada.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Available at http://www.assessment-reform-group.org/CIE3.PDF.

Birenbaum, M., & Tatsuoka, K. K. (1993). Applying an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied Measurement in Education, 6*(4), 255–268.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5,* 7–74.

Broaddus, A. (2011). *Modeling student understanding of foundational concepts related to slope: An application of the attribute hierarchy method*. Unpublished manuscript. Retrieved from https://aai.ku.edu/sites/cete.ku.edu/files/docs/Presentations/2012_04_Broaddus%20Modeling%20Student%20Understanding.pdf.

Buck, G., & Tatsuoka, K. K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing, 15,* 119–157.

Burny, E., Valcke, M., & Desoete, A. (2009). Towards an agenda for studying learning and instruction focusing on time-related competences in children. *Educational Studies, 35*(5), 481–492.

Gierl, M. J., & Alves, C. B. (2010). *Using principled test design to develop and evaluate a diagnostic mathematics assessment in grades 3 and 6*. Paper presented at the annual meeting of the American Educational Research Association, CO, USA.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36,* 923–936.

Hartz, S., Roussos, L., & Stout, W. (2002). *Skill diagnosis: Theory and practice* [Computer software user manual for Arpeggio software]. Princeton, NJ: Educational Testing Service.

Harris, S. (2008). It's about time: Difficulties in developing time concepts. *Australian Primary Mathematics Classroom, 13*(1), 28–31.

Hubner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*(3), 1–7.

Jang, E. E. (2009). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for fusion model application to language assessment. *Language Testing LANG TEST* 01/2009*, 26*(1), 031–073. https://doi.org/10.1177/02655322080973.

Kelly, M. K., Miller, K. F., Fang, G., & Feng, G. (1999). When days are numbered: Calendar structure and the development of calendar processing in English and Chinese. *Journal of Experimental Child Psychology, 73,* 289–314.

Ketterlin-Geller, L., & Yovanoff, P. (2009). Cognitive diagnostic assessment in mathematics to support instructional decision making. *Practical Assessment, Research, & Evaluation, 14*(16), 1–11.

Lee, Y. W., & Sawaki, Y. (2009). Cognitive diagnostic approaches to language assessment: An overview. *Language Assessment Quarterly, 6*(3), 172–189. https://doi.org/10.1080/15434300902985108.

Leighton, J. P., & Gierl, M. J. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge, UK: Cambridge University Press.

Linn, R. L. (1986). Educational testing and assessment: Research needs and policy issues. *American Psychologist, 41,* 1153–1160.

McGuire, L. (2007). Time after time: What is so tricky about time? *Australian Primary Mathematics Classroom, 12*(2), 30–32.

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement, 21,* 215–237.

Mislevy, R. J. (1996). Test theory conceived. *Journal of Educational Measurement, 33,* 379–416.

Nichols, P. (1994). A framework of developing cognitively diagnostic assessments. *Review of Educational Research, 64,* 575–603.

Roberts, M., Alves, C., Gotzmann, A., & Gierl, M. J. (2009). *Development of cognitive models in mathematics to promote diagnostic inferences about student performance.* Paper presented at the 2009 annual meeting of the American Educational Research Association, San Diego, CA.

Russell, M., O'Dwyer, L., & Miranda, H. (2009). Diagnosing students' misconceptions in algebra: Results from an experimental pilot study. *Behavior Research Methods, 41*(2), 414–424.

Sinharay, S., & Almond, R. G. (2007). Assessing fit of cognitive diagnostic models: A case study. *Educational and Psychological Measurement, 67,* 239–257.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263–331). New York, NY: Macmillan.

Wang, C., & Gierl, M. J. (2011). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills in critical reading. *Journal of Educational Measurement, 48*(2), 165–187.

Wu, L. J., Chen, H. H., Sung, Y. T., & Chang, K. E. (2012). Developing cognitive diagnostic assessments system for mathematics learning. *Proceedings of the 12th IEEE International Conference on Advanced Learning Technologies, ICALT 2012*, 228–229.

Ye, F. (2005). *Diagnostic assessment of urban middle school student learning of pre-algebra patterns* (Unpublished doctoral dissertation). The Ohio State University, USA.

Author Biographies

Carolyn Jia Ling Sia is a mathematics teacher at Fairview International School, Penang, Malaysia. She is currently teaching the Middle Years Programme. She completed her Masters' study in Mathematics Education from the Universiti Sains Malaysia while working as a research assistant of Prof. Chap Sam Lim for the project "Developing a Two-Prongs Cognitive Diagnostic Assessment [CDA] for Primary Mathematics Learning." She has attended local and international conferences to share the findings from the project.

Chap Sam Lim is a Professor in mathematics education at the Universiti Sains Malaysia, Penang, Malaysia. She obtained a Ph.D. in Mathematics Education from the University of Exeter, United Kingdom. She currently teaches mathematics education research courses and qualitative research methods courses at the graduate level. Her research interests focus on public images of mathematics, teaching mathematics in second language, cross-cultural study and particularly Lesson Study as a professional development for mathematics teachers. She is an active collaborator on several international research projects, especially on Lesson Study and cross-cultural comparative studies of mathematics teaching and learning in schools.

Chapter 10
Validating and Vertically Equating Problem-Solving Measures

Jonathan D. Bostic and Toni A. Sondergeld

Abstract This paper examines the validation of a measure for eighth-grade students related to problem-solving. Prior work discussed validity evidence for the Problem-Solving Measure series, but it is uncertain whether linking items appropriately vertically equates the seventh- and eighth-grade measures. This research connects prior work to the development of linked measures with anchor items that assess problem solving within the frame of the Common Core in the United States. Results from Rasch modeling indicated that the items and overall measure functioned well, and all anchor items between assessments worked satisfactorily. Our conclusion is that performance on the eighth-grade measure can be linked with performance on the seventh-grade measure.

Keywords Assessment · Evaluation · Measurement · Middle-grades problem solving · Rasch Modeling

10.1 Introduction

Assessments should address the depth and focus of instructional standards (Wiliam 2011). Classroom assessments provide opportunities to promote learning and give teachers data about what and how students are learning (Black et al. 2004). Since 2009, a majority of states within the United States of America have adopted the Common Core State Standards for Mathematics (Common Core). The Common Core shares similarities with mathematics standards implemented around the world

J. D. Bostic (✉)
Bowling Green State University, 529 Education Building, Bowling Green, OH 43403, USA
e-mail: bosticj@bgsu.edu

T. A. Sondergeld
School of Education, Drexel University, 3401 Market St., Office #332, Philadelphia, PA 19104, USA
e-mail: tas365@drexel.edu

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*, ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_10

(e.g., Australia, Japan, and Germany) with respect to depth and coherence, specifically the way that problem solving is woven throughout the standards across grade levels (Commonwealth of Australia 2009; Mullis et al. 2016; Takahashi et al. 2009). The Common Core has a clear focus on problem solving (Common Core State Standards Initiative [CCSSI] 2010) and has two equally important components: content and practice standards. The Standards for Mathematics Content describe what students should learn in each grade level. The Standards for Mathematical Practice communicate behaviors and habits students should experience while learning mathematics in classroom contexts. The Standards for Mathematical Practice are primarily developed from the notions of mathematical proficiency (Kilpatrick et al. 2001) and the National Council of Teachers of Mathematics process standards (2000) and are shown in Table 10.1. Some researchers have started to identify what the Standards for Mathematical Practice look like in practice; the Revised Standards for Mathematical Practice Look-for Protocol (Bostic and Matney 2016) and Mathematics Classroom Observation Protocol for Practices (Gleason et al. 2017) are two published, evidence-based observation protocols with sufficient validity evidence grounding their use by educators and researchers.

If teachers are expected to engage students in problem solving during regular mathematics instruction, then a discussion on how students' problem-solving performance can be assessed in a valid and reliable manner alongside the new content and practice standards must be held. From a review of literature using multiple scholarly search engines (i.e., EBSCO, Google Scholar, and Science Direct), Bostic and Sondergeld (2015b) reported that the Common Core, much less state-level standards from the previous era, have not been integrated into published problem-solving measures. Since that time, there have been no additional reported problem-solving measures drawing upon the Standards for Mathematics Content or Standards for Mathematical Practice besides the Problem-Solving Measure for grade six (PSM6; see Bostic and Sondergeld 2015a, b). On the PSM6, students are expected to solve problems drawing on content found in the content standards in ways that engage them in the Standards for Mathematical Practice. Teachers,

Table 10.1 Titles of *Standards for Mathematical Practice*

SMP #	Title
1	Make sense of problems and persevere in solving them
2	Reason abstractly and quantitatively
3	Construct viable arguments and critique the reasoning of others
4	Model with mathematics
5	Use appropriate tools strategically
6	Attend to precision
7	Look for and make use of structure
8	Look for regularity in repeated reasoning

From: Common Core State Standards Initiative (2010) educational stakeholders, researchers, and even students need to know to what degree students can do this as evidence of understanding.

This paper attends to the notion that problem solving is complex and requires high-quality assessments that are useful for a wide audience in order to move student learning forward and inform classroom instruction. Our first objective is to briefly describe the psychometric results of a new assessment for eighth-grade students called the Problem-Solving Measure 8 (PSM8), which builds upon past problem-solving measures for seventh- and sixth-grade students (PSM7 and PSM6; Bostic and Sondergeld 2015a, b). Our second objective is to describe the vertical equating process and results with respect to the PSM7 and PSM8. As a result of examining a measure for English-speaking students, readers may learn how to develop similar measures for use with other students using the PSMs as a model. The development of the Problem-Solving Measures began by examining a problem-solving measure used with Dutch-speaking students, translating it from Dutch to American English, then starting the assessment development process (see Bostic and Sondergeld 2015a, b for more information). Moreover, this paper illustrates an example of vertical equating within a mathematics education context, which may foster conversations about developing sound measures using modern measurement approaches.

10.2 Related Literature

10.2.1 Synthesizing Literature on Problems and Problem Solving

For these measures, we characterized problem-solving as a process including "several iterative cycles of expressing, testing and revising mathematical interpretations—and of sorting out, integrating, modifying, revising, or refining clusters of mathematical concepts from various topics within and beyond mathematics" (Lesh and Zawojewski 2007, p. 782). Problem solving occurs only when learners work on a problem. Schoenfeld (2011) frames a problem as a task such that (a) it is unknown whether a solution exists, (b) the solution pathway is not readily determined, and (c) more than one solution pathway is possible. Problems are unique from exercises. Exercises are tasks intended to promote efficiency with a known procedure (Kilpatrick et al. 2001).

Researchers across multiple countries have argued for students to experience non-routine word problems as part of their typical mathematics instruction (Boaler and Staples 2008; Bostic et al. 2016; Matney et al. 2013; Palm 2006; Verschaffel et al. 1999). These word problems should be complex, open, and realistic (Bostic et al. 2016; Matney et al. 2013; Palm 2006; Verschaffel et al. 1999). *Complex* problems require reasoning and persistence because a solution or solution pathway is not clear. *Open* problems allow multiple viable problem-solving strategies and offer several entry points into the task. *Realistic* problems encourage problem solvers to draw on their experiential knowledge and connect mathematics in and out of the classroom. Given frames for problem solving and problems, coupled with a need for valid, reliable problem-solving assessments, we developed a series of classroom assessments to measure students' problem-solving performance that allow students' performances to be linked over time. We build upon prior published work on developing classroom based assessments (Bostic and Sondergeld 2015a, b) and share findings about the validity and reliability of the PSM8. Rasch modeling (also known as one-parameter item-response theory) and vertical equating were employed while examining the PSM8 within the broad context of its two-related measures, the PSM7 and PSM6.

10.2.2 Rasch Modeling

Assessments for learning should provide meaningful evidence to students and teachers about future directions for classroom instruction (Wiliam 2011). The Problem-Solving Measures were created with two goals in mind. The first goal is to inform students, teachers, parents, school personnel, and research communities about what abilities students bring to problem solving. The second goal is to support instructional implications based upon the results. To do this, Rasch modeling (1980) was used in the Problem-Solving Measure instrument development process. Rasch modeling is a suite of modern measurement methods considered more appropriate for use in cognitive and affective assessment development and validation processes than Classical Test Theory methods (Bond and Fox 2007). This is largely because of Rasch's strict adherence to unidimensional measurement, its ability to produce a conjoint measurement scale which allows for criterion-referenced interpretations, the probabilistic nature of the models, and Rasch's generation of more precise person and item measures.

All measurement models specify that the construct under study must be unidimensional. This means that an instrument can only measure one latent trait or variable at a time. Using our case as an example, it was important to make sure that the Problem-Solving Measures were measuring only problem-solving ability and not reading ability or some other construct simultaneously or instead. While in theory all measurement models require unidimensionality, Rasch is the only measurement model that has specific fit statistics designed to assess the meeting of this assumption (i.e., point biserial correlations, infit/outfit mean-square fit statistics, and infit/outfit z-statistics for individual people and items). See Linacre (2002) or Bond and Fox (2007) for a more detailed explanation of Rasch fit statistics, and Smith (1996) for a more expansive discussion on the notion of unidimensionality in measurement theory.

Criterion-referenced interpretations of results are made possible with Rasch modeling because of the conjoint measurement scale produced. Conjoint measurement means items and persons are placed on the same scale or ruler. The ruler produced is made up of Logits (Log Odds Units) and show test-taker performance (abilities) in direct comparison to item measures (difficulties). Students are lined up from least able to most able, and items are on the opposite side of the ruler going from least difficult to most difficult to answer correctly. When students are compared to items on the same scale, then it is possible to determine the content students have mastered as well as areas to further master because Rasch modeling allows individuals to be measured against the construct (i.e., criteria or items) rather than a normed sample as done with Classical Test Theory. Moreover, multiple populations can be compared to one another in relation to the content because results are not norm-referenced.

The probabilistic nature of Rasch modeled measures are also likely to offer more accurate estimates of problem-solvers' abilities (De Ayala 2009; Embretson and Reise 2000). For example, suppose two students—Connor and Josephine—both correctly answer seven questions on a 10-item test. Connor answers the seven most difficult items correctly, and Josephine answers the seven easiest items correctly. With Classical Test Theory, Connor and Josephine both earn a 70% because item difficulty is not considered, and each item is typically given the same value (e.g., one point). However, Rasch measurement's use of a conjoint model that places students and items on the same ruler allows for students' scores to be expressed more accurately. Instead of both students earning the same score, with Rasch-modeled measures Connor would earn a higher score than Josephine because item difficulty is taken into consideration when generating student ability measures.

Rasch's probabilistic nature using a conjoint item/person measure also makes it so missing data are not problematic (Bond and Fox 2007). A test taker could skip items they are unsure of or do not have enough time to complete, and Rasch modeling is still able to reliably estimate an accurate person-ability measure from data collected. Rasch modeling does this by relating correct item responses to their item difficulty. A final way that Rasch modeling produces more accurate measures of student abilities is related to standard errors. With Classical Test Theory there is one standard error applied to all test takers, regardless of their ability. In contrast, Rasch modeling produces individual estimated standard errors for each test completer, eliciting more precise participant ability measures (Embretson and Reise 2000).

10.2.3 Vertical Equating

Many tests are limited to providing scores aligned with a single grade level, or inappropriately compare test results from year-to-year to assess student growth because tests are not appropriately linked. If assessing the same construct at multiple grade levels, then a single ruler of the construct can be developed to produce test results that show reliable and valid growth measures over time for individual students or compare a test taker to different levels of the same latent variable (e.g., problem solving). Vertical equating (or scaling) with Rasch modeling can be performed for this purpose when exploring a single, unidimensional construct (Lissitz and Huyunh 2003; Wright and Stone 1979). To do this, each test is developed with its own set of items aligned to grade specific content. A small set of common items (anchor items) are also placed on two consecutive measures to connect the measures and allow for one measuring stick across grade levels to be employed (Wright and Stone 1979). Typically, anchor items measure at a moderately difficult level on the lower grade level assessment, and are then evaluated for appropriateness on the higher grade-level assessment. If the anchor items fall within ±.40 logits on the higher-level test (displacement), they are considered suitable anchors to link different grade-level tests (Kenyon et al. 2006). A second suggestion for anchor items is that they address fundamental content within a set of standards (Wright and Stone 1979). This suggestion helps to maintain fidelity of measures between two sets of standards so long as there is a common theme cutting across the standards. There is no required number of anchor items to vertically equate measures; however, using as few as three high quality items has been shown to be sufficient (Pibal and Cesnik 2011).

In summary, many social science researchers believe Rasch measurement methods are one of the best approaches for assessment development and refinement. This is due to the model's ability to convert ordinal data into equal interval hierarchical measures that conjointly place both item difficulties and person abilities on the same scale for direct comparison (see Bond and Fox 2007). Further, using Rasch measurement for vertical equating of various grade-level tests assessing the same content allows growth between grade specific content to be meaningfully measured.

10.3 Method

This paper describes research building from validation and linking studies of the Problem-Solving Measure 6 (PSM6) and the Problem-Solving Measure 7 (PSM7) and explicitly focuses on the vertical equating process. Readers interested in the prior validation studies should consult Bostic and Sondergeld (2015a, b) or Bostic et al. (2017). The two research questions for the present study are (a) What are the psychometric properties of the Problem-Solving Measure 8 (PSM8)? (b) Is there sufficient psychometric evidence to suggest that scores on the PSM7 and PSM8 are vertically equated successfully? We present necessary evidence supporting two sources of validity evidence needed for measures aiming to generate meaningful scores and useful score interpretations.

10.3.1 Instrumentation

Development process for the PSM7 is shared first to contextualize development of the PSM8. Prior peer-reviewed manuscripts describe the PSM6 (Bostic and Sondergeld 2015b) and linking between PSM6 and PSM7 (Bostic and Sondergeld 2015a). The PSM7 has 19 items, which includes four anchor items from the PSM6. There are three items representing each of the five domains within the seventh-grade content standards of the Common Core (i.e., Ratio and Proportions, Number Sense, Expressions and Equations, Geometry, and Statistics and Probability). Item writing involved intentionally developing low-, moderate-, and high-difficulty items for each content area. It also involved examining key standards that led to the development of algebraic thinking and reasoning, a core component of the Common Core (Smith 2014). Initial development included reflecting on standards within each domain and then gathering data from potential respondents about what constituted realistic contexts.

The second stage of item development led to conducting think-aloud interviews with a small group of potential respondents for response process evidence as well as feedback from mathematics teachers and terminally degreed mathematics educators. Feedback led to item revisions, which led to further interviews and expert panel feedback until there was sufficient validity evidence that the individual items might be put together to form a useful and meaningful measure addressing problem-solving performance related to the Common Core.

An expert panel evaluating the measure consisted of four seventh-grade mathematics teachers, one mathematician holding a Ph.D., and two mathematics educators with terminal degrees. This expert panel reviewed the items for connections with the content found in the Standards for Mathematics Content and mathematical behaviors and habits described by the Standards for Mathematical Practice, developmental appropriateness, use of complex, realistic, and open problems, and considered potential bias within the measure. The survey sent to the expert panel is shared in the Appendix. Results of their reviews communicated that the items adequately addressed the Standards for Mathematics Content and Standards for Mathematical Practice, were grade-level appropriate, and adhered to the ideas that problems should be complex, realistic, and open.

We also wanted a representative sample of seventh- and eighth-grade students to examine items for potential bias and response processes. Teachers provided suggestions for individuals from both grade levels representing a diverse set of genders, ethnicities, and academic abilities. From those suggestions, nine students volunteered to serve on a second panel following measure administration. This group of students represented both grade levels, male and female students, multiple ethnicities, and below-average, average, and above-average academic abilities according to their teachers' perception. During one meeting, students met in groups of two or three students and were asked to share their thinking aloud about each item. They were instructed to describe how they might solve each problem. When every participant on the panel shared their approaches, then the interviewer moved to the next task. These response process interviews were followed by whole-group interviews about bias within the items. Students were asked to express if they felt an item exhibited any bias to a group of students that might cause undue harm or undue influence on their problem-solving abilities. Both groups (i.e., expert panel and student panel) expressed that there was no known bias impacting students' performance.

Data from 654 seventh-grade students located in the United States of America took the PSM7 during a 90-min block. None of the students had limited English proficiency. Responses were scored as correct or incorrect. Rasch modeling for dichotomous responses (Rasch 1980) was employed to examine the psychometric properties.

Development of the PSM8 followed a similar process; three eighth-grade items from each of the five content area domains were included in the measure. Item descriptions for the PSM8 are shown in Table 10.2. Standards come from the Standards for Mathematics Content.

One item from each of the PSM7 and PSM8 are shown in Figs. 10.1 and 10.2, respectively.

Table 10.2 Item descriptions and connections with Standards for Mathematics Content (SMC)

Item #	Item description	Primary SMC	Secondary SMC
1	Acres of land	8.Expressions and Equations.4	
2	John/Eric Run	8.Expressions and Equations.8.A	
3	Festival	8.Expressions and Equations.8.C	
4	Chess board	8.Number Sense.2	
5	Beef Jerky	8.Number Sense.1	
6	Marbles	8.Number Sense.2	
7	Boys/Girls Lunch	8.Statistics and Probability.4	
8	Walleye	8.Statistics and Probability.1	
9	Diapers	8.Statistics and Probability.2	
10	Water Balloons	8.Geometry.9	
11	Map of Town	8.Geometry.8	
12	Capture The Flag	8.Geometry.7	
13	Theme Park	8.Functions.5	
14	Hockey Game Travel	8.Functions.2	
15	Cat Weight	8.Functions.3	8.Functions.4

A water tower contains 16,880 gallons of water. Each day half of the water in the tank is used and not replaced. This process continues for multiple days. How many gallons of water are in the tower at the end of the fourth day?

Fig. 10.1 Water tower item from the PSM7

A chess board is made of eight rows with eight squares in each row. Each square has an area of 3 inches². What is the exact length for one edge of the chess board?

Fig. 10.2 Chess board item from the PSM8

To meet the aim of the present study, we also selected five seventh-grade items that represented core ideas and were moderately difficult for seventh-grade students. One item from each domain (i.e., Ratio and Proportions, Number Sense, Expressions and Equations, Geometry, and Statistics and Probability) was added to the PSM8 to make a total of 20 items. The rationale for item selection was (a) items represent content areas identified as critical areas of seventh-grade mathematics content standards (CCSSI 2010) and (b) support a trajectory to developing deep-rooted proficiency for understanding algebra (Smith 2014). These items were within one standard deviation of mean item difficulty for the PSM7 and respondents' mean; hence, they were moderately difficult items. As an example, an average performing sixth-grade test taker had approximately 50% likelihood of correctly answering the anchor items placed on PSM7. Thus, the items were appropriate within the broader scope of mathematics progressions because the average student should be able to correctly respond to the anchor items related to the previous year's mathematics content.

10.3.2 Data Collection and Analysis

It is expected that a sample of more than 30 respondents representing the target population is needed to participate in a validation study and/or to gather meaningful data about measure functionality when using Rasch modeling (Bond and Fox 2007; Embretson and Reise 2000). We exceed the minimum for the present study. Three hundred eighty-four students from a Midwest state in the US completed the PSM8 in approximately 90 min. Responses were scored as correct or incorrect and then examined using Rasch modeling for dichotomous responses as with previous iterations of the Problem-Solving Measure. Winsteps version 3.74.0 (Linacre 2012) was employed for executing Rasch analysis. Psychometric evaluation of validity evidence, reliability, assessment of anchor item functioning, and displacement were explored similarly to analyses conducted in previous studies of the measures.

10.4 Results

10.4.1 Psychometric Findings

First, we frame our findings within appropriateness for a fundamental quality of measurement—unidimensionality. The results respond to the first research question exploring the psychometric properties of the PSM8. For the purpose of our study, we will only assess item quality and not person-response quality because our focus is on measure creation and not evaluation of person-ability measures. Items with infit mean-square (MNSQ) statistics falling outside 0.5–1.5 logits, outfit MNSQ statistics greater than 2.0 logits, or negative point biserial correlations detract from high-quality measurements (Linacre 2002). Infit statistics provide information about unexpected patterns of responses that are closer to the mean. Outfit statistics give information about whether items have unexpected patterns of responses in outlier items. Point biserial correlations indicate whether an item is measuring the construct under investigation.

No item on the PSM8 had a negative point biserial correlation. Further, all infit and outfit statistics were appropriate. Item reliability was .95, indicating strong internal consistency. Item separation was high (4.16) suggesting that five distinct groups of items can be separated along the variable (minimum separation of 2.0 is acceptable). Item difficulties ranged from −2.87 to 3.11 logits. Synthesizing these results, we concluded that the items functioned well together to form a unidimensional construct assessing a broad range of problem-solving abilities within eighth-grade students.

10.4.2 Vertical Equating Findings

Related to our second research question examining evidence suggesting that scores on the PSM7 and PSM8 are vertically equated, four of the five anchor items from the PSM7 for the PSM8 functioned well as they were within the appropriate range for displacement (±.40 logits). We concluded that four of the five items should remain on the PSM8 as a means to link the PSM8 with the PSM7. Thus, the PSM8 was revised and future administrations of this assessment were as a 19-item test instead of 20 items. A variable map of the PSM8 with the anchoring items is provided in Fig. 10.3 showing items ordered on the right side of the map from easiest at the bottom to more challenging at the top. Students are on the left side of the map ordered from less able on the bottom to more able on the top. PSM7 items kept as anchors on the PSM8 are boxed on the map.

10.5 Discussion

The present study is grounded within the context of other research on the Problem-Solving Measures. Initially, the PSM6 was developed and validated for use as a classroom assessment useful for researchers, administrators, teachers, and students (Bostic and Sondergeld 2015b). A logical next step was to create a Problem-Solving Measure for use with seventh-grade mathematics students and vertically equate the PSM6 and PSM7. For a complete discussion of validating the

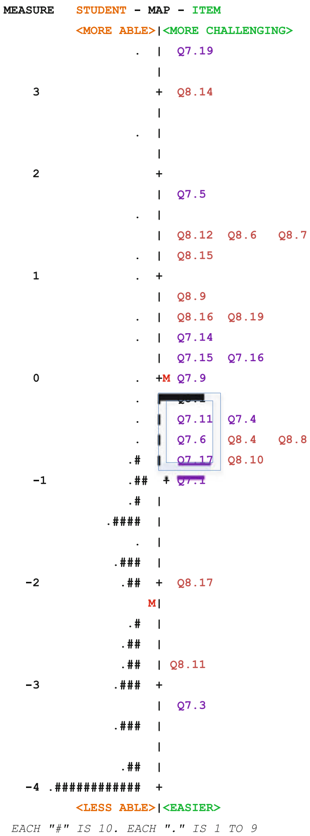


Fig. 10.3 Variable map for PSM8

PSM7 and linking it with the PSM6, see Bostic and Sondergeld (2015a). In summary, there was sufficient evidence along five validity sources for the PSM8 and four items link the PSM7 with the PSM8. The present study provides the next step in developing a suite of middle grades problem-solving assessments: communicating psychometric validity evidence for the PSM8 and linking it with the PSM7. Thus, these studies collectively provide sufficient grounding for using the Problem-Solving Measure series in grades six, seven, and eight for assessing students' problem-solving performance.

Standards and assessments should align (Wiliam 2011) and provide meaningful evidence to test administrators and respondents (Black et al. 2004). Vertically equating is often discussed but rarely performed in practice (Lissitz and Huynh 2003). Its use on the Problem-Solving Measures with linking items provides meaningful formative assessment data including mathematical understanding of standards that are part of classroom instruction as well as connections to mathematical practices (CCSSI 2010). The Problem-Solving Measures meet Wiliam's (2011) call for classroom assessments that align with content and practice standards that students learn.

Results from this study also provide researchers examining mathematics education constructs a means to gather data about students' problem-solving abilities. Problem solving and mathematics should not be considered as two separate fields of mathematics learning (National Council of Teachers of Mathematics 2000, 2014); hence, we frame the Problem-Solving Measures as not only a tool to examine mathematical problem-solving abilities, but also an instrument to gather data about students' mathematics learning. We caution use of the series as a sole determinant of mathematical understanding, because it does not necessarily provide evidence of students' knowledge related to all standards that should be taught during the academic year. Moreover, it is not a high-stakes test and should not be used to make high-stakes test decisions. Instead, it should be used as a formative assessment tool to capture information about students' growth during an academic year and/or growth across multiple years.

Considering the international audience of this monograph, we advocate for following best practices for creating measures and validating them (American Educational Research Association [AERA] et al. 2014). The Problem-Solving Measures were modeled after a problem-solving instrument used by Verschaffel et al. (1999) for a single grade level. Adhering to the validation process described by Bostic and Sondergeld (2015b) was essential before we considered vertical equating the PSM7 and PSM8. Thus, scholars working with respondents who are not fluent English speakers should not implement the Problem-Solving Measures series before conducting a small validation study. It may be necessary to revise contexts and/or language before implementing it on a large scale in such situations. Researchers working with English-speaking respondents should feel confident that they are using an instrument that measures performance on grade-level items as well as material found in the previous grade-level and is grounded in robust validity evidence. A second facet is that different standards from different countries may not necessarily align with the Common Core. Thus, researchers intending to use this measure in English-speaking countries outside of the US should explore the content connections between the Common Core and the standards in their country before administering the measure. Otherwise, there is potential for a misuse of the PSM7 and PSM8 because the validity evidence is grounded with a specific set of content and practice standards consistent with policies found in most states within the US.

10.6 Limitations

Like any study, the present work has limitations. The first limitation is that the PSM7 and PSM8 address all of the domains found on the Common Core but do not represent every standard within a particular domain. A measure containing an item whose primary standard is addressed by one and only one item would require a test with more than 25 items and likely take an entire week of school to complete. Hence, the Problem-Solving Measure in general broadly addresses the mathematics content found in grade-level standards.

A second limitation is that, while results from Rasch modeling are sample-independent, it may be possible to have slight fluctuations in results (usually within a standard error) due to a sample. We met evaluation standards (AERA et al. 2014) for measure development with our sample, but it is possible that a different respondent sample for this study might produce slightly different results. However, because we met the requirements for measure development (AERA et al. 2014), results from different samples should produce similar statistics because of the sample-invariant nature of Rasch modeling.

10.7 Future Directions for Research on the Problem-Solving Measure Series

Three directions for research on and with the Problem-Solving Measure series are shared here. Data from students in 65 teachers' classrooms are currently being analyzed to explore students' growth during one academic year using the series. Teachers in grades six, seven, and eight administered the measures at the beginning of an academic year (first two weeks of school) and again near the end of it (within the last month of school). Authors of the present study are leaders of this project, and will extend data collection into a second year, thus collecting two years of data on students' problem-solving performance. Goals with this project are three-fold. The first goal is to examine changes in students' problem-solving performance related to a set of mathematics standards during one academic year. Respondents are unlikely to recall items that are verbally dense when there is sufficient time between test administrations and the thinking required to solve the items requires critical, complex thinking. A rough guideline for problem-solving items like those on the PSM7 and PSM8 is to allow more than two weeks for respondents to forget their earlier responses, longer durations between test administrations significantly decrease likelihoods for retaining this information in working or short-term memory (Schacter 1999). Thus, there is little concern that at the end of the academic year students might respond identically to how they did at the beginning of the year.

The second goal is to examine students' growth across years using the series. Because the PSM7 and PSM8 are vertically equated, growth is measured along one meter stick. Respondents' performance is measured and can be discussed easily in terms of year-to-year growth, which is unlike tests without anchor items. In the latter case, respondents' growth can be described on a single measure but not necessarily across measures (or grade-levels).

A third goal is to illustrate how researchers might successfully document growth across one year as well as multiple years at-scale. There is a need for more longitudinal studies of problem-solvers' performance; at the present time, there are few such reports, and hence, little is known about how students' development as problem solvers is related to the content they learn in classroom contexts. The Problem-Solving Measures provides a means to capture those data effectively and efficiently. Moreover, score interpretations and uses are deeply grounded with validity evidence (see Bostic and Sondergeld 2015b; Bostic et al. 2017).

Extending the Problem-Solving Measure assessments into elementary grade levels (e.g., grades three, four, and five) is a second direction for future research. Such an extension might provide students, teachers, parents, and researchers with useful information about changes in students' problem-solving performance over time and across grade-levels. This extension may involve a closer examination of testing students of these ages across two days. For instance, a 90-min test might require students to have two or three days (e.g., 45 and 30-min blocks) for testing.

A third viable direction for research has two different but related pursuits. In every instance, a Problem-Solving Measure has been administered without a calculator and as a paper-and-pencil measure. Problem solving may focus on appropriate strategy-use for an open, realistic, complex task. Arriving at a correct result is important but not necessarily more important than employing an appropriate mathematical strategy to the task. At times, problem solvers express incorrect answers to problem-solving tasks because of calculation errors that might or might not be remedied by the use of a calculator. To that end, a study of students' responses with and without a calculator might offer some idea about the role of a calculator for the assessments in the series. The second direction involves an online format for test administration. Students' responses to the Problem-Solving Measure are scored as correct or incorrect. If each answer has a finite number of correct responses, then could it be delivered as an online instrument and still maintain its rigorous validity evidence? Again, this question requires follow-up in its own research study.

10.8 Conclusions

The PSM8 had reasonable psychometric properties and results supported linking the PSM8 with the PSM7. These results, in combination with our past research, suggest we have created a series of classroom-based measures for grades six, seven, and eight. Such measures reflect the complexity of assessing mathematical problem solving within the context of grade-level content standards. Linked measures also respond to a call for valid and reliable assessments that might be useful across a wide audience. Teachers in the United States have already used results from the series to inform instruction and modify classroom practices (Folger and Bostic 2015). Additional teachers and researchers might consider the series as a foundation for assessments that can be employed to derive meaningful feedback for learning about students' problem-solving abilities and informing instructional practices.

Appendix

Is the task a problem? (YES or NO).
Is the task open? (YES or NO).
Is the task realistic? (YES or NO).
What seventh- or eighth-grade Common Core Standard(s) for Mathematics Content are addressed by this task? Please list the primary standard and/or secondary standard, if applicable.
(NOTE: The primary standard is one that is best addressed by this task. The task may also address a secondary standard. Finally, the task may address a tertiary standard that is somewhat connected but not necessarily the clearest standard. For example, a student could solve the problem in a certain manner that employs knowledge from the tertiary standard but is not the most likely problem-solving approach that will be used.)
What Standard(s) for Mathematical Practice primarily are addressed by this task?
What bias, if any, do you perceive within this task that might inappropriately influence a respondent's ability to answer this question?
FOR MATHEMATICIAN ONLY
Please show at least two viable approaches to solve this task.
Is the mathematics correct?
Is there a well-defined solution?

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 9–21.

Boaler, J., & Staples, M. (2008). Creating mathematical future through an equitable teaching approach: The case of railside school. *Teachers College Record, 110,* 608–645.

Bond, T., & Fox, C. (2007). *Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Bostic, J., & Matney, G. (2016). Leveraging modeling with mathematics-focused instruction to promote other standards for mathematical practice. *Journal of Mathematics Education Leadership, 17*(2), 21–33.

Bostic, J., Pape, S., & Jacobbe, T. (2016). Encouraging sixth-grade students' problem-solving performance by teaching through problem solving. *Investigations in Mathematics Learning, 8*(3), 30–58.

Bostic, J., & Sondergeld, T. (2015a). Development of vertically equated problem-solving measures. In T. Bartell, K. Bieda, R. Putnam, K. Bradfield, & H. Dominguez (Eds.), *Proceedings of the 37th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 395–398). East Lansing, MI: Michigan State University. Retrieved from http://www.pmena.org/pmenaproceedings/PMENA%2037%202015%20Proceedings.pdf.

Bostic, J., & Sondergeld, T. (2015b). Measuring sixth-grade students' problem-solving: Validating an instrument addressing the mathematics common core. *School Science and Mathematics Journal, 115*(6), 281–291.

Bostic, J., Sondergeld, T., Folger, T., & Kruse, L. (2017). PSM7 and PSM8: Validating two problem-solving measures. *Journal of Applied Measurement, 18*(2), 151–162.

Common Core State Standards Initiative. (2010). *Common core standards for mathematics.* Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.

Commonwealth of Australia. (2009). *Shape of the Australian curriculum: Mathematics.* Retrieved from http://docs.acara.edu.au/resources/Australian_Curriculum_-_Maths.pdf.

De Ayala, R. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Folger, T., & Bostic, J. (2015). Using the PSM6 to adjust math instruction. *School Science and Mathematics Journal, 115*(6). Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/ssm.12130/abstract.

Gleason, J., Livers, S., & Zelkowski, J. (2017). Mathematics Classroom Observation Protocol for Practices (MCOP^2): A validation study. *Investigations in Mathematics Learning*. Advance online publication: http://dx.doi.org/10.1080/19477503.2017.1308697.

Kenyon, D. M., MacGregor, D., Ryu, J. R., Cho, B., & Louguit, M. (2006). *Annual technical report for ACCESS for ELLs English language proficiency test, Series 100, 2004–2005 Administration.* WIDA Consortium. Retrieved from https://www.wida.us/get.aspx?id=142.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

Lesh, R., & Zawojewski, J. (2007). Problem-solving and modeling. In F. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763–804). Charlotte, NC: Information Age Publishing.

Linacre, M. J. (2002). What do infit and outfit, mean-square, and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

Linacre, M. J. (2012). *WINSTEPS Rasch measurement computer program*. Chicago, IL: MESA Press.

Lissitz, R. W., & Huynh, H. (2003). Vertical equating for state assessments: Issues and solutions in determination of adequate yearly progress and school accountability. *Practical Assessment, Research & Evaluation, 8*(10). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=10.

Matney, G., Jackson, J., & Bostic, J. (2013). Connecting instruction, minute contextual experiences, and a realistic assessment of proportional reasoning. *Investigations in Mathematics Learning, 6,* 41–68.

Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (Eds.). (2016). *TIMSS 2015 encyclopedia: Education policy and curriculum in mathematics and science.* Retrieved from: http://timssandpirls.bc.edu/timss2015/encyclopedia/.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2014). *Principles to action: Ensuring mathematical success for all*. Reston, VA: Author.

Palm, T. (2006). Word problems as simulations of real-world situation: A proposed framework. *For the Learning of Mathematics, 26,* 42–47.

Pibal, F., & Cesnik, H. S. (2011). Evaluating the quality-quantity trade-off in the selection of anchor items: A vertical scaling approach. *Practical Assessment, Research & Evaluation, 16*(6). Retrieved from pareonline.net/pdf/v16n6.pdf.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Denmarks Paedagoiske Institut.

Schacter, D. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist, 54*(3), 182–203.

Schoenfeld, A. (2011). *How we think: A theory of goal-oriented decision making and its educational applications*. New York, NY: Routledge.

Smith, R. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling, 3,* 25–40.
Smith, T. (2014, September). *Curricular alignment to support student success in algebra 1.* (Research Report). Retrieved from United States Department of Education website: http://www2.ed.gov/programs/dropout/instructionalpractices092414.pdf.
Takahashi, A., Watanabe, T., & Yoshida, M. (2009). *English translation of the Japanese mathematics curricula in the course of study*. Madison, NJ: Global Education Resources.
Verschaffel, L., De Corte, E., Lasure, S., Van Vaerenbergh, G., Bogaerts, H., & Ratinckx, E. (1999). Learning to solve mathematical application problems: A design experiment with fifth graders. *Mathematical Thinking and Learning, 1,* 195–229.
Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37,* 3–14.
Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: Mesa Press.

Author Biographies

Jonathan D. Bostic is an associate professor of mathematics education in the school of Teaching and Learning at Bowling Green State University in Bowling Green, Ohio (USA). He has two primary research foci. The first is exploring mathematics instruction that promotes problem solving and more broadly, mathematical proficiency. The second research focus is developing and investigating measures in mathematics education that produce quantitative data. Individuals interested in using the Problem Solving Measures are encouraged to contact Jonathan or Toni Sondergeld for more information.

Toni A. Sondergeld is an associate professor of assessment, research, and statistics in the School of Education at Drexel University in Philadelphia, Pennsylvania (USA). Her research interests include assessment development and use (cognitive and affective), program and grant evaluation, STEM education, and urban education.

Part V
Engaging Teachers in Formative Assessment

Chapter 11
French Primary Teachers' Assessment Practices: Nature and Complexity of Assessment Tasks

Nathalie Sayac

Abstract Lack of regulation regarding grading and autonomy granted to teachers in the classroom characterize evaluation in France. Therefore, we know little about the assessment practices of French teachers. This paper reports results of a study designed to explore the assessment practices of French primary school teachers in mathematics with a didactic approach. Using a didactic tool to analyse assessment tasks given to students and studying classroom assessment practices of 25 primary school teachers, I found that, even if the tasks proposed in assessments are different, most are connected to low-levels of complexity and competencies. Great diversity is a characteristic of assessment practices among primary school teachers in France concerning grading, design intent, collaboration, etc.

Keywords Assessment · Mathematics · Schoolteacher · Tasks
Didactics

11.1 Introduction

How do French primary school teachers assess their students in mathematics? It is the question I asked myself as a researcher in didactics of mathematics and teacher training. To deal with this issue, we need to focus on the specific activities of teaching and consider the cultural and institutional context within which they occur. This paper answers this call by focusing on the assessment activity of primary school teachers in France to increase awareness and understanding of their assessment practices and ways to enhance those practices. Given the link between assessment and learning and concerns about student achievement, the question of how students are assessed in classrooms is a major concern. In a number of countries, classroom assessment is strongly supported by objectives set out in the

N. Sayac (✉)
Laboratoire de Didactique André Revuz, Université Paris-Est Créteil,
ESPE de Créteil, 45 avenue Jean Zay, 93381 Livry-Gargan, France
e-mail: Nathalie.sayac@u-pec.fr

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_11

curriculum (in terms of competencies) or in education policies; these objectives may or may not be accompanied by guidelines. Classroom assessments should therefore be studied within the context in which they exist or evolve because this may provide more elements for interpretations of observed patterns (Nortvedt et al. 2016).

In the case of France, classroom assessment has been largely left to teachers. However, over the past fifteen years, France has begun to join other countries in moving toward some regulation of assessment and grading. An appendix provided by The Superior Council for the Curriculum recommends considering assessment as a 'toolbox' with resources for both students and teachers. However, there is little information and training as to how to apply these recommendations, so assessment practices appear to be implemented in ways that are individual to the teacher.

In fact, very little is known about the assessment practices of teachers in general, and in particular with respect to assessment in mathematics. Thus, the question arises: how do French teachers really assess their students in mathematics in their day-to-day classroom practice? The study reported here was developed to gain further insight into French assessment practices in mathematics by exploring these practices among primary school teachers.

To study the assessment practices in mathematics classrooms in France, I conducted a study as part of a collaborative research project with practitioners, which allowed a close investigation of the day-to-day reality faced by teachers (Jaworski 2006). In the next section, I present results related to summative assessment in mathematics at school, by studying various data that highlight the kinds of assessment tasks given by primary school teachers, in terms of content and competencies, in a specific mathematical domain.

11.2 Theoretical Background

In this study, I consider the assessment of student learning from a didactic approach. A didactic approach towards assessing student performance focuses on the "relationships between teaching, learning, and content, and relationships between assessment and subject matter construction, in other words between didactic and pedagogy" (Reuter et al. 2013, p. 101). This didactic approach to assessment lies within the context of French didactics, the main elements of which need to be specified. Assessment is not a central subject in French mathematics didactic theories. In the Theory of Didactic Situations (TDS) (Brousseau 1997), students learn by confrontation with fundamental situations and an appropriate didactical contract. Assessment has no place in this leading French theory, which could explain why there are so few studies on assessment in France. In the Anthropological Theory of Didactic (ATD) (Chevallard 1999), assessment is the sixth moment of the 'study' (teaching situation). Chevallard considers that "the evaluative events that may be observed are not merely incidental existing events, a necessary evil that may be ignored, but instead constitute one of the determining aspects of the didactic process that sets and regulates everything: teachers' behavior and students' learning alike"[1] (1986, p. 4). For Roditi (2011), as part of the 'double approach' theory (Robert and Rogalski 2002), assessment is one of the five organizational activities in teaching practices. For my part, I consider assessment as a specific activity that is part of the teaching and learning process that should be studied through different evaluative episodes (both formal and informal) planned by the teacher during the learning process. These evaluative episodes may play either formative or summative purposes. They are influenced by the way that the teacher conducts them and on the didactic contract[2] involved (Brousseau 1988; Schubaeur-Leoni 1991). To study classroom assessment, I chose to focus on the main evaluative episode, the one proposed by primary school teachers in France at the end of a teaching situation (summative). My examination focused on:

- **Teachers' assessment activity**, in terms of complexity of their practices, their professional judgement and their assessment skills. How do they design their tests? Which kind of grading do they use? How do they connect their tests to the other moments of the teaching situation?
- **The content being assessed**, in terms of the nature and the validity of assessment tasks. What can we say about the assessment tasks proposed by teachers in their tests? Is there enough variety and complexity regarding the mathematical content?

In the following, I explain why I focused specifically on these elements.

11.2.1 Teachers' Activity

11.2.1.1 Complexity of Practices

In France, teachers have full freedom in how they assess their students (Nortvedt et al. 2015). Yet, I share the view, held by Black and Wiliam (1998, 2009), Clarke (1996), Nunziati (1990) and Stiggins (1988), that classroom assessment is a complex teaching–learning process that is dependent on various parameters. Further to this, in the double approach theory (Robert and Rogalski 2002), which I use as a framework for analyzing teaching practices in mathematics, teachers' practices are

[1]In French: "*les faits d'évaluation qu'il peut alors y (la classe) observer ne sont pas simplement un existant contingent, un mal nécessaire que l'on pourrait ignorer, mais bien l'un des aspects déterminants du processus didactique qui règle et régule tout à la fois les comportements de l'enseignant comme l'apprentissage des élèves*."

[2]From the Encyclopedia of Mathematics, edited by Lerman (2014): A "*didactical contract*" is an interpretation of the commitments, the expectations, the beliefs, the means, the results, and the penalties envisaged by one of the protagonists of a *didactical situation* (student, teacher, parents, society) for him- or herself and for each of the others, *à propos of the mathematical knowledge being taught* (Brousseau and Otte 1989; Brousseau 1997).

also considered to be complex and dependent on several components (cognitive, evidential, personal, social, and institutional). Thus, the assessment practices of primary school teachers in mathematics, being part of overall teaching practices, are also complex. Although very little is known about these assessment practices, certain findings about classroom assessment practices in other countries may also apply to practices in French classrooms. For instance, like teachers in other countries, French primary school teachers may use a “hodgepodge” of factors when assessing and grading students (McMillan et al. 2002). Teachers’ decision-making in the classroom is influenced by a variety of external factors (e.g., accountability testing, parents’ expectations) and classroom realities (e.g., absenteeism, disruptive behaviour, heterogeneity) (McMillan 2003), which presumably is also true for teachers in France.

11.2.1.2 Professional Judgement

Professional judgement plays a large role in teachers’ assessment practices (e.g., Allal 2013; Barr and Cheong 1995; Laveault 2009; Morgan and Watson 2002; Wyatt-Smith et al. 2010). Professional judgement includes both cognitive processes and social practice (Mottier Lopez and Allal 2008), which is not the same as a ‘mechanical gesture of measurement’ (Wyatt-Smith et al. 2010), but must be considered as a “flexible dynamic process comprised of middle and final judgements” (Tourmen 2009). Allal considers that “teachers’ judgement in assessment is analogous to clinical judgement in the medical professions in that it implies establishing a relationship between the singular (everything the evaluator knows about a particular individual) and the general (formal and tacit professional knowledge, as well as institutional norms and rules) in order to formulate the most appropriate course of action possible” (2013, p. 31). The professional judgement of teachers could be viewed as an act of discernment and as the ability to build an intelligibility of the phenomenon of assessment, while taking into account the epistemic, technical, social, ethical and paradigmatic dimensions of classroom assessment practices (Tessaro 2013). According to Noizet and Caverni (1978), Chevallard and Feldmann view professional judgement as “an act in which we make a judgement about an event, individual or object by referring to one or more criteria, regardless of the criteria or object involved” (1986, p. 24). They also assert that assessment is a message from a teacher to their students that goes beyond the sole objective of assessing learning. In the didactic approach of assessment I propose, the professional judgement in assessment could be considered as a kind of “didactic vigilance” (Charles-Pézard 2010) specifically applied to the assessment activity of teachers allowing them two things: to give a valid verdict (Chevallard 1989) on students’ mathematical knowledge, individually and collectively, from data collected during the different evaluative episodes; and to mutually articulate the different moments of the teaching situation (connecting evaluative episodes to the other moments of the learning process), based on data collected during the different evaluative episodes.

This professional judgement in assessment depends on individual factors, such as beliefs about learning and assessment and also professional and personal experiences with assessment (Brady and Bowd 2005; Di Martino and Zan 2011; Jong and Hodges 2015). Professional judgement in assessment is also related to teachers' mathematical and didactical knowledge as they interpret students' misconceptions and errors (Bennett 2011; Vantourout and Maury 2006).

11.2.1.3 Teachers' Assessment Skills

Classroom assessment requires a great deal of time and effort and is a central task of teachers (Stiggins 1988) but many researchers have shown that teachers may be under-skilled and lack confidence in carrying out assessment (Black and Wiliam 2010; Moss and Brookhart 2014) and in using assessment techniques (Christoforidou et al. 2014). Martinez et al. (2009) have highlighted that teachers use normative perspectives on a widespread basis without understanding the real issues at stake, while Kilday et al. (2011) indicate that a number of teachers are misinterpreting students' results and undervaluing their skills. Furthermore, Stiggins (2007) argues that too much attention is paid to high-stakes standardized tests and not enough to day-to-day assessments. Although France does not engage in high-stakes standardized tests at the primary level, I am convinced that there is a lack of attention to day-by-day assessments, particularly because teachers receive little or no training in student assessment.

11.2.2 Content Being Assessed

11.2.2.1 Nature of Assessment Tasks

Focusing on mathematical tasks to study how students are assessed is a common practice in mathematics education (Clarke 1996; Senk et al. 1997; Van Den Heuvel-Panhuizen 1996) because "the quality and consistency of teachers' informal assessment of their students are also dependent on the quality of the tasks used" (Morgan and Watson 2002, p. 82).

In this study, tasks are analyzed via Activity Theory, which was first developed by researchers who adopted Vygotsky's approach, was subsequently used in professional didactics (*la didactique professionnelle*), and was later linked to a didactical approach to mathematics teaching in the 'double approach' (Robert and Rogalski 2002). This theory involves the notions of task and activity. By task, we refer to the definition proposed by Leontiev (1976, 1984) and developed by Leplat (Leplat 1997; Leplat and Hoc 1983), that is, the "goal to be attained under certain circumstances" (Rogalski 2013, p. 3). The activity is what a subject engages in while performing the task. This includes external actions as well as any inferences, hypotheses, decisions, and actions that the subject decides not to carry out.

In collaboration with Grapin, I developed a tool for analyzing assessment items in mathematics, guided by didactic findings as well as the theory of activity (Sayac and Grapin 2015). This tool, which takes into account the class level, is composed of three factors:

- **A complexity factor related to the wording and the context of the items (FC1)**, and to the way in which the task is presented to students. In this factor, the language level used in the formulation of the task and the nature of the information are considered (text, chart, diagram, etc.). What matters is how students manage to understand the question and what they must do to perform the task. Sometimes, the formulation is not clear or is misconstrued. Context and the provision (or absence) of an operational representation of the task are also considered in this factor. Indeed, even if the effectiveness of real-life problems is far from established (Beswick 2011; Boaler 1993), I agree with Van Den Heuvel-Panhuizen that:

> Their crucial feature is that they be situations that students can imagine, that are 'real' in students' minds. (2005, p. 10)

For example, consider three complexity levels when asking students to decompose a number:

Level 1—Decompose every number as in the example: 4567 = 4000 + 500 + 60 + 7.
Students could easily understand what can be done through the example.
Level 2—Write every number by decomposing it unit-by-unit (thousand, hundred, ten, unit).
This instruction is clear, but students need to interpret the *units* words.
Level 3—Decompose every number by writing its additive decomposition giving the value of every digit.
This instruction is convoluted and more difficult for students to understand.

- **A complexity factor related to the difficulty of the mathematics involved in the tasks (FC2)**, in terms of the scope of the work performed in the didactics of mathematics.
  In this factor, the mathematical difficulty of the tasks is considered and highlighted by didactic findings that include some of the factors used by Henningsen and Stein (1997) to study appropriate or inappropriate mathematical tasks. This factor is directly connected to mathematical knowledge as well as to difficulties encountered by students or identified in didactics research. From this factor, a task could be simple or not, and regarded as more or less easy. For instance, many studies have shown the difficulties of learning and teaching the decimal number system for whole numbers, particularly in the concept of unit (Fosnot and Dolk 2001; Houdement and Chambris 2013; Mounier 2010). Tempier (2016) has also shown that, in the case of numbers larger than one hundred, complexity is partly due to the multitude of possible relationships between units.

For example, if you ask students to write numbers, the mathematical difficulty depends on various didactic components.

Level 1—5 hundreds, 3 tens, 6 units: the way you write this number is the way you read it. Students should be able to do this without any difficulty.
Level 2—7 hundreds, 8 units: The absence of ten units involves the presence of a zero in the number. This may present a difficulty for the student.
Level 3—8 hundreds, 12 tens, 45 units: students need to compose units before writing the number. This makes the task of writing the number more difficult.

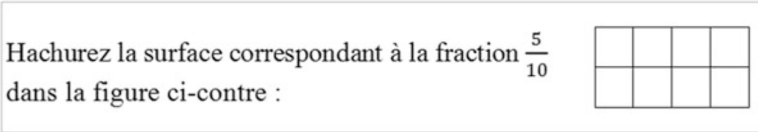
- **A competency factor related to the way in which students must combine knowledge to carry out the task (NC).**
  Mathematical competency is defined as the "capacity to act in an operational way when faced with a mathematical task, which may be unfamiliar, based on knowledge that is autonomously mobilised by the student" (Sayac and Grapin 2015, p. 113). This means that the familiarity with the tasks must be considered despite the difficulty involved in the task. This competency factor is illustrated below by means of three items corresponding to different competency levels of the tool relating to a particular task, that of associating a hatched area with a given fraction.

Level 1: Shade in the area corresponding to the fraction ¼ in the given figure.

This level requires the direct application of knowledge pertaining to fractions in a regular task.
Level 2: Shade in the area corresponding to the fraction ¼ in the given figure.

At this level, students must first divide the figure into four equal parts before shading in the area corresponding to the fraction ¼. It is not exactly a regular task, but students would certainly have encountered similar ones before.

Level 3: Shade in the area corresponding to the fraction 5/10 in the given figure.


Hachurez la surface correspondant à la fraction $\frac{5}{10}$ dans la figure ci-contre :

At this level, the student must convert $\frac{5}{10}$ to $\frac{1}{2}$ before shading in half of the eight parts that constitute the figure. This conversion is left entirely to the student and requires a high level of competency.

To study classroom assessment through a didactic approach, I analyze the knowledge being assessed and the complexity of the tasks used by teachers to assess their students by examining these three factors.

11.2.2.2 Validity of Classroom Assessment

The validity of assessments is an important component of research on classroom assessment. My understanding of validity is that conceptualised by European researchers (De Ketele 1989; De Ketele and Gerard 2005; De Landsheere 1988; Laveault and Grégoire 2002):

> Validity is the adequacy level between what we say we are doing (assessing such and such a dimension) and what we are actually doing, so between what the tool is measuring and what it is supposed to measure. (De Ketele and Gerard 2005, p. 2)

Among the various definitions of validity by these authors, some important components that are consistent with classroom assessment are *curriculum validity* (what is assessed refers to the curriculum) and *pedagogical validity* (what is assessed refers to the mathematical content taught) (De Landsheere 1988). Researchers in didactics also pay attention to the techniques used by students to perform tasks, such as Grapin (2015) who considers that to be valid, the goals defined for tasks need to correspond to the techniques that are expected to be used to accomplish these tasks. With regard to summative assessment, Moss (2013) suggests these conditions for validity: the appropriateness of the assessment tasks in relation to the learning being assessed, the format of the question in relation to the nature of the learning being assessed, and the correspondence between curriculum goals and the learning that occurs in class.

11.3 Methodology

11.3.1 Overall Approach

As a first step towards studying the assessment practices of French primary school teachers, I decided to focus on evaluative episodes proposed by teachers at the end of a teaching situation, because they constitute the main evaluative episodes proposed by teachers in France. To limit the scope of the study, while allowing mathematics assessment to be explored at all primary school levels (grades 1–5), I chose to focus on one specific area of knowledge: whole-number learning assessment (knowledge and writing of numbers).

These choices led me to study, more specifically, the assessment tasks designed by teachers to assess their students. To do that, I collected test(s) given by teachers to assess their students in numeracy and whole numbers as well as exercises and homework on numeracy and whole numbers given to students during the learning process.

Each assessment task in every test was examined in relation to the three factors defined previously, so as to determine the corresponding levels. I also compared these assessment tasks with those given during the teaching situation (exercises and homework) in order to measure the gaps between tasks conducted during this period and tasks conducted during the evaluative episode. In this way, I could explore the validity of the classroom assessment provided by the teachers in my sample. I also collected complementary data allowing me to investigate some of the dimensions of the assessment practices of the primary school teachers:

- *Biographical and professional data* based on a short questionnaire in addition to the other data collection for the study.
- *Marked tests with comments and grades*: successful, less successful and unsuccessful students' tests.
- *Additional information* based on interviews.

After analyzing all the tasks given by the teachers in their test(s), I identified questions to be asked in the interviews pertaining to specific tasks that I wished to explore in greater depth or in terms of the rationale behind the task selection. For example, I asked a teacher to justify the chosen numbers with different writing proposed within the same exercise, I asked another one if she always used such presentations to ask her students to compare numbers. I wanted to know if the tasks were usual and if teachers picked them knowingly. Each interview was transcribed and the teachers' responses were coded so as to enable the identification of what I considered to be the most crucial elements with regard to their approach, namely when they created their assessments (before or after planning their lesson), what resources they used (websites, textbooks, curriculum, etc.), whether they designed assessments with other teachers or alone, how much importance they placed on the competencies defined in the curriculum, what grading system they used (numerical, alphabetical, competency level), and how they utilised assessments (group or individual correction, identification of groups with specific needs, etc.). In doing so, I could explore the teachers' assessment activity of my sample and learn more about their assessment practices.

The marked tests along with interview data provide information about the teachers' professional judgement and their relationship with mathematics (Vantourout and Maury 2006). After gathering all the data, I computed simple statistics to explore the different components of the assessment practices and the nature of the assessment tasks.

11.3.2 Data

The data were collected during the 2014–2015 school year with the help of field educators involved in the collaborative research, who approached colleagues within their school sector (the Seine-et-Marne department). Twenty-five primary school teachers formed our sample. Among these were regular teachers (22), fledgling teachers (3), and teachers working in disadvantaged areas with special educational needs (8). They all agreed to take part in our study because they were interested in the research subject and the potential results. The teachers provided us with the data mentioned above, with each test corresponding to an evaluative episode of summative assessment. I obtained the exercises and homework during the teaching situations themselves by photocopying the exercise book (*cahier du jour*) of one student per class.

Our analytical tool was only applied to the mathematical tasks given in tests as the other data were solely useful for comparing the nature of test tasks with tasks given during teaching situations. A total of 884 tasks were studied, and each teacher gave tests composed of various numbers of tasks, ranging from 16 to 53.

11.4 Results

In this section, the results relating to episodes of summative assessment by French primary school teachers are studied through the aspects investigated: content being assessed and teachers' activity.

11.4.1 Content Being Assessed

As indicated earlier, the content of evaluative episodes at the end of teaching situations concerning numeracy in Grades 1–5 classrooms were analyzed by using the three analytical factors of the above-mentioned tool. This tool provides insight into the levels of complexity and competency associated with assessment tasks based on our sample of primary school teachers in France.

11.4.2 Overall Result

On the whole, the assessment tasks given by the teachers in their tests are characterised by low levels in both complexity factors related to the wording and the context of the tasks (FC1), to the mathematics involved in the tasks (FC2), and the competency factors (NC) as presented in Table 11.1.

Notice that most of the assessment tasks in the sample correspond to the lowest level in the complexity factor related to the wording and the context of the tasks (FC1). That means that, broadly, teachers give simple instructions to their students to assess them. However, there is significant variability among the twenty-five teachers, regardless of grade. For example, level 3 tasks are observed at all grades, from grade 1 to the upper grades. However, observe that more than 75% of the tasks are related to the first level of FC1 and this result is valid regardless of the grade level. It is a reassuring outcome to the extent that students need to understand the assessment task in order to perform it. If a student doesn't understand the assessment task, the teacher cannot assess him and determine his knowledge level of content.

The majority of assessment tasks pertaining to the second factor (FC2), which is directly related to mathematical knowledge, are also connected to low levels. Slightly over two-thirds of assessment tasks correspond to the lowest complexity level, approximately 30% at the intermediate level, while only 2.5% related to the highest level. This finding is not surprising as the vast majority of the teachers stated their belief that assessment must not be so difficult as to penalize students and must be designed fairly. Regarding the nature of the assessment tasks, complexity is often associated with the presence of zeros, the absence of certain units, number length, or the transition to higher-order or lower-order units.

If we distinguish assessment task levels by grade, we observe that teachers from higher grades (cycle 3[3]: grades 3–5) set more difficult assessment tasks in relation to this factor related to mathematical knowledge than teachers from lower grades (cycle 2: grades 1–2) (see Table 11.2).

It seems as though the higher-grade teachers wish to prepare their students for junior secondary school, as the higher-grade teachers are more likely to provide more difficult assessment tasks in relation to mathematical knowledge. This might be regarded as legitimate given the increase in the level of requirements in accordance with the class level, but the allocation of the complexity levels of this factor already takes into account this increase. That means that teachers seem to

[3]In France, school levels are cut in cycles: cycle 1 for kindergarten, cycle 2 for grades 1–2, and cycle 3 for grades 3–5.

Table 11.1 Percent of content analysis of assessment tasks

FC1 (level 1)	FC1 (level 2)	FC1 (level 3)	FC2 (level 1)	FC2 (level 2)	FC2 (level 3)	NC (level 1)	NC (level 2)	NC (level 3)
75.5	21.7	2.7	68.5	29	2.5	92.6	7.4	0

Table 11.2 Percent of tasks at each level of FC2 by grade level

Mathematical knowledge	FC2 = 1	FC2 = 2	FC2 = 3
Grade 1–2 (cycle 2)	85.5	14.5	0
Grade 3–5 (cycle 3)	58.5	37.1	3.9
Average cycle 2 and cycle 3	68.5	29	2.5

expect more of their students only when they are teaching in higher grades at school. Concerning the competency factor (NC), we note that no level 3 was identified in any of the assessment tasks. This finding supports the fact that during the interviews, the teachers said that they did not want their students to fail; therefore, they preferred to wait until they believed that the students were ready before assessing them. Overall, the first competency level is associated with 92.6% of the tasks given in the assessment sessions. Only a small percent (7.4%) of the assessment tasks deal with the second level. Furthermore, the setting of assessment tasks corresponding to competency level 2 does not seem to be linked to grade levels but rather to professional judgement and personal conceptions. I also examined the nature of tasks given by the teachers in tests concerning the domain of numeration. More specifically, I counted the number of tasks related to 3 typical aspects of numeration: ordering of numbers, equivalence of writing registers, and the cardinal number of a collection, distinguishing them by cycle (see Table 11.3).

The hierarchy of the tasks corresponding to the category “Equivalence of writing in the same register or in different registers” is very marked. This nature of tasks represents more than half of the tasks proposed in the whole of the collected tests, but the distribution between cycles is uneven (26% in cycle 2 against 66% in cycle 3). The category “Ordering of numbers” is rather homogeneous in both cycles of primary school and represents a little more than a third of all tasks. In contrast, the category “Cardinality of a collection without operation” is absent from cycle 3, although it corresponds to more than a third of the tasks proposed in cycle 2 tests.

I had considered that the validity of classroom assessments could be ensured by a variety of assessment tasks and through the connection of assessment tasks to tasks given during the teaching process. My results indicate that the validity of the tests given by the teachers to their students may be questioned; although the tasks given during assessment align with the tasks given during teaching situations, the predominance of low-level assessment tasks regarding the three factors has implications for the validity of classroom assessment. If only low-level tasks are included in the tests, how could we consider that students are adequately assessed?

Table 11.3 Number (percent) of numeration tasks by content nature

Nature of tasks	Cycle 2	Cycle 3	Total
Equivalence of writing numbers in the same register or in different registers	83 (26%)	388 (65.6%)	471 (52%)
Ordering of whole numbers	125 (39%)	203 (34.4%)	328 (36%)
Cardinality of a collection without operation	111 (35%)	0	111 (12%)
Total	319	591	910

This predominance of low-level assessment tasks could be explained by the will of primary school teachers to not defeat their students, but teachers also need to know exactly what their students have learned and how students might extend their knowledge to more complex tasks. The validity of classroom assessment could be ensured with a range of simple to complex assessment tasks and not only by matching assessment tasks with tasks proposed during the teaching situation.

11.4.3 Teachers' Activity

11.4.3.1 Assessment Practices

A majority of the teachers in this study designed their tests alone (72%) based on online resources (72%) or mathematics textbooks (68%). Teachers appear to have picked exercises from various sources, which they then compiled to create the tests in accordance with both institutional and professional expectations as well as the exercises given during previous sessions.[4] This supports the idea that classroom assessment usually uses a range of teacher-selected or teacher-made assessments that are closely aligned with what and how the students are learning (Baird et al. 2014).

Furthermore, nearly half of the teachers designed their tests at the end of the teaching situation, which was when they considered that the students were "ready to be assessed". Thirty-six percent of teachers (36%) designed their tests prior to the teaching situation, but they admitted to making adjustments towards the end of the teaching cycle. In France, all primary school teachers must fill in a skills booklet for each student at the end of each term. They have choice as to what data they draw on. Some of them choose to base their information on day-by-day assessments, while others use specific tests. In the interviews, twenty-eight percent (28%) of the teachers mentioned the skills booklet in relation to tests and how to design tests. Of course, this does not imply that the other teachers did not link their tests with the skills booklet. Sixty four percent (64%) of the teachers involved in the study

[4]80% of the teachers gave exercises corresponding 'exactly' to those given during previous sessions.

designed their tests based on the competencies listed in the curriculum and 68% of them used them for grading and filling in skills booklets.

Twenty percent (20%) of the teachers expressed a certain degree of benevolence or encouragement, either during the interviews or through the comments they made when marking tests. For example, one of the marked tests bore the comment 'Well done, you've made progress, keep it up'; during the interviews, several teachers said that they did not wish to 'trick' their students or that they took care not to stigmatise weaker students.

11.4.3.2 Assessment Skills

Only sixteen percent (16%) of the teachers involved in this study had a background in mathematics. This could mean that many teachers may lack mathematical skills or are uncomfortable with mathematics. The analysis of teachers' grading of tasks demonstrated this lack of mathematical knowledge in many instances. For example, in the exercise[5] below (Fig. 11.1) given in Grade 5, a teacher has corrected the student answer concerning the figure C (triangle) by proposing the number 8 for the perimeter (in red). She didn't realise that neither she nor the students could answer this question, at this grade level, as the Pythagorean theorem would be required.

Surprisingly, most of the teachers did not grade tests in collaboration with other teachers, with only eight percent (8%) of them doing so. Nearly a third of the teachers (32%) indicated they only went back to a test if they felt it was necessary, for example when many students failed an exercise. This seems to suggest that teachers only use assessments to pick out students who are struggling, with the intention of helping them by means of specific devices outside of classroom sessions. These findings can be explained by the fact that the focus was on summative evaluative episodes. During the interviews, only four teachers spontaneously discussed formative assessment and two mentioned diagnostic assessment. A likely reason for this is that in France, until recently, there were no institutional recommendations promoting formative assessment in the classroom and there is still no training in formative assessment.

11.4.3.3 Grading Practices

According to a report by the General Inspectorate of the Ministry of Education (July 2013), French teachers use a variety of grading systems, ranging from colour codes (in nursery school), literal appraisals such as 'mastered' (A), 'not mastered' (NA), 'to be reviewed' (AR) and 'in the process of being mastered' (ECA), to marks

[5]Students must find out the area and the perimeter of each figure and put results in a table.

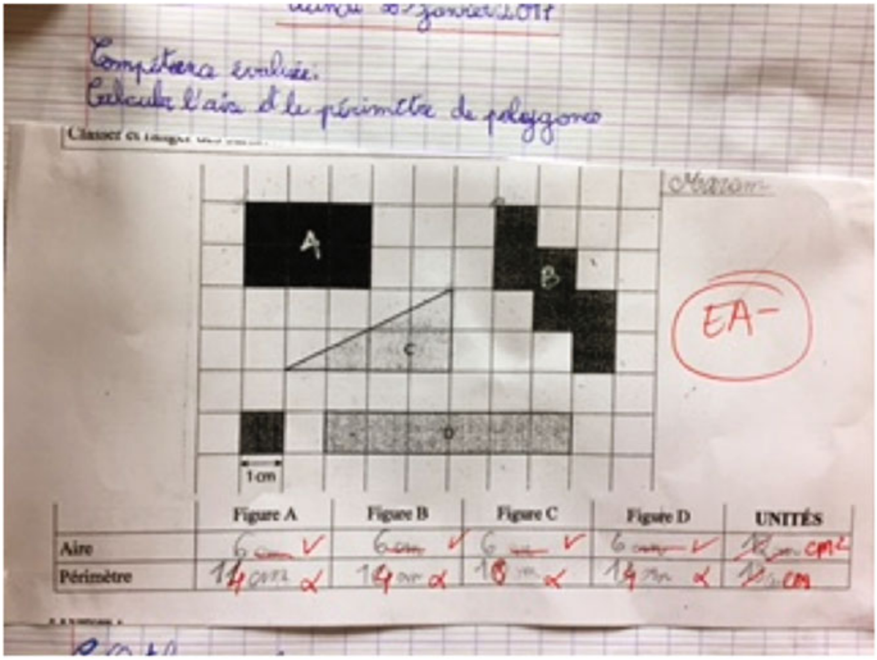


Fig. 11.1 Corrected grade 5 exercise

denoted by numbers (from 0 to 10 or from 0 to 20) or letters (A, B, C, D). The grading system differs greatly among class levels, but there is a marked increase in the use of the number system from grade 4 onwards (p. 3).

In our sample, forty percent (40%) of the teachers used numerical marks (mainly in cycle 3), twenty-four percent (24%) of them used letters, and a quarter carried out appraisals related to competencies; some combined two methods of grading. This study demonstrates that pedagogical freedom granted to French teachers leads them to design their tests and grade their students in a variety of ways, regardless of the class level involved. This finding is consistent with the highly variable results obtained by Cizek et al. (1996) and McMillan and Nash, thereby confirming that "an important characteristic of classroom assessment and grading practices is that they are highly individualised and may be unique from one teacher to another, even in the same school" (McMillan and Nash 2000, p. 31).

11.5 Conclusion

To broaden or deepen our understanding of mathematics teacher practices and effective teacher education as called for by Chapman (2016), it is necessary to study real practices of teachers. This is what I proposed to do with this study investigating specific evaluative episodes of classroom assessment practices of a group of French primary school teachers through the following perspectives:

- What kinds of mathematical assessment tasks, in terms of complexity and competencies, are given to students in French primary schools?
- What can be said about the summative episodes conducted by French primary school teachers? What kinds of resources do they use? Do they collaborate with other teachers to design their tests? To what extent do they follow institutional recommendations and use the competency approach?

The didactic tool used to analyze the assessment tasks provided by the teachers has revealed that, notwithstanding the differences in presentation, context, and formulation among these assessment tasks, practically all of them are connected to low levels of complexity and competency, even if, on average, teachers teaching higher grades designed assessment tasks with higher levels of difficulty. This implies that French students regularly face low-level assessment tasks in summative evaluative episodes; even if these tasks are connected to those tasks studied during the teaching situation, the results challenge the validity of classroom assessment in mathematics. Therefore, if assessment and learning are indeed closely related, one wonders if French students are properly equipped to support high-level mathematical thinking and reasoning (Henningsen and Stein 1997). Another conclusion that can be drawn is that the assessment practices of primary school teachers in France are extremely diverse in terms of such characteristics as grading, design intent, and collaboration, which is also the case in many other countries (Cizek et al. 1996; McMillan 2003; Remesal 2007; Stiggins 2001; Suurtamm et al. 2010). This diversity is likely due to the pedagogical freedom granted to French teachers but also to a lack of training in assessment and the absence of grading and assessment instructions in the curriculum.

Classroom assessment among French primary teachers appears to be more related to societal function, paying attention to social and personal behavior of students (Coll and Remesal 2009) than to pedagogical function (Coll et al. 2000; Coll and Martin 1996). That is why in France, formative assessment is not the common model of assessment in the classroom although recent educational policies have promoted it through the curriculum and the introduction of certain standards.

In addition to these new recommendations to incorporate more formative assessment, I believe that French teachers need to be trained in didactics so as to be better equipped to change their classroom assessment practices as well as enhance their professional judgement in order to better support both student learning and equity. We share the view of Morgan, who considers that:

> Assessing students in the subject of mathematics is a complex endeavor that relies on different understandings of the purposes of assessment, as well as what it means to know and/or do mathematics and whether and how this knowledge and these activities can be observed and evaluated. (Morgan 1999, p. 18)

It is well known that assessment, especially classroom assessment, can promote student learning and improve teachers' instruction (e.g., Hao and Johnson 2013;

Stiggins 2001, 2002), not only in France but also in all countries in the world. Enhancing classroom assessment is, however, a considerable challenge, particularly in France with respect to its national assessment culture.

References

Allal, L. (2013). Teachers' professional judgement in assessment: A cognitive act and a socially situated practice. *Assessment in Education: Principles, Policy & Practice, 20*(1), 20–34.

Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review: Assessment and learning*. Oxford, England: Norwegian Knowledge Centre for Education, Oxford University Centre for Educational Assessment.

Barr, M. A., & Cheong, J. (1995). Achieving equity: Counting on the classroom. *Evaluation in Education and Human Services, 40,* 161–184.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Beswick, K. (2011). Putting context in context. An examination of the evidence for benefits of 'contextualised' tasks. *International Journal of Science and Mathematics Education, 9,* 367–390.

Black, P., & Wiliam, D. (1998). *Inside the black box: Raising standards through classroom assessment*. London, England: King's College.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81–90.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more 'real'? *For the Learning of Mathematics, 13*(2), 12–17.

Brady, P., & Bowd, A. (2005). Mathematics anxiety, prior experience and confidence to teach mathematics among preservice education students. *Teachers and Teaching, 11*(1), 37–46.

Brousseau, G. (1988). Les différents rôles du maître. *Bulletin de l'A.M.Q. Montréal, 23*, 14–24.

Brousseau, G. (1997). *Theory of didactical situations in mathematics* (Didactique des mathematiques, 1970–1990). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chapman, O. (2016). Approaches and challenges in supporting mathematics teachers' change. *Journal of Mathematics Teacher Education, 19*(1), 1–5.

Charles-Pézard, M. (2010). Installer la paix scolaire, exercer une vigilance didactique. *Recherches en didactique des mathématiques, 30*(2), 197–261.

Chevallard, Y. (1989). Le concept de rapport au savoir. Rapport personnel, rapport institutionnel, rapport officiel. *Séminaire de didactique des mathématiques et de l'informatique, 108,* 103–117.

Chevallard, Y. (1999). L'analyse des pratiques enseignantes en théorie anthropologique du didactique. *Recherches en didactique des mathématiques, 19*(2), 221–266.

Chevallard, Y., & Feldmann, S. (1986). Pour une analyse didactique de l'évaluation. *IREM d'Aix-Marseille.*

Christoforidou, M., Kyriakides, K., Antoniou, P., & Creemers, B. P. M. (2014). Searching for stages of teacher's skills in assessment. *Studies in Educational Evaluation, 40,* 1–11.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1996). Teachers' assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment, 3*(2), 159–179.

Clarke, D. (1996). Assessment. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Vol. 1, pp. 327–370). Dordrecht, The Netherlands: Kluwer Academic.

Coll, C., Barberà, E., & Onrubia, J. (2000). La atención a la diversidad en las prácticas de evaluación. *Infancia y Aprendizaje, 90,* 111–132.

Coll, C., & Martín, E. (1996). La evaluación de los aprendizajes: una perspectiva de conjunto. Signos. *Teoría y práctica de la educación, 18,* 64–77.

Coll, C., & Remesal, A. (2009). Concepciones del profesorado de matemáticas acerca de las funciones de la evaluación del aprendizaje en la educación obligatoria. *Infancia y Aprendizaje, 32*(3), 391–404.

De Ketele, J-M. (1989). L'évaluation de la productivité des institutions d'éducation. *Cahier de la Fondation Universitaire; Université et Société, le rendement de l'enseignement universitaire, 3,* 73–83.

De Ketele, J.-M., & Gerard, F.-M. (2005). La validation des épreuves selon l'approche par les compétences. *Mesure et évaluation en éducation, 28*(3), 1–26.

De Landsheere, V. (1988). *Faire réussir, faire échouer*. Paris, France: Presses Universitaires de France.

Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: A bridge between beliefs and emotions. *ZDM: The International Journal on Mathematics Education, 43*(4), 471–482.

Fosnot, C. T., & Dolk, M. L. A. M. (2001). *Young mathematicians at work*. Portsmouth, NH: Heinemann.

Grapin, N. (2015). *Étude de la validité de dispositifs d'évaluation et conception d'un modèle d'analyse multidimensionnelle des connaissances des élèves de fin d'école* (Thèse de doctorat). Université Paris Diderot, Paris.

Hao, S., & Johnson, R. L. (2013). Teachers' classroom assessment practices and fourth-graders' reading literacy achievements: An international study. *Teaching and Teacher Education, 29,* 53–63.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28,* 524–549.

Houdement, C., & Chambris, C. (2013). Why and how to introduce numbers units in 1st and 2nd grades. In *Actes du colloque CERME, Turkey* (Vol. 8).

Jaworski, B. (2006). Theory and practice in mathematics teacher development: Critical inquiry as mode of learning in teaching. *Journal of Mathematics Teacher Education, 9,* 187–211.

Jong, C., & Hodges, T. E. (2015). Assessing attitudes toward mathematics across teacher education contexts. *Journal of Mathematics Teacher Education, 18*(5), 407–425.

Kilday, C. R., Kinzie, M. B., Mashburn, A. J., & Whittaker, J. V. (2011). Accuracy of teacher judgements of preschoolers' math skills. *Journal of Psychoeducational Assessment, 29*(4), 1–12.

Laveault, D. (2009). L'amélioration de l'efficacité du système éducatif: sur quels indicateurs s'appuyer? In X. Dumay & V. Dupriez (Eds.), *L'efficacité dans l'enseignement: promesses et zones d'ombre* (pp. 177–194). Bruxelles, Belgium: De Boeck.

Laveault, D., & Grégoire, J. (2002). *Introduction aux théories des tests: en psychologie et en sciences de l'éducation*. De Boeck Supérieur.

Leontiev A. (1976). *Le développement du psychisme. Problèmes*. Paris, France: Editions sociales (1ère édition, 1972, en russe).

Leontiev A. (1984). *Activité Conscience Personnalité*. Moscou, Russia: Editions du Progrès (1ère édition, 1975, en russe).

Leplat, J. (1997). *Regards sur l'activité en situation de travail*. Paris, France: Presses Universitaires de France.

Leplat, J., & Hoc, J.-M. (1983). Tache et activité dans l'analyse psychologique des situations. *Cahiers de Psychologie Cognitive, 3*(1), 49–63.

Lerman, S. (Ed.). (2014). *Encyclopedia of mathematics education*. Berlin, Germany: Springer.

Martinez, J. F., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgements and student achievement in mathematics: Evidence in the ECLS. *Educational Assessment, 14,* 78–102.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34–43.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*(4), 203–213.

McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading practices decision making*. Raserche Report, Virginia Commonwealth University.

Morgan, C. (1999). Assessment in mathematics education: A critical social research perspective. In J. Portela (Ed.), *Actas do IX Seminario de Investigaçao em Educaçao Matematica* (pp. 5–23). Guimaraes: Associaçao de Professores de Matematica.

Morgan, C., & Watson, A. (2002). The interpretative nature of teachers' assessment of students' mathematics: Issues for equity. *Journal for Research in Mathematics Education, 33*(2), 78–110.

Moss, C. (2013). Research on classroom summative assessment. In J. McMillan (Ed.), *Handbook of research on classroom assessment* (pp. 235–256). Thousand Oaks, CA: SAGE Publisher.

Moss, C., & Brookhart, S. (2014). *Learning targets*. Alexandra, VA: Association for Supervision and Curriculum Development.

Mottier Lopez, L., & Allal, L. (2008). Le jugement professionnel en évaluation: quelles triangulations méthodologiques et théoriques? In L. Paquay, C. Van Nieuwenhoven, & P. Wouters (Eds.), *L'évaluation, levier du développement professionnel? Tensions, dispositifs, perspectives* (pp. 237–250). Bruxelles, Belgium: De Boeck.

Mounier, E. (2010). *Une analyse de l'enseignement de la numération. Vers de nouvelles pistes* (Doctoral dissertation). Université Paris-Diderot-Paris VII.

Noizet, G., & Caverni, J.-P. (1978). *Psychologie de l'évaluation scolaire*. Paris, France: Presses Universitaires de France.

Nortvedt, G. A., Santos, L., & Pinto, J. (2016). Assessment for learning in Norway and Portugal: The case of primary school mathematics teaching. *Assessment in Education: Principles, Policy & Practice, 23*(3), 377–395.

Nunziati, G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers Pédagogiques, 280,* 48–64.

Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance. Building upon Black and Wiliam (2005). *The Curriculum Journal, 18*(1), 27–38.

Reuter, Y., Cohen-Azria, C., Daunay, B., Delcambre, I., & Lahanier-Reuter, D. (2013). *Dictionnaire des concepts fondamentaux des didactiques*. Bruxelles, Belgium: De Boeck.

Robert, A., & Rogalski, M. (2002). Comment peuvent varier les activités mathématiques des élèves sur des exercices? le double travail de l'enseignant sur les énoncés et sur la gestion de la classe. *Petit x, 60,* 6–25.

Roditi, E. (2011). *Recherches sur les pratiques enseignantes en mathématiques: apports d'une intégration de diverses approches et perspectives*. HDR. Université René Descartes-Paris V.

Rogalski, J. (2013). Theory of activity and developmental frameworks for an analysis of teachers' practices and students' learning. In F. Vandebrouck (Ed.), *Mathematics classrooms: Students' activities and teachers' practices* (pp. 3–22). Rotterdam: Sense Publishers.

Sayac, N., & Grapin, N. (2015). Evaluation externe et didactique des mathématiques: un regard croisé nécessaire et constructif. *Recherches en didactique des mathématiques, 35*(1), 101–126.

Schubaeur-Leoni, M.-L. (1991). L'évaluation didactique: une affaire contractuelle. In J. Weiss (Ed.), *L'évaluation, problème de communication* (pp. 79–95). Cousset, Suisse: Delval.

Senk, S. L., Beckman, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28,* 187–215.

Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69*(5), 363–368.

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758–765.

Stiggins, R. J. (2007). Conquering the formative assessment frontier. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 8–28). New York, NY: Teachers College Press.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy, & Practice, 17*(4), 399–417.

Tempier, F. (2016). New perspectives for didactical engineering: An example for the development of a resource for teaching decimal number system. *Journal of Mathematics Teacher Education, 19*(2), 261–276.

Tessaro, W. (2013). Améliorer la qualité des pratiques évaluatives des enseignants: une articulation entre formation initiale et formation continue, Evaluation et orientation à l'école. *Enjeux pédagogiques, 21* (février).

Tourmen, C. (2009). Evaluators' decision making: The relationship between theory, practice and experience. *American Journal of Evaluation, 30*(1), 7–30.

Vantourout, M., & Maury, S. (2006). Quelques résultats relatifs aux connaissances disciplinaires de professeurs stagiaires dans des situations simulées d'évaluation de productions d'élèves en mathématiques. *Revue des sciences de l'éducation, 32*(3), 759–782.

Van Den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the Learning of Mathematics, 25*(2), 2–23.

Van den Heuvel-Panhuizen, M. H. A. M. (1996). *Assessment and realistic mathematics education* (Vol. 19). Utrecht, The Netherlands: Utrecht University.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice, 17*(1), 59–75.

Other Publications

Rapport de l'Inspection Générale de l'Education Nationale. (2013, Juillet). *La notation et l'évaluation éclairées par des comparisons internationales*. n° 2013–072. http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/134000726.pdf.

Author Biography

Nathalie Sayac is a French lecturer in Didactics, member of the LDAR (Laboratory of Didactics of André Revuz, University Paris Diderot) and professor in the Teacher Training School where she has been teaching math to primary school teachers (University Paris-Est Créteil) for 20 years. She was a maths teacher for 10 years in junior and high school before becoming involved in primary teachers' training in mathematics. Her research focuses on mathematical teachers' practices and on teachers' training practices in math. For some years, she has been particularly interested in assessment in mathematics, both large-scale assessment and classroom assessment in France. She has developed a didactical approach to study mathematics assessment and classroom assessment. This approach integrates French didactic work and various other work on assessment from both francophone and anglophone traditions. She is involved in many collaborative research projects and promotes teacher training through research and collaboration with practitioners (e.g., field trainers, teachers).

Chapter 12
Assessing Visualization: An Analysis of Chilean Teachers' Guidelines

Melissa Andrade-Molina and Leonora Díaz Moreno

Abstract The aim of this paper is to describe how visualization should be assessed in schools as recommended by official curricular guidelines. We contend that researchers and policy makers have granted spatial abilities with the status of a key element to improve students' mathematics performance. However, this importance seems to fade when developing curricular guidelines for assessing students while learning school mathematics and geometry. We conducted an analysis of the teacher's official guidelines for the assessment of school mathematics in Chile. The analysis of two of those guides is considered in this paper. The results revealed that these guidelines do not provide sufficient guidance to teachers related to assessing visualization in schools; rather, their focus is on a culture of evaluation based on large-scale assessment techniques that leads to less emphasis on spatial abilities and more emphasis on calculation.

Keywords School geometry · Visualization · Spatiality · Assessment Guidelines for teachers

12.1 Introduction

As Suurtamm et al. (2016, p. 25) claim, assessing what students learn in schools "is a fundamental aspect of the work of teaching." They draw awareness to the complexity of the connection between large-scale assessment, for example PISA, and classroom assessment that is usually teacher-selected or teacher-made. Such

M. Andrade-Molina (✉)
Faculty of Engineering and Science, Aalborg University, Aalborg, Denmark
e-mail: melissa@plan.aau.dk

L. Díaz Moreno
Department of Mathematics, Faculty of Science,
Valparaiso University, Valparaiso, Chile
e-mail: leonora.diaz@uv.cl

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_12

interplay leads to external large-scale assessment often influencing assessment practices in the classroom.

> Although classroom teachers have long used various forms of assessment to monitor their students' mathematical learning and inform their future instruction, increasingly external assessments are being used by policy makers throughout the world to gauge the mathematical knowledge of a country's students and sometimes to compare that knowledge to the knowledge of students in other countries. (Op. cit., p. 1)

Assessments, both external large-scale and classroom assessments, as a means of monitoring and regulating teaching and learning practices of school mathematics are believed to "define what counts as valuable learning and assign credit accordingly" (Baird et al. 2014, p. 21), with valuable learning becoming "what is important to know and to learn" (Suurtamm et al. 2016, p. 6). In other words, the monitoring and regulating features of assessments determine what is considered the important and necessary mathematical knowledge that students should learn in schools and the feedback students should receive to improve their own performances. As stated by Swan and Burkhardt (2012), assessments not only have influence on the content of instruction, but additionally on how tasks are regularly presented to students, for example the preference for multiple-choice problems. Large-scale assessments, as pointed out by Suurtamm et al. (2016), are not concerned with examining students' thinking and communication processes—the learning itself—as they frequently use mathematical problems leading to only one correct answer. In contrast, formative assessment in classrooms is "informal assessments that teachers might do as part of daily instruction as well as more formal classroom assessments used to assess the current state of students' knowledge" (Op. cit., p. 14). Thus, formative assessment is "a social practice that provides continual insights and information to support student learning and influence teacher practice" (Suurtamm et al. 2010, p. 400). Building on Pellegrino et al.'s notion of assessment as "a process of reasoning from evidence" (2001, p. 2), the very nature of students' results taken as "evidence" leads to imprecise views regarding what students are able to do, what they have learned, and what they know; Pellegrino and colleagues add that such assessment results are just estimates of students' performances.

The impact that large-scale assessment has on policy, curriculum, and classroom practice influences the nature of classroom instruction, for example, by shaping curricular reforms (e.g., Barnes et al. 2000). This impact prompts Suurtamm et al. to recognize "the challenges teachers face in engaging in assessment practices that help to provide a comprehensive picture of student thinking and learning" (2016, p. 25). Schleicher, as a result of a study conducted among OECD countries, stated that formative-assessment techniques are "infrequently used in mathematics classes, with the exception of telling students what is expected of them for tests, quizzes and assignments" (2016, p. 20). Considering teachers' challenges regarding the assessment of school mathematics as a political dilemma, as described by Windschitl (2002), helps position the discussion over the struggles teachers face concerning their own views on classroom assessment and the educational policies and standards on national assessment (e.g., Engelsen and Smith 2014).

In this paper, we explore how official guidelines for teachers, released by the Chilean Ministry of Education (MINEDUC), recommend teachers assess school geometry in the classroom and what they should expect from students' performances. Particularly in school geometry, these recommendations expose a double discourse about, on the one hand, the importance of visualization as a necessary skill in problem-solving tasks, and on the other hand, expectations of students' learning that seem to be focused solely on calculation techniques. This paper seeks to explore how assessment ought to occur in the classroom according to teachers' official guidelines. That is, how do Chilean curricular guidelines inform teachers about how to assess visualization in schools?

12.2 Visualization, Spatiality, and School Geometry

Since the 1980s, research in the field of mathematics education has focused on understanding and grasping the complexity of visualization for the learning of school mathematics (see e.g., Swoboda and Vighi 2016). Probably one of the most prominent and most cited definitions is:

> Visualization is the ability, the process and the product of creation, interpretation, use of and reflection upon pictures, images, diagrams, in our minds, on paper or with technological tools, with the purpose of depicting and communicating information, thinking about and developing previously unknown ideas and advancing understandings. (Arcavi 2003, p. 217)

Visualization has become a form of reasoning that has particular significance, both in the learning and teaching of school geometry and school mathematics (Sinclair et al. 2016). On one hand, some studies advocate for the inclusion of this ability as part of the school mathematics curriculum (Sinclair and Bruce 2015). On the other, Presmeg (2014), for example, suggests that it is worth arguing whether visualization is accepted, encouraged, and valued in the classroom. In connecting school geometry and visualization, Andrade-Molina and Valero (2015) argue on the existence of a gap between the curricular aims of school geometry, in terms of the teaching of spatial abilities—"typically defined as spatial perception, visualization, and orientation" (Lee and Bednarz 2012, p. 15)—and a school geometry rooted in Euclid's axioms and Cartesian system. Skordoulis et al. (2009) claim that the reduced understanding of space in schools might lead to students' misconceptions in school geometry.

Often, space in schools is taken as Cartesian because of its connections with the vector model: students "are placed in real three-dimensional situations [which provide] new tools to make spatial and flat depictions, such as the vector model" (MINEDUC 2004, p. 68, our translation). To place students in real three-dimensional situations implies the visualization of an optically perceived world (see e.g., Crary 1992), which is different from the abstract reduction of school space into a Cartesian system. The concept of spatiality could be constructed from the interaction between concrete practices of human beings and of mathematical knowledge. Lefebvre (1991) considers space as a social production. In his understanding, space is experienced in three forms: as physical form that is generated and used (*Perceived*); as an instrumental form of knowledge—*savoir*—and logic (*Conceived*); and as a form of knowing—*connaissance*—that is produced and modified over time and through its use (*Lived*). Following the express claims of MINEDUC (2010), school geometry should pursue the connection of all three forms of space described by Lefebvre. The "real world" becomes the physical form of space, the tools and mathematical models become the instrumental form of knowledge and logic, and the context in which both interact becomes the *lived* form of space. However, usually

> Students are presented with the properties of shapes and theorems for proof …all the information needed is given in the problem, and the students are asked to apply the theorems in what has to be proven. …The skills needed to solve these types of problems are limited, and teaching these skills usually consists of demonstrating the appropriate technique followed by a series of similar problems for practice. (Mevarech and Kramarski 2014, p. 24)

This disconnection occurs even in initial years of schooling, in which Clements and Sarama (2011) have highlighted that the teaching of spatial abilities has been largely ignored in formal school settings. One possible explanation is that most geometric content, including skills such as visualization in school geometry, was removed from school as one of the consequences for the need of 'workers with particular skills' as an agenda of industrialization, urbanization, and capitalism (Whiteley et al. 2015). Others argue that students encounter difficulties "as a result of their spontaneous processes of visual perception in cases in which they contradict the geometric concepts/knowledge aimed at by the teacher and the tasks" (Gal and Linchevski 2010, p. 180). But, within research on geometry education, there has been little discussion about assessment in school geometry, and on how national curriculums address visualization in assessment guidelines for teachers and standardized tests (see e.g., Sinclair et al. 2016).

12.3 Exploration of Teachers' Guidelines

The Chilean Ministry of Education (MINEDUC) constantly releases guidelines for teachers to enhance their practices and to improve the learning of school mathematics. There exist diverse guidelines for teachers from MINEDUC: school textbooks with instructions for teachers, maps of learning progress, curricular materials and programs, learning standards, and more. All of those resources have recommendations about how teachers should evaluate and assess students' performance in the classroom. Here, we explore the recommendations for the assessment of school geometry to determine the advice the guidelines provide teachers related to assessing visualization and spatiality in the classroom. We take Lefebvre's understanding of space to search for expressions of spatiality and Arcavi's definition of visualization in the instructions for teachers and problem samples presented to teachers to assess school geometry. We present the analysis of the Learning Standards and Curricular program, both from the eighth year of compulsory education.

The learning standards (MINEDUC 2013) are aimed at teachers to guide them to evaluate what students "should know and can do to reach, in national tests, appropriate levels of achievement according to the fixed learning objectives in the current curriculum" (MINEDUC 2013, p. 4, our translation). These standards are a means to correlate what students have learned with the national syllabus (the school mathematics curriculum), and are aimed at helping teachers to determine what students need to learn and to monitor their progress (MINEDUC 2013). The standards present students' ideal answers by classifying them within three categories of accomplishment. The categories compare students' performances in the Chilean assessment system for measuring the quality of education (SIMCE) with the respective performance in the classroom. The categorization is given to the teacher explicitly in the learning standards. The learning standards also enable teachers to anticipate the score students would obtain in SIMCE given that this information might be gathered by the teacher through formative or summative assessment. As specified by the learning standards, if the performance of a student is categorized as *adequate*, the outcome of the test should be more than 297 points; if the performance is categorized as *elemental*, the outcome should be more than 247 and less than 297 points; and, if the performance is categorized as *insufficient*, the outcome should be less than 247 points. SIMCE's scores range from 0 to 400.

The three levels of the learning standards for the eighth year of compulsory education (twelve years in total) are described as:

Learning level *adequate* (Fig. 12.1): Students who are placed in this learning level in SIMCE demonstrate that they have accomplished the compulsory requirements to achieve this level, and also, that they have exceeded those compulsory requirements (MINEDUC 2013, p. 14, our translation).

Learning level *elemental* (Fig. 12.2): Students who are placed in this learning level in SIMCE demonstrate that they have accomplished the minimum compulsory requirements, and also, that they have exceeded those requirements but not enough to meet the requirements to reach the adequate learning level (MINEDUC 2013, p. 30, our translation).

Learning level *insufficient* (Fig. 12.3): Students who are placed in this learning level in SIMCE do not meet the compulsory requirements to achieve the elemental learning level. This level gathers those students who are far from reaching the requirements, but, also, those who are close to reaching the requirements (MINEDUC 2013, p. 38, our translation).

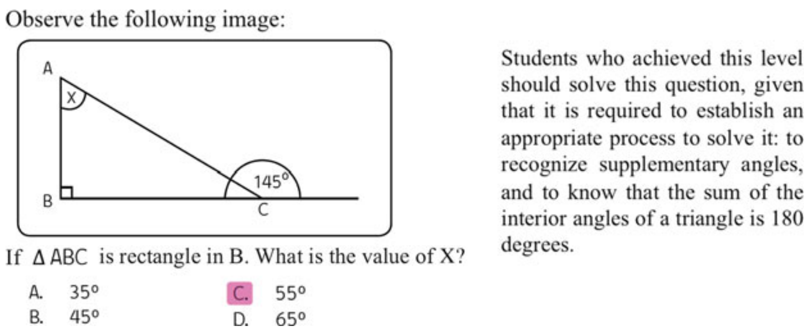


Fig. 12.1 Example of the "adequate" level (MINEDUC 2013, p. 20, our translation)

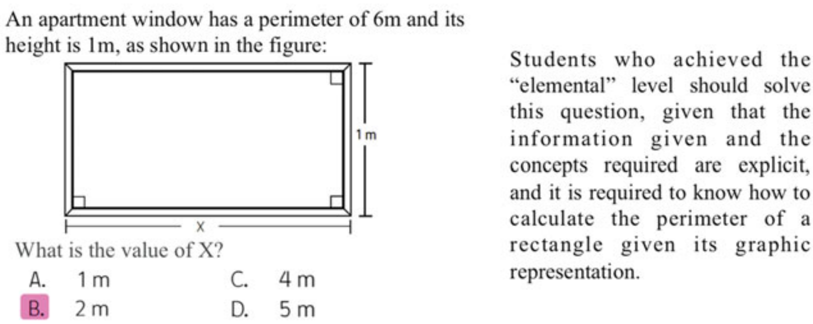


Fig. 12.2 Example of the "elemental" level (MINEDUC 2013, p. 35, our translation)

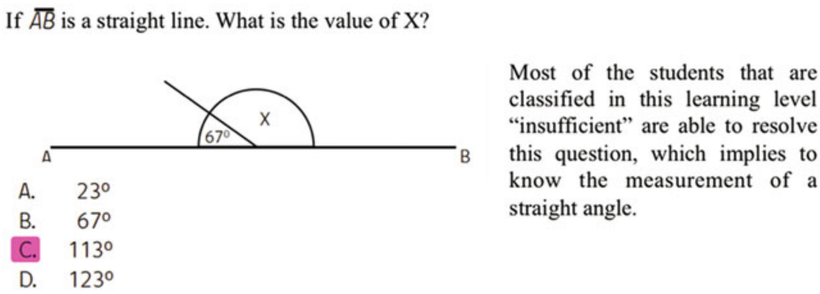


Fig. 12.3 Example of the "insufficient" level (MINEDUC 2013, p. 39, our translation)

Each level displays the students' ideal performances. Such performances are represented with a series of examples and their corresponding analysis. The analysis presents to teachers the detailed competencies students should use to obtain the correct answer. None of the examples mention the use of visualization as a possible ability to solve those problems. Figures 12.1, 12.2 and 12.3 show the actual recommendation for teachers to assess geometry in the classroom with the purpose of informing teachers on how students might do on SIMCE. The ideal performance (Fig. 12.1) corresponds to the "adequate" level, which enables teachers to predict that this student will score more than 297 in SIMCE.

The curricular programs (MINEDUC 2016) aim to guide teachers during the planning, coordination, evaluation, and so forth of each topic of the syllabus in mathematics. In these guidelines, teachers are presented with organized topics and recommendations on how to incorporate each unit of content into the classroom. The guide details the purpose of each topic, the mathematical knowledge required, abilities developed, attitudes that should be promoted, learning objectives, assessment indicators, and examples of tasks for each topic. The curricular program presents examples of problems teachers should use in the classroom to assess students or to introduce new content. These examples are accompanied by instructions such as "students should deduce the formulas to calculate the areas and volumes of straight prisms" (MINEDUC 2016, p. 140), and also by the expected performance of students. For example, Fig. 12.4 shows how the curricular program recommends teachers assess students' skills while calculating areas and volumes. The instructions expressed are in terms of deconstructing prisms and characterizing the resulting plane shape, calculating the volume of the prism (3D), deducing the formula to calculate the volume, and measuring the sides to calculate the area. The instructions are focused on calculation techniques; they do not express the use of visualization abilities. These examples inform teachers on the capabilities students should show when solving the task. There are not recommended questions, therefore, it is assumed by these examples that MINEDUC expects teachers to pose questions to students in terms of the indicators.

Figure 12.4 illustrates how teachers are guided to evaluate students' "intuitive estimation of the surface's area and volume" (MINEDUC 2016, p. 140), along with the formula to calculate the area and volume of the figure's surface, its proper application, and the result. There is no mention of visualization, for example as a skill needed for the interpretation of the deconstructed figure.

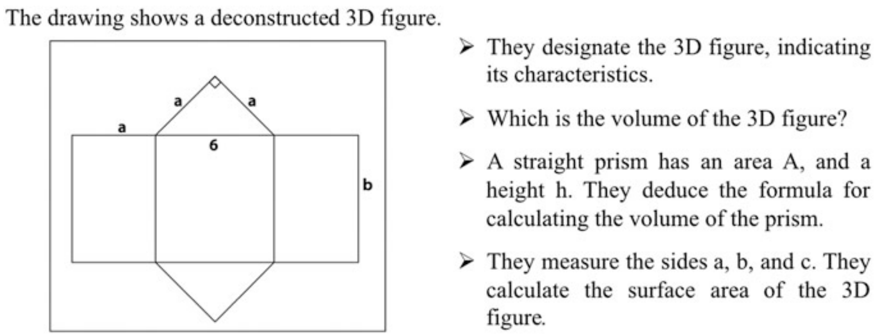


Fig. 12.4 Indicators for assessing areas and volumes of prisms (MINEDUC 2016, p. 140, our translation)

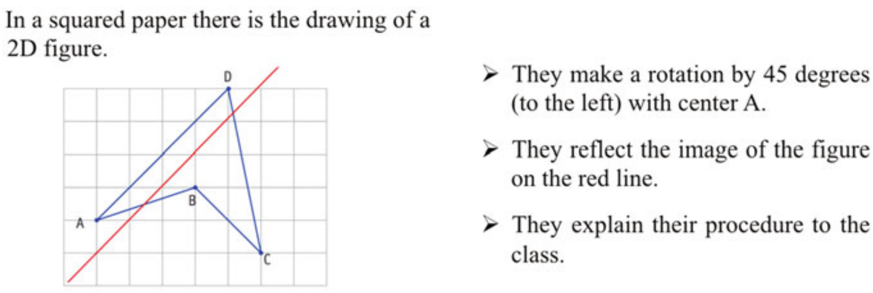


Fig. 12.5 The assessing of rotation skills (MINEDUC 2016, p. 158, our translation)

In another task (Fig. 12.5), the instructions for teachers assert the prompting of spatial abilities, such as spatial imagination. Therefore, students should be encouraged to use "pictorial representations" of 2D figures, both concretely and mentally (MINEDUC 2016). However, the considerations for assessment (translated in the image) are not directed towards visualizing (in terms of Arcavi), neither to spatiality (in terms of Lefebvre). These are closer to the idea of a space inscribed in a Cartesian system, or to the use of techniques like counting squares (explored in more detail in Andrade-Molina and Valero 2017).

12.4 The Assessment of School Geometry

In the previous five examples, visualization plays no apparent role while solving each problem. Although visualization is an ability that could be used by students while solving the tasks above, it is not considered as a part of the skills students should use. In the expressed guidelines for teachers, visualization and spatiality are not explicit, and therefore not a necessary topic for evaluation. As illustrated in the figures, MINEDUC expresses the importance of visualizing in school because it helps link the "real world"—the optically perceived world—and mathematics, for example, by stating that students "use their spatial abilities… to visualize figures in 2D and 3D" (MINEDUC 2016, p. 40, our translation). Visualization, as a form of reasoning, is reduced only to techniques of seeing and recognizing necessary information (Fig. 12.6).

For example, in 8th grade, in the "adequate" level students are able to solve problems by using geometrical axioms and applying the properties of triangles and squares to measure the interior and exterior angles. In the guidelines, teachers are advised to assess if students, for example, recognize supplementary angles and also the basic properties of a triangle, such as the sum of the interior angles is 180°, while calculating the interior angle of a right triangle by knowing the measurement of the opposite exterior angle (Fig. 12.1). Students are categorized as "insufficient" if they are only able to solve problems involving the measurement of a straight

To represent: To relate and to contrast information between different levels of representation. **To model:** To use models to solve problems	A rectangle is formed with 10 sticks of length x and 6 sticks of length y. The sticks of the same length are used in opposite sides. ➢ They deduce an algebraic expression to calculate the area of the rectangle. ➢ They transform algebraically the above expression in another product. ➢ They visualize with rectangles the algebraic transformation.

Fig. 12.6 Visualization in guideline tasks (MINEDUC 2016, p. 104, our translation)

angle. Teachers are advised to ask students the measurement of an angle by knowing its supplementary angle (Fig. 12.3). Students in this category cannot solve problems beyond this argument. But what if they use other types of techniques besides axioms and theorems? By only recognizing the results and not the procedures, school space turns into a mathematical space of *savoir* in which students should navigate on a Cartesian system by following Euclidean metrics. Therefore, it seems that students, from the curricular guidelines for teachers, are not confronted with events in which they develop tools outside mathematical axioms and theorems to navigate in space. In relation to Lefebvre's classification, the perceived space is not connected with the lived space. School geometry is taken only as a conceived space, which leaves the perceived space only reachable through axiomatics, and through reason and logic. Visualization takes the same connotation as depicting figures. Even though spatial thinking and reasoning are considered skills students should develop in early stages of schooling, there is no expressed guide on how to assess such skills.

> Students should learn to recognize, visualize and depict figures, and to describe the characteristics and properties of static and dynamic 3D shapes and 2D figures. Concepts are given for students to understand the structure of space and to describe with precise language what they know of their environment. The early study of objects' movement—reflection, translation and rotation—seeks to develop students' spatial thinking and reasoning. (MINEDUC 2012, p. 91, our translation)

When it comes to assessment in school geometry, visualization seems to blur, prompting what Andrade-Molina and Valero (2015) called the "sightless eyes of reason," in which the optically perceived world is transformed into a 'coordinate system world' inside the classroom. Students should use visualizing tools, according to MINEDUC, but in a space reduced to XYZ. As they contend, in MINEDUC activities involving visual–spatial skills, the reduction of space leads to model reality only in terms of axiomatical deductions. Visualization becomes a skill that is not assessed in schools, but is considered key in discourse produced by the Ministry of Education in Chile.

> Students should comprehend the representations of coordinates in a Cartesian system and use their visual–spatial skills. In this process of learning, students should use diverse instruments to visualize figures in 2D and 3D and manual and IT tools are recommended. (MINEDUC 2015, p. 100, our translation)

In the curricular guidelines analyzed in this exploratory study, geometry in school mathematics in Chile seeks the development of skills on vectors from axiomatic methods. In this sense, students perform successfully if they are able to solve problems by using geometry axioms and theorems. As stated in the previous section, assessment in geometry from official guidelines promotes a particular type of ideal answer. It is possible to state that there is a culture of evaluation based on large-scale assessment techniques—SIMCE and PSU—in Chile. These types of assessments are designed to certify students' competencies and skills while solving problems in optimum time, to test their knowledge and to grade students. MINEDUC, in order to achieve the desired score, guides teachers to use similar problems and styles of questions. Preiss describes the teaching and learning practices in school mathematics in Chile as a

> '[P]rivate appropriation of terms and procedures' because of the Chilean emphasis on individual work. The emphasis of Chilean teachers on practicing concurrent problems may indicate an overemphasis on skill drilling instead of mathematical understanding. (2010, p. 350)

12.5 Final Remarks

We are aware that there exist *real life situations* in which visualization and spatiality are not enough to solve a problem. Consider, for example, the following situation:

> An engineer analyses oscilloscopes' graphics to determine the functioning of an electronic component. He "cannot see" directly the waves, its resistances and other properties of the device. Oscilloscopes' graphics become an instrument for the design and evaluation of electronic components. (Arrieta and Díaz 2015, p. 34, our translation)

In this kind of situation, other teaching-learning strategies should intervene, such as variational thinking (see e.g., Carrasco et al. 2014).

Recall the main concern: how assessment ought to occur in the classroom according to teachers' official guidelines in Chile. Although existent research has shown the importance of visual–spatial abilities and, also skills for the sciences and for problem solving in the learning of school mathematics (Sinclair et al. 2016), there is no certainty on how these skills should be assessed in school. According to MINEDUC (2004), students are placed in 'real' three-dimensional situations to develop spatial thinking, but there is no guidance for teachers to assess these skills in the classroom. From the materials analyzed, spatial abilities are not officially assessed in schools. The analyzed materials show the importance of a spatiality constituted by a Cartesian coordinate system aimed at vector modeling. As students move forward in the levels of learning progress, they learn how to navigate in space only in terms of XYZ, or on Lefebrve's conceived space. And they are classified, according to the learning standards, for their expertise on geometry axioms and deductions. This classification help teachers identify how students might perform in large-scale assessments, which suggests these guides have the potential to align the curriculum, instruction, and assessment. However, from our analysis the official guidelines for teachers do not help teachers assess visualization. Apparently, this is a task for teachers to solve, even though visual–spatial abilities and skills are important for school mathematics (see e.g., Bruce et al. 2015). Therefore, it is possible to question whether there are more skills, considered to be important by MINEDUC, that have no recommendation for teachers on how they should be assessed in the classroom because of the impact that large-scale assessment has had on policy, curriculum, classroom practice and instruction, and assessment in Chile.

Acknowledgements This research is funded by the National Commission for Scientific and Technological Research, CONICYT PAI/INDUSTRIA 79090016, in Chile.

References

Andrade-Molina, M., & Valero, P. (2015). The sightless eyes of reason: Scientific objectivism and school geometry. In K. Krainer & N. Vondrová (Eds.), *Proceedings of the Ninth Conference of the European Society for Research in Mathematics Education* (pp. 1551–1557). Prague, Czech Republic: Charles University in Prague.

Andrade-Molina, M., & Valero, P. (2017). The effects of school geometry in the shaping of a desired child. In H. Straehler-Pohl, N. Bohlmann, & A. Pais (Eds.), *The disorder of mathematics educations. Challenging the sociopolitical dimensions of research* (pp. 251–270). Cham, Switzerland: Springer International Publishing.

Arrieta, J., & Díaz, L. (2015). Una perspectiva de la modelación desde la socioepistemología [A modeling perspective from socioepistemology]. *Revista Latinoamericana de Matemática Educativa, 18*(1), 19–48.

Arcavi, A. (2003). The role of visual representations in the learning of mathematics. *Educational Studies in Mathematics, 52,* 215–241.

Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review: Assessment and learning*. Oxford, England: Oxford University Centre for Educational Assessment, Norwegian Knowledge Centre for Education.

Barnes, M., Clarke, D., & Stephens, M. (2000). Assessment: The engine of systemic curricular reform? *Journal of Curriculum Studies, 32*(5), 623–650.

Bruce, C., Sinclair, N., Moss, J., Hawes, Z., & Caswell, B. (2015). Spatializing the curriculum. In B. Davis & Spatial Reasoning Study Group (Eds.), *Spatial reasoning in the early years. Principles, assertions and speculations* (pp. 85–106). New York, NY: Routledge.

Carrasco, E., Díaz, L., & Buendía, G. (2014). Figuración de lo que varia [Figuration of what varies]. *Enseñanza de las ciencias, 32*(3), 365–384.

Clements, D., & Sarama, J. (2011). Early childhood teacher education: The case of geometry. *Journal of Mathematics Teacher Education, 14*(2), 133–148.

Crary, J. (1992). *Techniques of the observer: On vision and modernity in the nineteenth century*. Cambridge, MA: MIT Press.

Engelsen, K. S., & Smith, A. (2014). Assessment literacy. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 91–107). Dordrecht, The Netherlands: Springer.

Gal, H., & Linchevski, L. (2010). To see or not to see: Analyzing difficulties in geometry from the perspective of visual perception. *Educational Studies in Mathematics, 74,* 163–183.

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*(1), 15–26.

Lefebvre, H. (1991). *The production of space* (D. Nicholson-Smith Trans.) Oxford, UK & Cambridge, MA: Blackwell.

Mevarech, Z., & Kramarski, B. (2014). *Critical maths for innovative societies: The role of metacognitive pedagogies.* OECD Publishing. https://doi.org/10.1787/9789264223561-en.

MINEDUC. (2004). *Matemática. Programa de Estudio. Cuarto Año Medio*. Santiago, Chile: MINEDUC.

MINEDUC. (2010). *Mapas de progreso del aprendizaje. Geometría*. Santiago, Chile: Gobierno de Chile.

MINEDUC. (2012). Matemática. In Ministerio de Educación & Bases Curriculares (Eds.), *Primero básico a sexto básico* (pp. 85–135). Santiago, Chile: Gobierno de Chile.

MINEDUC. (2013). *Estándares de Aprendizaje. Matemática. Octavo Básico*. Santiago, Chile: Gobierno de Chile.

MINEDUC. (2015). Matemática. In Ministerio de Educación, *Bases Curriculares. Séptimo básico a segundo medio* (pp. 93–103). Santiago, Chile: Gobierno de Chile.

MINEDUC. (2016). *Programas de Estudio. Matemática octavo año básico.* Santiago, Chile: MINEDUC.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science of design and educational assessment*. Washington, DC: National Academy Press.

Preiss, D. D. (2010). Folk pedagogy and cultural markers in teaching: Three illustrations from Chile. In D. D. Preaiss & R. J. Sternberg (Eds.), *Innovations in educational psychology: Perspectives on learning, teaching, and human development* (pp. 325–356). New York, NY: Springer.

Presmeg, N. (2014). Contemplating visualization as an epistemological learning tool in mathematics. *ZDM: The International Journal on Mathematics Education, 46*(1), 151–157.

Schleicher, A. (2016). *Teaching excellence through professional learning and policy reform: Lessons from around the world, International Summit on the Teaching Profession*. Paris, France: OECD Publishing.

Sinclair, N., Bartolini Bussi, M., de Villiers, M., Jones, K., Kortenkamp, U., Leung, A., et al. (2016). Recent research on geometry education: An ICME–13 survey team report. *ZDM: The International Journal on Mathematics Education*, *48,* 691–719.

Sinclair, N., & Bruce, C. (2015). New opportunities in geometry education at the primary school. *ZDM: The International Journal on Mathematics Education, 51*(3), 319–329.

Skordoulis, C., Vitsas, T., Dafermos, V., & Koleza, E. (2009). The system of coordinates as an obstacle in understanding the concept of dimension. *International Journal of Science and Mathematics Education, 7*(2), 253–272.

Suurtamm, C., Koch, M., & Arden, A. (2010). Teachers' assessment practices in mathematics: Classrooms in the context of reform. *Assessment in Education: Principles, Policy, and Practice, 17*(4), 399–417.

Suurtamm, C., Thompson, D. R., Kim, R. Y., Diaz Moreno, L., Sayac, N., Schukajlow, S., et al. (2016). *Assessment in mathematics education: Large-scale assessment and classroom assessment*. Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-32394-7.

Swan, M., & Burkhardt, H. (2012). A designer speaks: Designing assessment of performance in mathematics. *Educational Designer: Journal of the International Society for Design and Development in Education, 2*(5), 1–41.

Swoboda, E., & Vighi, P. (2016). *Early geometrical thinking in the environment of patterns, mosaics and isometries*. Cham, Switzerland: Springer International Publishing.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research, 72*(2), 131–175.
Whiteley, W., Sinclair, N., & Davis, B. (2015). What is spatial reasoning? In B. Davis & Spatial Reasoning Study Group (Eds.), *Spatial reasoning in the early years. Principles, assertions and speculations* (pp. 3–14). New York, NY: Routledge.

Author Biographies

Melissa Andrade-Molina has a Ph.D. in mathematics education from the Faculty of Engineering and Science at Aalborg University, Denmark. She holds a Master in Science with major in Mathematics Education from the Center for Research and Advanced Studies of the National Polytechnic Institute in Mexico, and a Bachelor in Education with major in Mathematics from the Silva Henriquez Catholic University in Chile. Her main research interests are sociopolitical perspectives of mathematics and science education, and philosophy of mathematics education. She researches how school geometry as part of the curriculum has been historically related to the production of mathematical-scientific subjectivities.

Leonora Díaz Moreno is Professor at the Department of Mathematics, Valparaiso University, in Valparaíso, Chile. She has a Ph.D. in Educational Sciences from the Pontifical Catholic University of Chile, a Master in Mathematics Education and a Master in Mathematics, both from the University of Santiago, Chile. Her main research interests are mathematical modelling and mathematics of change, teachers' training, assessment for the learning of mathematics, and design of undergraduate and postgraduate programs.

Chapter 13
Formative Assessment and Mathematics Teaching: Leveraging Powerful Linkages in the US Context

Megan Burton, Edward A. Silver, Valerie L. Mills, Wanda Audrict, Marilyn E. Strutchens and Marjorie Petit

Abstract Despite compelling evidence of the benefits of formative assessment on student learning, it is infrequently or unsystematically implemented in many U.S. classrooms. Consequently, the National Council of Supervisors of Mathematics (NCSM) and the Association of Mathematics Teacher Educators (AMTE) collaborated to relate formative assessment to other aspects of effective mathematics teaching, rather than treating it as an isolated topic. The Formative Assessment Initiative makes explicit the connection between formative assessment strategies and other instructional frameworks and tools intended to promote improved teaching and learning of mathematics. Because of its focus on promoting high quality mathematics teaching, the work of this U.S.-based project transcends boundaries and offers ideas that should be useful to mathematics teacher educators across the globe.

M. Burton (✉)
Auburn University, 5020 Haley Center, Auburn, AL 36849, USA
e-mail: megan.burton@auburn.edu

E. A. Silver
University of Michigan, Ann Arbor, MI 48109, USA
e-mail: easilver@umich.edu

V. L. Mills
Oakland Schools, 2111 Pontiac Lake Rd., Waterford, MI 48328, USA
e-mail: Valerie.Mills@oakland.k12.mi.us

W. Audrict
TLLR, Inc, 1590 Linksview Way, Stone Mountain, GA 30088, USA
e-mail: kuji00@hotmail.com

M. E. Strutchens
Auburn University, 5010 Haley Center, Auburn, AL 36849, USA
e-mail: strutme@auburn.edu

M. Petit
Ongoing Assessment Project, 1722 Airport Road, Moretown, VT 05660, USA
e-mail: mpetit@gmavt.net

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_13

Keywords Formative assessment · Professional development
Mathematics teacher education · Instructional strategies

13.1 Introduction

Assessment is an ongoing, informative process that is integral to effective instruction. It is implicated in teacher-student interactions that support students' communication and their thinking and the development of their understanding of mathematical ideas. A complete classroom assessment program contains both summative and formative assessments. Summative assessment focuses on assessment *of* student learning for evaluation (Black et al. 2004); formative assessment is referred to as assessment *for* learning (Broadfoot 2008; Stiggins 2005). This paper shares ideas gained from an initiative of the National Council of Supervisors of Mathematics (NCSM) and the Association of Mathematics Teacher Educators (AMTE) to promote an intentional and systematic approach to implementing formative assessment in U.S. mathematics classrooms. Although this paper is about the process of this initiative in the United States, the call to make formative assessment more explicit in mathematical professional development is a global issue. Linking this central practice to effective mathematics instruction to other professional development or educational experiences for teachers helps make visible the interlinking nature of effective teaching practices.

The National Council of Supervisors of Mathematics (NCSM) and the Association of Mathematics Teacher Educators (AMTE) advocate for instructional leaders to promote the use of formative assessment in effective mathematics instruction (NCSM/AMTE 2014). Although effective instruction involves deeply integrated formative assessment, this isn't always made visible during professional development and instructional discussions. Therefore, a joint task force was formed to promote and support the attention paid to formative assessment practices by mathematics instructional leaders as they worked with preservice and inservice teachers. In addition, the task force sought to better understand members' current thinking about and attention to formative assessment, other popular instructional frameworks, tools, and approaches (which for the purposes of this paper, will be called approaches) and the connections that might exist among them. Toward this end, the task force developed a joint position paper on formative assessment (NCSM/AMTE 2014), conducted a survey of its membership, and shared information about formative assessment in sessions at national and international conferences and through publications (Petit and Bouck 2015; Silver and Smith 2015).

In addition, international experts on mathematics teacher education and professional development, who had worked with and/or contributed to the development of selected approaches to teaching, participated in a working meeting at the University of Michigan funded from the U.S. National Science Foundation (DRL1439366). The five approaches of focus at this meeting were *Culturally Responsive Pedagogy* (Gay 2013), *Cognitively Guided Instruction* (Carpenter et al.

2014), *Classroom Discourse Tools* (Smith and Stein 2011), *Response to Intervention* (Gersten et al. 2009), and the *Mathematical Tasks Framework* (Stein et al. 2009). The approaches that were examined were selected based on feedback from the survey administered to mathematics teacher educators and supervisors about their use of assessment in teacher education courses and professional development opportunities. The membership overwhelming saw formative assessment as important to their work and effective teaching, and these five approaches were identified as the most widely utilized. Therefore, the meeting focused on whether and how formative assessment might be a more explicit focus in the work of these popular approaches.

A significant outcome of this working meeting was that experts familiar with these selected approaches recognized important connections between formative assessment practices and the approach for which they are associated, while acknowledging this connection hasn't always been explicit in the professional development. Further, it appears to a growing group of experts that explicitly making the role of formative assessment within their approach visible to educators might both advance understanding and use of their framework and deepen educators' understanding and use of formative assessment practices. This paper focuses on the importance of formative assessment in instruction by examining how it is seen in various common instructional approaches. Below we first discuss important elements of formative assessment practices and then make connections to its presence within the additional approaches we have studied. Similar connections between formative assessment and other important approaches utilized globally can also be found. This paper focuses on approaches utilized in the United States because this is where the work was conducted, but many of these (e.g., Cognitively Guided Instruction, Mathematical Tasks Framework) are also known and used across the globe.

13.2 Formative Assessment

Formative assessment focuses on using information about student thinking to inform the instruction so as to improve learning (Black et al. 2004). Ideally, it would be a prominent part of lesson planning and instructional enactment. It is a deliberate process which involves teachers and students gathering and utilizing information about what students know and can do. It is cyclical and provides feedback to students about their progress and guides decisions about the next instructional steps to take. It includes eliciting information about what students know and are learning, and uses this information to guide decisions about short-term, mid-range, and long-range instructional issues. Formative assessment—eliciting and using information about student thinking—is one of eight effective mathematics teaching practices emphasized in the *Principles to Action* (National Council of Teachers of Mathematics [NCTM] 2014). By understanding what students know and how they are thinking, teachers can adjust instruction to maximize learning potential for all students.

Wiliam (2011, p. 45) described three instructional processes associated with formative assessment: finding out where the students are and what they are thinking, knowing where they will be going, and determining how to get them to this destination. These three processes are found in the five aspects of instruction that characterize effective formative assessment in classrooms described by Leahy et al. (2005). These five aspects include:

(1) Sharing clear learning goals and benchmarks for success;
(2) Designing lessons that involve successful classroom discussions, questions, and instructional tasks;
(3) Enhancing and progressing learning through feedback;
(4) Fostering student ownership of learning and;
(5) Creating an environment where students serve as resources for one another.

In classrooms where teachers regularly employ formative assessment strategies, student learning is enhanced (Black and Wiliam 2010; Ehrenberg et al. 2001; Popham 2013). For example, Ehrenberg et al. (2001) estimated the impact of formative assessment on student achievement to be four to five times greater than the effect of reducing class size. However, despite its obvious importance, formative assessment has not taken hold in U.S. mathematics instruction. We hypothesize that this is due, at least in part, to inadequacies in its treatment in the initial preparation and continuing professional development of mathematics teachers. In these settings, formative assessment is typically addressed in isolation, rather than holistically connected to other aspects of effective mathematics teaching. Formative assessment is much more than the addition of an exit slip, test, or observation; it is interwoven into the fabric of teaching and learning. Effective instruction and formative assessment are indivisible (Black and Wiliam 2010). Because formative assessment is such a critical element of effective teaching within any frameworks, tools, and approaches (FTA), teacher leaders, supervisors, mathematics coaches, and teacher educators need to firmly understand formative assessment strategies, be able to effectively implement them in the classroom, and promote the strategies throughout their work with teachers and others.

13.3 Making Formative Assessment Visible for Teachers

One way for teacher educators, teacher leaders, and mathematics coaches to promote formative assessment strategies is through providing sustained, meaningful professional development opportunities which support and model the effective use of formative assessment in the classroom. Making formative assessment explicit in professional development allows teachers to see how integral it is in effective teaching. Teachers may develop and/or enhance their use of the five formative

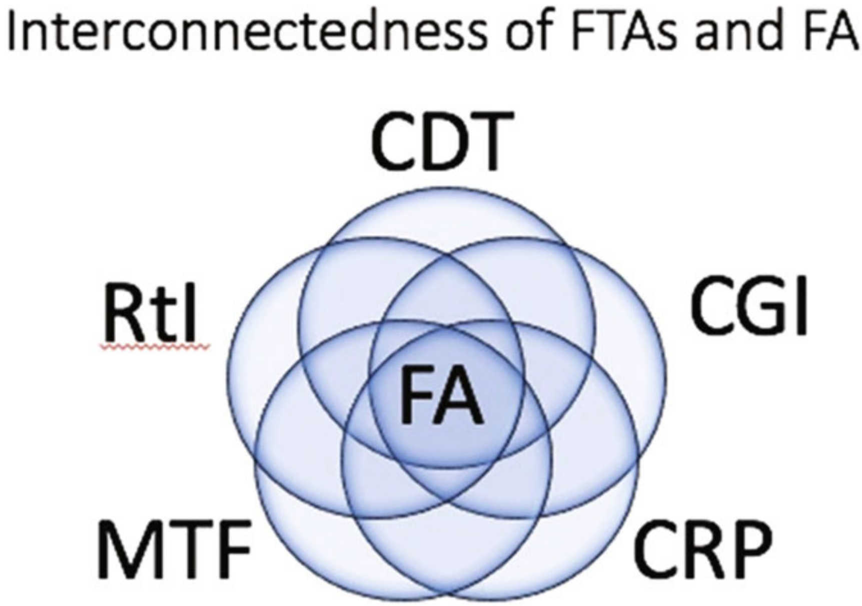


Fig. 13.1 The interconnectedness of frameworks, tools, or approaches (FTA) and formative assessment (FA)

assessment strategies described in Leahy et al. (2005) through professional development opportunities that connect to their own classrooms.

When providing professional development, teacher leaders, supervisors, coaches, and educators need to explicitly discuss the role formative assessment plays in the classroom instructional frameworks they are utilizing, such as Culturally Responsive Pedagogy (CRP), Cognitively Guided Instruction (CGI), Classroom Discourse Tools (CDT), Response to Intervention (RtI), and the Mathematical Tasks Framework (MTF) (see Fig. 13.1). Although each of the approaches examined may have some overlap with other approaches, formative assessment is central to each and is a common thread. This could be said of many other common approaches that are utilized around the world. Teachers need to see the integral role formative assessment plays in each of these research-based frameworks.

Below are examples of how formative assessment is interwoven into the approaches examined by this initiative. However, connections can be made to other approaches that focus on effective teaching and learning.

13.3.1 Culturally Responsive Pedagogy (CRP)

Culturally Responsive Pedagogy places focus on the local context and culture of the students and thus empowers students. This notion of connecting to student culture and context is relevant globally and “exemplifies the notion that instructional practices should be shaped by the sociocultural characteristics of the settings in which they occur, and the populations for whom they are designed” (Gay 2013, p. 63). Communication is key in this approach. For teachers to understand their students, they need to listen and observe. This involves utilizing formative assessment. When Culturally Responsive Pedagogy is implemented, the need for diversity in thinking about mathematics (e.g., style, pace, connections) is appreciated. Therefore, the evidence will be diverse as well. By noting how students think about and approach mathematical situations, teachers are able to adjust and build the most effective learning opportunities. Formative assessment allows students the opportunity to share their views, experiences, and thinking and it is a natural part of this approach.

When offering professional development on Culturally Responsive Pedagogy, it is important to help teachers identify the cultural aspects of mathematics and the importance of building upon students’ strengths. Considering ways to collect evidence that honors the cultural diversity in the classroom and how this can be used to drive instruction can be a useful element of professional development in this approach. Exploring how a teacher’s own experience shapes the instruction and assessment of student learning is critical in professional development. For example, in a classroom where debate of mathematical ideas is encouraged, it would be important for a teacher to recognize that some families do not support children debating adults rather than interpret silence as lack of engagement by a student. The experience of justifying and critiquing reasoning may need to be altered to empower these students (by having students debate each other or finding alternate ways that develop classroom community while supporting individual cultural norms). For more information about Culturally Responsive Pedagogy in mathematics education, the reader may find *The Impact of Identity in K-8 Mathematics: Rethinking Equity-based Practices* (Aguirre et al. 2013) to be a useful resource.

13.3.2 Cognitively Guided Instruction (CGI)

Cognitively Guided Instruction is a framework that focuses on teachers understanding student thinking and utilizing this thinking to guide instruction. The framework utilizes problem types and solution strategies to help inform teachers and students about student thinking and make appropriate plans to build upon student knowledge (Carpenter et al. 2014). Using this framework, teachers are able to examine the difficulty of problems and scaffold instruction based on student responses to various problems.

Consider the following scenario. Composing numbers greater than 10 is a learning goal for a first-grade class based on state standards. Individual students have more specific learning goals, based on analysis of their individual prior work and understanding of the framework (e.g., using the correct strategy to solve a problem, counting by ones, making groups of ten to solve the problem, utilizing place value). Cognitively Guided Instruction focuses on individual learning; each child has a specific learning goal. The problem given will be open enough that it contains multiple entry points to accommodate the diversity in the classroom. The instructional decisions for each child are based upon analysis of prior student work. The teacher will ask specific questions of each learner, based on his or her individual learning goals. Either way, teachers circulate among the room as students work independently. Teachers take anecdotal notes, pose questions, and utilize these observational notes to engineer whole class discussions at the end. Although all five formative assessment strategies (Leahy et al. 2005) are present in this framework, students as owners of their learning is central to Cognitively Guided Instruction.

When providing professional development related to this approach, it is essential to practice gathering information about student thinking, interpreting this information, and using it to plan future instruction. During professional development, teachers learn about the problem types and their progressive difficulty. They utilize student samples to begin to understand student thinking and identify scaffolding necessary for future learning. Teachers begin to see the way the evidence of student thinking can effectively guide instruction. During professional development, teachers see videos of students solving problems, conduct interviews with students, and examine classroom embedded work. These experiences are all useful to the formative assessment cycle. Professional development leaders need to explicitly ensure teachers recognize the experiences as examples of formative assessment. For more information on this framework and other Cognitively Guided Instruction Frameworks see Carpenter et al. (2014).

13.3.3 Classroom Discourse (CD)

Classroom Discourse focuses on understanding the various forms of discourse in the classroom. Moreover, Classroom Discourse involves creating more productive, student-centered discussions, intentionally planning questions that elicit thinking, asking students to clarify their thinking, and encouraging others to engage. Thus, this approach involves planning the discourse, but also being purposeful about the discourse based on observations during instruction and reflecting on discourse needed to further the learning in future lessons. Students need to be empowered to make sense of their own mathematical thinking, as well as the thinking of their peers. Discourse focuses on uncovering both the understanding and misunderstandings that are present and utilizing this information to inform future instruction. An example of providing feedback to move the learning forward would be a teacher saying, “One group found 32 and another group found 34. Does anyone have a prediction of why their answers might be different?” Helping students own their learning might involve asking, “Does anyone want to revise his or her answer? If so, can you explain why?”

When offering professional development around Classroom Discourse, it is important for teachers to see how discourse moves connect to formative assessment. When selecting tasks, teachers can analyze opportunities for formative assessment. While learning discourse strategies, teachers need to explicitly see how these strategies can be utilized during the instruction cycle to formatively assess student learning. Professional development can include using case studies to consider evidence of learning and how data can be used to form future instruction. For more information on this approach see *Five Practices for Orchestrating Productive Mathematical Discussions* (Smith and Stein 2011).

13.3.4 Response to Intervention (RtI)

Response to Intervention is a multi-tiered instructional support system aimed at meeting the diverse needs of students. A triangle figure is often used as a visual for this 3-level support system. It is based on the belief that if all students receive universal, high quality, engaging lessons, that approximately 80% of the students will have their needs met. This universal level of instruction is known as tier 1. Students who do not respond to this core instruction receive high quality supplementary instructional strategies, which is the only additional instruction needed by approximately 15% of the students; this targeted instruction is known as tier 2. Approximately 5% of classroom students will need intensive, individualized strategies, because they don't respond to tier 2 interventions; this level of instruction is known as tier 3.

Conducting diagnostic interviews and planning interventions based on evidence of student struggles is at the heart of Response to Intervention. Ensuring teachers have shared expectations about the use of formative assessment when utilizing this approach is essential. Each tier of instruction requires teachers to formatively assess progress to inform whole group, small group, or individual instruction. Formative assessment at all levels allows the teacher to identify both the strengths as well as areas of focus for instruction of all learners. Formative assessment goes beyond exploring if students have the right answer and explores student thinking, which is key to address issues they may have. Professional development of Response to Intervention should include various types of formative assessment strategies that could be useful at each tier. In addition, discussing which Response to Intervention assessments are more summative rather than formative is key. Connecting the five formative assessment strategies (Leahy et al. 2005) to the formative assessment that should occur in classrooms using this approach allows teachers to note the important role it plays in quality instruction at all levels. For more information, see *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (Gersten et al. 2009).

13.3.5 Mathematical Tasks Framework (MTF)

The Mathematical Tasks Framework is a tool that allows teachers to discuss various tasks that occur during instruction and the different learning opportunities they provide. Teachers examine and identify the cognitive demand involved in tasks and the processes a task goes through (Stein et al. 2009). Formative assessment happens everywhere in the Mathematical Tasks Framework. However, it is especially visible as teachers determine when to move to the next phase of the lesson, how the evidence collected indicates cognitive demands, and how student work relates to the goal and informs future instruction. The Mathematical Tasks Framework supports teachers in analysis of tasks to determine the level of thinking possible for students and the most effective way to set up a task to reach its maximum potential of student growth. The set-up or launch of a task communicates the expectations and learning goals for students, which is one of the five strategies listed by Leahy et al. (2005).

When implementing professional development on the Mathematical Tasks Framework, it is important that leaders provide opportunities for teachers to explore the formative assessment evidence that can be gained from various tasks and discuss how this evidence can be utilized to move learning forward. For more information see *Implementing standards-based mathematics instruction: A casebook for professional development, Second Edition* (Stein et al. 2009).

13.4 Discussion

To address and dispel this view of formative assessment as something “extra,” we advocate that teacher educators and professional development specialists explicitly connect formative assessment to other frameworks, tools, and approaches utilized in their work. Despite their differences, the approaches used in our work emphasize important aspects of formative assessment, such as eliciting students’ thinking and using this information to inform instructional decisions.

We argue that the explicit foregrounding of formative assessment in connection with these approaches can both help support the increased, effective use of formative assessment in mathematics teaching and bring greater coherence to professional development. The National Council of Teachers of Mathematics (2014) highlights the importance of a coherent curriculum that organizes and integrates important content to empower students to make connections and build upon existing ideas. This same coherence is important in professional development opportunities for teachers. Providing experiences that make visible the connections among important instructional ideas is key to developing new understandings.

The teacher’s role in formative assessment takes many forms including facilitating classroom discussions, questioning, eliciting student thinking, analyzing student work, providing feedback to students, and using formative assessment data to make instructional decisions. “In a classroom that uses assessment to support learning, the divide between instruction and assessment blurs” (Leahy et al. 2005, p. 22). Teaching plans can be adjusted based on the information gathered through questioning, classroom discussions, observation, and other formative assessment strategies. Just as the line between instruction and assessment blurs, the lines between the eight mathematics teaching practices (NCTM 2014) blur, because each impacts the effectiveness of the other. This is why explicitly sharing the role of formative assessment in various approaches is needed (see Fig. 13.2) to enable teachers to appropriately interpret evidence from formative assessment data and respond in a manner that moves students forward in their thinking. Figure 13.2 shares where various approaches fall in relation to elements of instruction. However, it illustrates that formative assessment is seen in all three elements of instruction on the chart.

Evidence from participants in the working meeting and responses from those who have attended conference sessions regarding this approach to treating formative assessment have been very encouraging. For example, 18 of the 19 working meeting participants indicated that they had developed a new and increased appreciation of the importance of formative assessment and its connection to the approaches they use, and they planned to implement these ideas in their work. This work has informed multiple presentations and publications. However, the true test is the impact this approach has on teachers and their ability to see formative assessment within the approaches and instruction they implement daily in the classroom. Future research on the impact of implementing professional development of a framework, tool, or approach with specific, explicit attention to formative assessment is needed. Does this provide coherence in teacher views on instruction and/or professional development? Does it change practices in formative assessment? How does this approach impact participants’ views of the frameworks, tools,

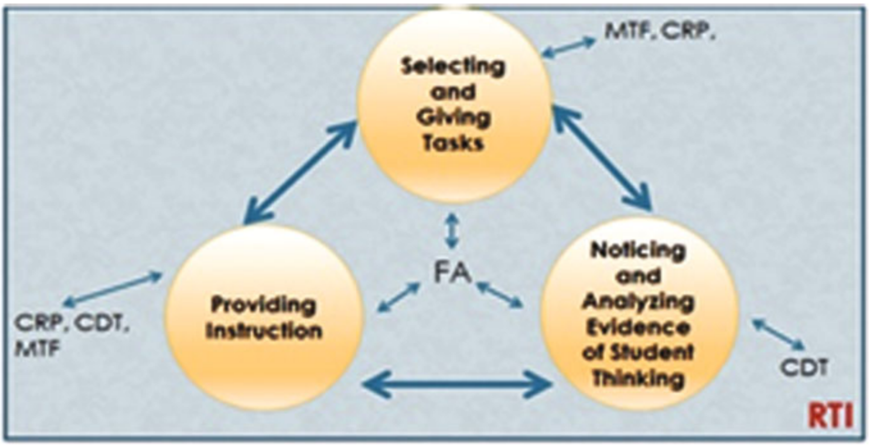


Fig. 13.2 Sample of connections of frameworks, tools, and approaches to formative assessment

and approaches? Each of these questions could be explored in future studies. In addition, examinations of the place of formative assessment in frameworks, tools, and approaches that are widely used in other countries would be useful. It would provide more strength to global arguments on the integral role of formative assessment in instruction.

Acknowledgements Preparation of this paper was supported in part by a grant (DRL1439366) from the National Science Foundation. The views expressed herein are those of the authors and do not necessarily reflect those of the Foundation.

References

Aguirre, J., Mayfield-Ingram, K., & Martin, D. (2013). *The impact of identity in K-8 mathematics: Rethinking equity-based practices*. Reston, VA: National Council of Teachers of Mathematics.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Assessment for learning: Putting it into practice*. New York, NY: Open University Press.

Black, P., & Wiliam, D. (2010). Inside the back box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 80–90.

Broadfoot, P. (2008). *An introduction to assessment.* London, England: Continuum.

Carpenter, T., Fennema, E., Franke, M., Levi, L., & Empson, S. B. (2014). *Children's mathematics, second edition: Cognitively Guided Instruction*. Portsmouth, NH: Heinemann.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Williams, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2*(1), 1–30.

Gay, G. (2013). Teaching to and through cultural diversity. *Cultural Diversity and Multicultural Education, 43,* 48–70.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., et al. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 19–24.

National Council of Supervisors of Mathematics/Association of Mathematics Teacher Educators. (2014). *Improving student achievement in mathematics through formative assessment in instruction: An AMTE and NCSM joint position paper.* Retrieved from https://amte.net/resources/formative-assessment.

National Council of Teachers of Mathematics. (NCTM). (2014). *Principles to action: Ensuring mathematics success for all.* Reston, VA: Author.

Petit, M., & Bouck, M. (2015). The essence of formative assessment in practice: Classroom examples. *NCSM Journal of Mathematics Education Leadership, 16*(1), 19–28.

Popham, J. (2013). Formative assessment's 'advocatable moment'. *Education Week, 32*(15), 29.

Silver, E. A., & Smith, M. S. (2015). Integrating powerful practices: Formative assessment and cognitively demanding mathematics tasks. In C. Suurtamm & A. Roth McDuffie (Eds.), *Annual perspectives in mathematics education (APME) 2015: Assessment to enhance learning and teaching* (pp. 5–15). Reston, VA: National Council of Teachers of Mathematics.

Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematical discussions*. Reston, VA: National Council of Teachers of Mathematics.

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2009). *Implementing standards-based mathematics instruction: A casebook for professional development* (2nd ed.). New York, NY: Teachers College Press.

Stiggins, R. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324–328.
Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Author Biographies

Megan E. Burton is an associate professor and the elementary education program coordinator at Auburn University, Alabama (USA). She teaches and advises undergraduate and graduate students in elementary education and conducts research related to elementary mathematics education, with focus on elementary teacher change, inclusion, and rural education. As a former elementary teacher with experience in inclusion and English Language Learners, Burton is committed to classrooms that allow all students to encounter strong mathematics instruction in meaningful, differentiated ways.

Edward A. Silver is the Senior Associate Dean for Research and Graduate Studies and the William A. Brownell Collegiate Professor of Education in the School of Education and professor of mathematics in the College of Literature, Science & the Arts at the University of Michigan (USA). He teaches and advises graduate students in mathematics education and conducts research related to the teaching and learning of mathematics, with particular attention to the study of mathematical thinking, enhancing the knowledge and effectiveness of mathematics teachers, and productive interchange between education research and practice.

Valerie L. Mills is a Mathematics Education Consultant for Oakland Schools, a Michigan (USA) regional resource center. Beginning as a high school mathematics teacher, she has held various leadership positions in Michigan school districts and served professional organizations, including the National Council of Supervisors of Mathematics as president. Mills was Principal Investigator on five large grants supporting high-needs districts, the author of numerous articles and professional development resources, and a frequent speaker at professional meetings. She is the recipient of the Outstanding Achievement award from the Michigan Council of Teachers of Mathematics, the Presidential Award for Excellence in Mathematics and Science Teaching, and the Milken National Educator Award.

Wanda Audrict is Cofounder/Executive Director for Executive Consultants—TLLR, Inc. Her work supports effective Teaching, Learning, Leadership, and Research. She earned a B.S. in Mathematics/Physics; a M.S. in Management/Administration; an Ed.S. in Curriculum and Instruction; and course work for a doctorate in Organizational Leadership. Her career began as a Telecommunications Senior Network Design/Equipment Engineer for 20 years. Subsequently, for the last 25 years, she has been a dedicated education professional—mathematics teacher, K-12 District Mathematics Coordinator, Mathematics Consultant, and a Director on the National Council of Supervisors of Mathematics Board (NCSM). Her focus is increasing student engagement through effective formative assessment.

Marilyn E. Strutchens is the Emily R. and Gerald S. Leischuck Endowed Professor and the Mildred Cheshire Fraley Distinguished Professor of Mathematics Education in Curriculum and Teaching, Auburn University, Alabama (USA). She is the secondary mathematics education coordinator and director of the Professional Mathematics Learning Communities Project.

Strutchens is a Board member for the National Council of Teachers of Mathematics and the Education and Human Resource Directorate for the National Science Foundation. She is an editor and author of many educational publications. Strutchens is the 2017 Judith Jacob's Lecturer for the Association of Mathematics Teacher Educators and a past president of the organization.

Marjorie Petit works primarily in supporting the Ongoing Assessment Project (OGAP)—a formative assessment project focused on additive reasoning, fractions, multiplicative reasoning, and proportionality. She is currently co-director of OGAPMath LLC. Marge Petit brings to this work three decades of research and development in standards-based restructuring efforts at the classroom, district, state and national levels, and 23 years of experience as a classroom teacher. She served as Vermont's Deputy Commissioner of Education from 1996 to 2000, and was a Senior Associate for the Center for Assessment from 2000 to 2004, focused on supporting states in the use of classroom assessment.

Chapter 14
Designing for Formative Assessment: A Toolkit for Teachers

David Wright, Jill Clark and Lucy Tiplady

Abstract This paper describes the outcomes of the three-year design research project: Formative Assessment in Science and Mathematics Education (FaSMEd) (http://www.fasmed.eu). Its goals: to develop a toolkit for teachers, and to research the use of technology in formative assessment in mathematics and science. Countries in the partnership included the United Kingdom, Republic of Ireland, Norway, The Netherlands, France, Germany, Italy, and South Africa. The project used an iterative, collaborative, process-focused approach. Case studies drew on a wide variety of researcher and teacher obtained evidence. The paper draws on analysis of the case studies and country reports to provide policy recommendations and suggestions for further research.

Keywords Formative assessment · Technology · Design study
Toolkit · Raising attainment

14.1 Introduction

The report *Knowing What Students Know* (Pellegrino et al. 2001) identified progress in the science of designing assessments as a key factor in enhancing classroom assessment. This paper describes a design research project focused on the design of assessments and its outcomes.

The Rocard report (2007) identified widespread concern across the European Union (EU) about the economic consequences and social impact of under-achievement in mathematics and science education and recommended the adoption of an inquiry based pedagogy. Consequently, a range of research projects were

D. Wright (✉) · J. Clark · L. Tiplady
Research Centre for Learning and Teaching, Newcastle University,
Newcastle upon Tyne NE1 7RU, UK
e-mail: wrightdavidg@gmail.com

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_14

commissioned by the European Commission (EC), for example: SAILS—Strategies for Assessment of Inquiry Learning in Science[1]; MASCIL—Mathematics and Science for Life[2]; PRIMAS—Promoting Inquiry in Mathematics and Science Education across Europe[3], and ASSIST-ME—Assess Inquiry in Science, Technology and Mathematics Education.[4]

Formative Assessment in Science and Mathematics Education[5] (FaSMEd) was commissioned in the FP7[6] programme with a focus on researching the application of technology to facilitate Formative Assessment (FA) in the classroom. The project investigated the conditions and requirements for promoting sustainable, appropriate and innovative socio-technical approaches to the raising of achievement in mathematics and science education, developed exemplars and support for teachers in the form of a 'toolkit', and provided policy recommendations and suggestions for further research.

The approach adopted within FaSMEd was based on knowledge of complex organizational work design focused on the interaction between people (in this case teachers and students) and technology in workplaces (classrooms). The aim was to explore how technology can raise attainment in mathematics and science classes using formative assessment, given that it had been evidenced to produce substantial student learning gains and to have a greater effect size[7] on achievement than many other interventions in the classroom (Black and Wiliam 1996; Hattie 2009).

FaSMEd is a collaborative development project, in a partnership consisting of the United Kingdom, The Republic of Ireland, Norway, The Netherlands, France, Germany, Italy, and South Africa. The project adapted the principles of design research (Swan 2014) in its methodology. This is a formative approach in which a product or process (or 'tool') is envisaged, designed, developed, and refined through cycles of enactment, observation, analysis and redesign, with systematic feedback from end-users. Educational theory is used to inform the design and refinement of the tools, and is itself refined during the research process. Its goals are to create innovative tools for others to use, to describe and explain how these tools function, account for the range of implementations that occur, and develop principles and theories that may guide future designs. Ultimately, the goal is

[1]http://www.sails-project.eu/.

[2]http://www.mascil-project.eu/.

[3]http://www.primas-project.eu/.

[4]http://assistme.ku.dk/.

[5]http://www.fasmed.eu/.

[6]**FP7** is the short name for the Seventh Framework Programme for Research and Technological Development. This is the European Union's main instrument for funding research in Europe and it ran from 2007–2013. **FP7** was designed to respond to Europe's employment needs, competitiveness, and quality of life.

[7]**Effect size** is a simple way of quantifying the difference between two groups that has many advantages over the use of tests of statistical significance alone. **Effect size** emphasises the **size** of the difference rather than confounding this with sample **size**.

transformative; we seek to create new teaching and learning possibilities and study their impact on end-users.

A key element of teaching using assessment and intervention relates to the quality of the information generated by the various feedback loops that exist in the classroom setting and the involvement of the students within this process. By introducing innovative technology to create a digital environment which enhances connectivity and feedback to assist teachers in making more timely formative interpretations, the FaSMEd project explored the potential to amplify the quality of the evidence about student learning both in real-time and outside the classroom, for access by both students and teachers.

As presented in our position paper[8], the following are the innovative features of connected classroom technologies that, as outlined by researchers, make them effective tools to develop formative assessment:

- *they give immediate information to teachers, enabling them to monitor students' incremental progress and keep them oriented on the path to deep conceptual understanding, providing appropriate remediation to address student needs* (Irving 2006; Shirley et al. 2011);
- *they support positive student thinking habits, such as arguing for their point of view* (Roschelle et al. 2007)*, seeking alternative representations for problems, comparing different solution strategies, explaining and describing problem-solving strategies* (Irving 2006);
- *they create immersive learning environments that highlight problem-solving processes and make student thinking visible* (Looney 2010);
- *they enable most of the students to contribute to activities, taking a more active role in discussions* (Roschelle and Pea 2002; Shirley et al. 2011);
- *they display aggregated student results, giving powerful clues to what students are doing, thinking, and understanding* (Roschelle et al. 2004*) and enable teachers to "take the pulse" of learning progress for the classroom as a whole* (Roschelle and Pea 2002);
- *they provide students with immediate private feedback, encouraging them to reflect and monitor their own progress* (Looney 2010; Roschelle et al. 2007);
- *they provide opportunities for independent and collaborative learning* (Looney 2010), fostering classroom discourse (Abrahamson et al. 2002; Dufresne et al. 2000; Roschelle et al. 2007; Shirley et al. 2011);
- *they offer potentially important avenues for enlarging the types of cultural practices used as resources for learning and foster students' dynamic engagement in conceptual, collective activities which are more akin to practices of professional communities, making them become knowledge producers rather that consumers* (Ares 2008);
- *they enable a multi-level analysis of patterns of interactions and outcomes, thanks to their potential to structure the learning space to collect the content of*

[8]https://research.ncl.ac.uk/fasmed/positionpapers/The+use+of+technology+in+FA+to+raise+achievement_Revision+UNITO-FINAL.pdf.

students' interaction over longer timespans and over multiple sets of classroom participants (Roschelle and Pea 2002).

These potentialities were explored in the variety of approaches adopted across the partners in the project.

14.2 The FaSMEd Framework

During the first year of the project, time was allocated to establish a common understanding of the key concepts of FaSMEd. These were articulated through a series of position papers[9] and an agreed glossary.[10]

We recognized that an approach to learning through active participation in, and reflection on, social practices, internalisation and reorganization of experiences to activate pre-existing concepts and ideas would be desirable. Hence, FaSMEd activities stimulate 'conflict' or 'challenge' to promote re-interpretation, reformulation and accommodation. (See FaSMEd position paper.[11]) In the case of under-achieving students, the aim was to re-engage them rather than attempt to re-teach previously learned material. In addition, the aim was to devolve problems to learners so that learners articulated their own interpretations and created their own connections. (See Swan and Foster (this volume), for further details of some of the activities used.)

Partners were encouraged to create and adopt activities from their own contexts which reflected this approach to learning. However, because this approach could increase the cognitive load for students, it was important that the learning environment was engineered to support students and FaSMEd included technology as part of the design of the environment to provide such support. The FaSMEd project case studies provide examples of where this approach has worked successfully with under-achieving students. (See Cusi et al. (this volume) for a detailed example.)

Wiliam and Thompson (2007) focus on three central processes in teaching and learning: (a) Establishing where the learners are in their learning; (b) Establishing where the learners are going; and (c) Establishing how to get there. Considering all agents within the learning processes in a classroom (teacher, students and peers), they indicate that formative assessment can be conceptualized in five key strategies:

(1) Clarifying/Understanding/Sharing learning intentions and criteria for success,
(2) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding,
(3) Providing feedback that moves learners forward,

[9]https://research.ncl.ac.uk/fasmed/positionpapers/.

[10]https://microsites.ncl.ac.uk/fasmedtoolkit/theory-for-fa/glossary/.

[11]https://research.ncl.ac.uk/fasmed/positionpapers/Cognitive+conflict_Nottingham_ude_revised.pdf.

(4) Activating students as instructional resources for one another,
(5) Activating students as owners of their own learning.

The key strategies by Wiliam and Thompson (2007) constitute the foundation of the theoretical framework that has developed within the FaSMEd project and intersect with the central processes. They represent, indeed, the starting point for the development of a three-dimensional framework (see Fig. 14.1) aimed at extending their model to include the use of technology in formative assessment processes.

The FaSMEd framework (see Fig. 14.1) considers three main dimensions which enabled the project team to characterize technologically enhanced formative assessment processes: (1) the five key strategies of formative assessment introduced by Wiliam and Thompson (2007); (2) the three agents that intervene in the formative assessment processes and that could activate these strategies, namely the teacher, the student and the peers; and (3) the functionalities of technology.

The third dimension, **Functionalities of Technology**, was introduced with the aim of highlighting how technology could support the three agents involved in formative assessment processes when they activate the different formative assessment strategies. The *functionalities of technology* are subdivided into three categories: sending and displaying, processing and analysing, and providing an interactive environment. This subdivision was based on the FaSMEd partners' experience in the use of technology to support formative assessment processes.

The ***Sending and Displaying*** category includes those functionalities of technology that support communication and fruitful discussions between the agents of formative assessment processes. For example, the teacher can send questions to the students or display a student's screen to show his/her work to the whole class. Several other functionalities, such as sending messages, files, answers, or displaying screens or students' worksheets, belong in this category.

The functionalities that support the agents in the processing and analysis of the data collected during the lessons are included in the category ***Processing and Analysing***. This could include software that generates feedback based on a learner's answer or an application which creates statistical overviews of solutions of a whole class, e.g. in a diagram or table. Other examples are the generation of statistics of students' answers to polls or questionnaires as well as the tracking of students' learning paths.

The third category, ***Providing an Interactive Environment***, refers to those functionalities of technology that provide a shared interactive environment within which students can work individually or collaboratively on a task, or a learning environment where mathematical/scientific content can be explored. This category includes, for example, shared worksheets, Geogebra© files, graph plotting tools, spreadsheets, dynamic representations, or ChemSketch© models.

Figure 14.1 highlights how the subdivision of each dimension into different sub-categories identifies small cuboids within the diagram. Each cuboid helps to locate specific formative assessment practices, highlighting the agents involved in this practice, the main formative assessment strategies that are activated, and the functionalities of the technology that are involved. The framework has been used to

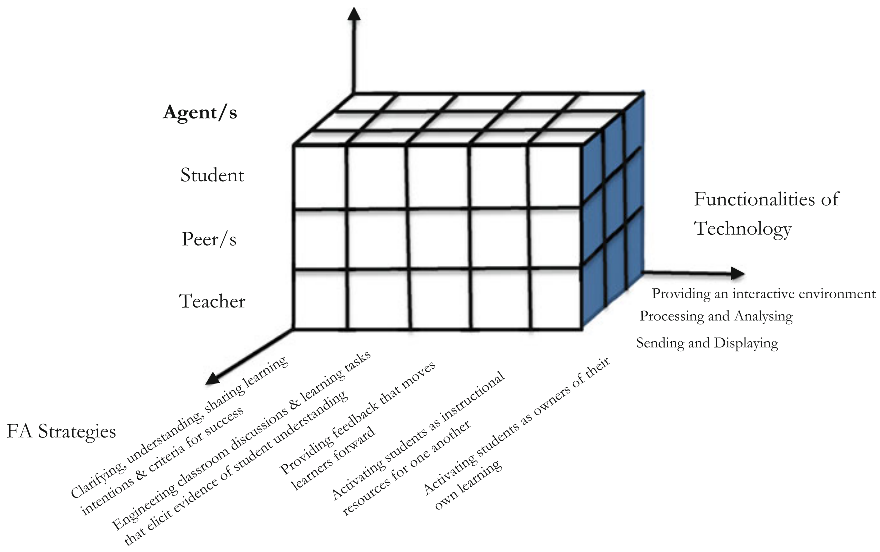


Fig. 14.1 Overview of the FaSMEd framework

identify and locate each of the cases reported by the partners in the project and has been the focus of several published papers and presentations at international conferences.[12] (Cusi et al. in this volume provide further discussion of the use of this framework.)

14.2.1 *Sending and Displaying in Practice*

A school working with Newcastle University implemented interactive whiteboards with Reflector[13] technology into classrooms. While students worked on the activity *Designing Candy Cartons*[14] on their iPads, the technology made it possible to display a student's screen to the whole class, sharing his/her work, and making it possible to annotate and comment visibly in real time.

Another example is provided by the University of Maynooth. They used Schoology[15], a learning management and social network system in classrooms as a way for teachers and students to communicate: sharing materials, uploading their work, teachers sending out tasks, but also to give feedback and ask questions.

[12]See our dissemination activities at: https://research.ncl.ac.uk/fasmed/disseminationactivity/.

[13]http://www.airsquirrels.com/reflector/.

[14]http://map.mathshell.org/lessons.php?unit=6300&collection=8.

[15]https://www.schoology.com/.

14.2.2 Processing and Analysing in Practice

In the tool *Unit of length* developed by the University College of Trondheim, Norway, the applet Kahoot[16] is used for sending questions to students, sending their answers to the teacher, and the teacher displaying the students' solutions to discuss and give feedback. What is more, the technology produces a statistical overview represented in a bar diagram of the whole class' answers (see Fig. 14.2), and therefore, helping students and the teacher to grasp all students' solutions at once.

Another example of this functionality is the tool *Equivalence of fractions* developed at Ecole Normale Superieure De Lyon, France. It uses a student response system ('Je leve la main'[17]) to display a question to the whole class, which each learner then answers individually via a remote control. Then, the technology analyses the answers, indicating in green or red color whether a student's solution was correct, and displays the answer of each individual student (see Fig. 14.3). The teacher can finally display all the sent-in solutions to discuss the problem with the whole class and give feedback.

14.2.3 Providing an Interactive Environment in Practice

The digital self-assessment tool *Can I sketch a graph based on a given situation?*[18], developed at the University of Duisburg-Essen, functions as an interactive environment, in which students can explore the mathematical content of sketching a graph dynamically and assess their own work based on a presented check-list (see Fig. 14.4).

Another example of technology used for formative assessment is the functionality of Providing an Interactive Environment (see Fig. 14.5), designed by The Freudenthal Institute at Utrecht University. They created four different modules in an online Digital Assessment Environment (DAE).[19] Within this environment, learners work on a series of questions while being able to choose between several different tools to help them solve a problem, for example, tables, scrap papers, hints, and percentage bars. The technology then presents an overview of the students' work, their chosen tools, and answers to the teacher, who can use this data formatively.

[16]https://getkahoot.com/.

[17]https://www.jelevelamain.fr/en/.

[18]https://sourceforge.net/projects/fasmed-self-assessment-tool/ (TI-Nspire© software needed).

[19]https://app.dwo.nl/dae/ (use Firefox browser).

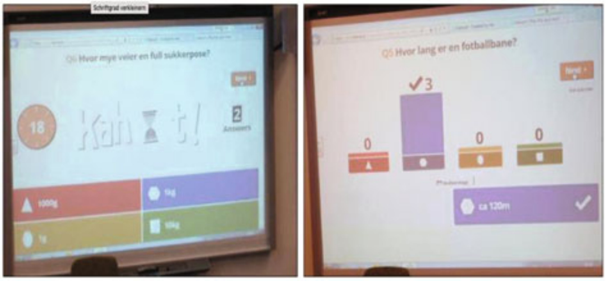


Fig. 14.2 Display of Kahoot!©

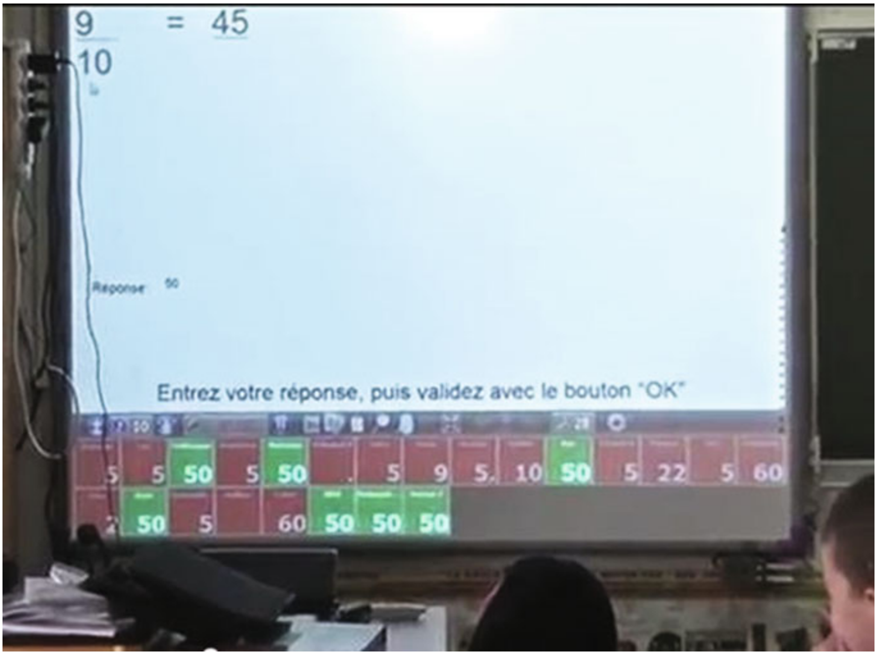


Fig. 14.3 'Je leve la main©' display for class

14.3 Raising Achievement

The FaSMEd project's aim was to address low achievement in science and mathematics education. Two surveys were completed by the partners at the beginning of the project to establish an understanding of the issues across the partnership. The first was to map the 'landscape' for low achievers in science and mathematics

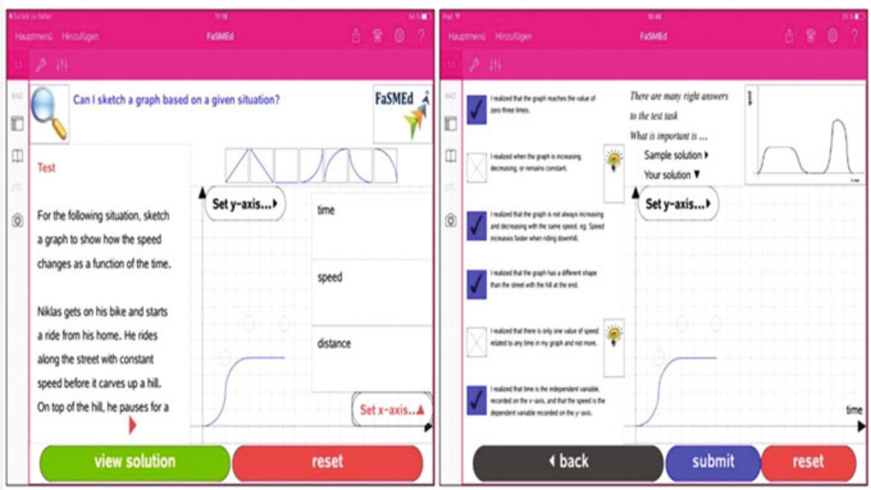


Fig. 14.4 TI-Nspire© app developed at the University of Duisburg-Essen on a tablet

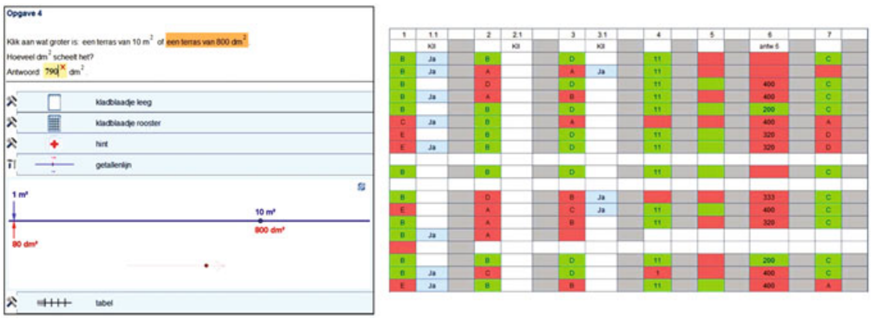


Fig. 14.5 Digital Assessment Environment showing the question (on left) for an individual student and the teacher's view of class responses to a range of questions (on right)

across the partner countries and their typical learning trajectories[20]; the second was to survey the systemic practices of partner countries for addressing the needs of low-achieving students.[21]

A disproportionate number of underachieving students come from disadvantaged social, cultural and ethnic groups, and in some countries, from groups without a good command of the first language of the classroom (Boaler et al. 2000; Ireson and Hallam 2001). Established approaches for working with such students are frequently characterised by a *deficit* model of their potential, which entails repeating

[20]http://research.ncl.ac.uk/fasmed/deliverables/D2%201.pdf.

[21]http://research.ncl.ac.uk/fasmed/deliverables/D2%202.pdf.

material from earlier years, broken down into less and less challenging tasks, focused on areas of knowledge in which they have previously failed and which involve step-by-step, simplified, procedural activities in trivial contexts. In contrast, the TIMSS seven-nation comparative study shows that high-achieving countries adopt approaches which preserve the complexity of concepts and methods, rather than simplifying them (Hiebert et al. 2003). In addition, evidence indicates that attitudinal factors on the part of both students and teachers can have a powerful impact on achievement, particularly with this group of students. Hence, FaSMEd partners were encouraged to develop resources, processes, and technological tools which would allow all students to engage with complex concepts and methods successfully and to improve motivation.

The FaSMEd project was based on the evidence that formative assessment strategies can raise levels of achievement for students (Black and Wiliam 1996). The project also builds on the evidence of research from, for example, the LAMP (Ahmed 1987), RAMP (Ahmed and Williams 1991), and IAMP (Watson et al. 2003) projects in mathematics teaching and the CASE (Shayer and Adey 2002) project in science teaching in the United Kingdom and elsewhere, which adopted approaches focused on the proficiencies of the students rather than their deficiencies. These projects adopt what Shulman (2002) calls *pedagogies of engagement*, characterised by revisiting student thinking, addressing conceptual understanding, examining a task from different perspectives, critiquing approaches, making connections, and engaging the whole class.

The main objectives for the project were to produce (through design research) a toolkit for teachers and teacher educators to support the development of practice (the expression 'toolkit' refers to a set of curriculum materials and methods for pedagogical intervention) and produce a professional development resource that exemplifies use of the toolkit through a process of design research. The toolkit is disseminated through a website produced by the partners[22]. Partners were encouraged to identify activities in science and mathematics which built on recent meta-analyses of the accumulated corpus of research on effective teaching that examined teaching components in mathematics and science (Seidel and Shavelson 2007), teaching strategies in science (Schroeder et al. 2007), and teaching programs in mathematics (Slavin and Lake 2008; Slavin et al. 2009). These provide clear indications of the relative effectiveness of some teaching components and the toolkit has numerous examples of such materials and approaches. (See also Swan and Foster (this volume) for further examples.)

[22]http://www.fasmed.eu/.

Table 14.1 Professional development modules in FaSMEd

Professional development modules	
Module number	Content
1	Introducing formative assessment
2	Using students' mistakes to promote learning
3	Improving questioning
4	Improving student collaboration
5	Students becoming assessors
6	Using technology for formative assessment

14.4 Professional Development Resources

The professional development package produced by FaSMEd reflects the range of ways in which partners have worked with teachers in their countries and offers examples for teachers and teacher educators to use. These include a set of six professional development modules[23] (see Table 14.1) designed to help teachers use Formative Assessment more effectively in their classrooms.

The resources also include a theoretical section on principles for effective professional development and a practical section on ways in which professional development can be organized. This section[24] is for people who are organizing professional development for teachers of mathematics and science but can also be used by teachers either individually or working with peers.

The FaSMEd position paper on the professional learning of teachers[25] warned that professional development (PD) is perceived and experienced differently across countries. Partners were aware, therefore, that it was important not to make assumptions about expectations and norms in other countries. However, the position paper then concludes that there is a high degree of convergence in descriptions of successful professional learning and the partners generally agreed. Typically, these include securing interest and engagement from the teachers, providing a theoretical framework for understanding of the innovation/strategy/programme, and offering some practical tools to apply to classroom practice.

The position paper also notes that *Professional Learning Communities* (PLCs) emerge as one of the most promising structures for professional learning. This is because the conditions for effective professional learning fundamentally require teachers to feel safe to experiment, to examine the impact of their innovations, to talk openly, and to establish principles about effective student learning. Partners were thus encouraged to engage with groups of teachers who were willing to

[23]https://microsites.ncl.ac.uk/fasmedtoolkit/professional-development/modules-new/.

[24]https://microsites.ncl.ac.uk/fasmedtoolkit/professional-development/.

[25]http://research.ncl.ac.uk/fasmed/positionpapers/TeacherProfessionalLearningPositionPaperRevised_Final.pdf.

collaborate as active participants in the design process of the resources for the toolkit and to support PLCs where possible.

Case studies produced by partners[26] in the project illustrate examples of teachers using the resources. The cross-comparison studies (both cases and countries) further provided an analysis of Formative Assessment practices across the partner countries.[27]

In FaSMEd, all partners used *active* involvement of the teachers in the design-based research process as professional development. Teachers were involved through cluster meetings and school visits throughout the intervention phase of the project (2014/2015). These meetings included dialogues with the FaSMEd researchers, sharing of practice with other teachers as well as participating in the 'design-do-review' cycles of classroom materials. The organization of the approach was different for each FaSMEd partner but essentially fell into three main types: courses; learning groups; and individual teachers.[28]

The *courses* varied considerably in most aspects. However, usually one or more experts led the course, choosing the content to cover and the order in which it was covered. The experts planned the sessions and prepared materials, such as handouts for teachers. A course had a beginning and an end and consisted of a series of meetings, during which the participants covered the course content which was pre-determined by the leader(s) of the course.

We used the term *learning groups* as an umbrella term which covered groups of teachers working together on their own professional development. Some of these groups called themselves professional learning communities, while others might be communities of practice; in general, the idea of these groups was that the members of the group learn together, usually by examining and inquiring into their own developing practice. Typically, the agenda was set by the group and there was no particular leader, although there may have been someone or some people who took responsibility for coordinating the group.

Working with *individual teachers* involved utilizing a professional development 'expert', tailoring the professional development to the needs of the individual teacher. For example, if they wanted to try something new in a particular context, the expert could plan a lesson with them, watch them teach, and talk to them about the lesson afterwards. They could also plan their own lesson, and again, the expert could watch them teach and discuss with them afterwards. The discussion with the teacher helps them to reflect deeply, not only on the way they taught the lesson, but also on their students' responses. It can also help them think in detail about the design of the task and the classroom lesson. For many teachers, this process leads to significant learning.

[26]http://research.ncl.ac.uk/fasmed/deliverables/.

[27]http://research.ncl.ac.uk/fasmed/deliverables/Deliverable%20D5.2%20Cross%20Comparative%20study%20of%20case%20studies.pdf.

[28]https://microsites.ncl.ac.uk/fasmedtoolkit/professional-development/approaches/.

14.5 Socio-technical Approaches to Raising Achievement in Mathematics and Science Education

Through FaSMEd, consortium partners (and teachers) engaged in the design process of socio-technical approaches aimed at raising achievement in mathematics and science education. Our schools and teachers, nevertheless, used many different technological tools in their mathematics and science classrooms, and worked under different conditions and environments. Hence, a true comparative analysis was not possible, as many variables changed with the use of different tools, change of environment, etc. Our intention was not to compare teachers internationally, but rather to develop deeper insights into how formative assessment strategies (particularly technology-based) help teachers and students follow better learning trajectories in a range of contexts.

The main findings from the cross-case study analysis (D5.2[29]) are as follows:

- The technology can provide immediate feedback, potentially useful for teachers and students. However, the usefulness depends on teachers' skills to benefit from it, as they often do not know how to helpfully build the formative feedback into their teaching.
- The technology potentially provides, and seems to encourage, ample opportunities for classroom discussions. Moreover, it appears that the technology helps to develop more cooperation within the class: teacher-student cooperation; and cooperation between individual students/within groups.
- Technology appears to provide an objective and meaningful way for representing problems and misunderstandings.
- Technology can provide opportunities for using preferred strategies in new or different ways.
- The technology helps to raise issues with respect to formative assessment practices (for teachers and students), which were sometimes implicit and not transparent to teachers. In nearly all the cases, the connection of formative assessment and technology tools helped teachers to re-conceptualize their teaching with respect to formative assessment.
- Different technological tools provide different outcomes. In principle, each tool can be used in different ways, and hence the processes of feedback to the individual, feedback to groups of students, feedback to whole class and discussion, are significant. Often a mix of technology was used, and the orchestration of the technology tools needed specific skills.

Looking across the cases, the technology tools provided immediate feedback for teachers about pupils' difficulties and/or achievement with a task. For example, the software 'Je leve la main' (see Fig. 14.3) provided opportunities for collecting and processing students' responses, and subsequently for further analysing individual

[29]https://research.ncl.ac.uk/fasmed/deliverables/Deliverable%20D5.2%20Cross%20Comparative%20study%20of%20case%20studies.pdf.

student work. As another example, a mathematics teacher mentioned that "*other effective moments are the polls, since they are immediate and interesting*." (Socrative©[30] was adopted in several schools in the United Kingdom for polling.)

We found that teachers see the technological tools as opportunities, such as opportunities for changing practices in the sense that teachers expanded their repertoire of strategies with the technological tools:

> [Before FaSMEd] the use of Formative Assessment was implicit. I had very low awareness of it. No specific tool was constructed or used for this purpose. [Now Formative Assessment is] gathering information at all steps of the teaching act.

Technological tools were also opportunities for adapting their preferred strategies in new or different ways. For example, one teacher reported that the tablet made her work more cooperatively with her class and removed her from the constraints of the whiteboard:

> It just means that I'm not at the front all the time.

Another teacher, who used questioning as his predominant approach, was aware that:

> not all students are comfortable to answer questions vocally or to be putting their hands up [….] sometimes you have to use other methods that are not as intrusive, things like using mini whiteboards where everyone can respond and no-one feels under pressure.

The tool (or resource), such as a 'clicker', can become an instrument for a Formative Assessment strategy as outlined by Wiliam and Thompson (2007)

(1) Clarifying learning intentions and criteria for success
(2) Engineering effective classroom discussions and other learning tasks that elicit evidence of student understanding
(3) Providing feedback that moves learners forward
(4) Activating students as instructional resources for one another
(5) Activating students as the owners of their own learning.

Within our cases, formative assessment practices were associated with functionalities of the technology tool/s: for example, with **sending and displaying** questions; and with **displaying students' answers**.

In our cases, several teachers reported that technical difficulties, e.g. setting up the technology for handling it with students, prevented them from using the technology tools more often. However, once they managed the tools successfully and saw the advantages of using them for formative assessment, they regarded them as beneficial both for their instruction and for student learning. One teacher commented:

> [Before FaSMEd]. The collection of information was done through conventional controls, activities at the beginning of the lesson, oral exchanges, observations of students in their activities. The quality and consistency of the treatment of such information varied widely.

[30]https://www.socrative.com/.

> There were some technical difficulties related to the handling of the material, during the first two months of the FaSMEd project. [But] Today I see only advantages of using digital technologies for formative assessment.

For students, there was an appreciation of the value of formative assessment. Through sharing and explaining work the teacher would "*know you haven't just copied, because if you had copied then you wouldn't have been able to explain the answer*". One student did explain that it was important not to be judged or humiliated. The classroom culture created by the teacher would therefore appear to be crucial if 'in the moment' formative assessment strategies are adopted (i.e., students need to feel it is safe to explain their ideas even if they might be wrong):

> If you're in class and you're doing a question on the tablet, if you get something wrong it's easier to tell than just writing it in your copy where you only can see, then the whole class can see and tell you where you went wrong.

Students thought that the technology also helped teachers to get a better (i.e., objective and observable) overview of how students were progressing:

> well, [teachers] can see what we've done better, it's hard to explain, if we do stuff on technology they can save it … they can see it …

Representing their knowledge in a meaningful way was perceived to be especially beneficial to low-achieving students, as it allowed them to represent their learning pictorially. Students could make sense of images and videos within an application (e.g., iPad application Popplet[31] or Classflow[32]).

Some students reported that working with these technology tools helped them to improve their learning, and facilitated their understanding of mistakes. It was reported that, after FaSMEd, students changed their minds on the utility of using clickers in math and science lessons, for example, using the projected answers for discussions with respect to their own results/answers. Selected students reconsidered the status of mistakes for their learning; they realized that mistakes could be useful in the learning process:

> You made a mistake, that's all, but [now] you know that you have understood.

Nearly all the case studies reported on the positive effect of technology in terms of facilitating and encouraging classroom discussions, either between teacher and students, or among students. Many students appeared to have had ample opportunities for peer interactions, partly due to the technology, in terms of paired discussions; students compared samples displayed, interpretations and strategies from peers, suggestions from peers, solutions, working and explanations from peers.

All the case studies reported an impact on student motivation and engagement. One teacher reported:

[31]http://popplet.com/.

[32]https://classflow.com/.

> I feel that my students are more confident in approaching unfamiliar tasks. They are more likely to 'have a go' at a task. The need to share work with their partner and to improve their own work, has helped them to appreciate the need to get something down on paper and to try things out. It has also helped their accountability in needing to complete a task, rather than just saying 'I don't know what to do'.

In some cases, teachers reported increased engagement and an improvement in the quality of student work due to the key role that technology played in displaying their work to their peers:

> If they know that they are going to have to present their work to the rest of the class they make much more effort with it.

In other words, it was not the technology itself, but the knowledge that the technology *could* be used which had an impact on the quality of some students' work.

14.6 Conclusions

The main objective for the project was to produce a toolkit for teachers and teacher educators and a professional development resource, implemented through a website, which would support their application of formative assessment strategies using technology. This output[33] is now ready for teachers and others to use. This resource now sits alongside the resources developed by SAILS[34], MASCIL[35], PRIMAS[36] and ASSIST-ME.[37]

It was found that most mathematics and science teachers in our study were not familiar with processing data (from students) for formative purposes using a range of technologies. In short, despite the widely recognized powerful impact of formative assessment on student achievement, formative assessment-utilising technology is not yet well developed. Improvement is required both in terms of ergonomics, with respect to the technology tools, as well as teacher professional development to build such tools into formative assessment instructional practices.

Through the FaSMEd project, selected teachers managed to build the formative assessment tools into their teaching, and reported a desire to embed these practices in future teaching. However, while most teachers used the technology to collect and store data, it was not always used formatively in subsequent steps. It became clear that unless teachers were experienced and confident teachers of mathematics/science, the combination of formative assessment practices and technology for

[33] http://www.fasmed.eu/.

[34] http://www.sails-project.eu/.

[35] http://www.mascil-project.eu/.

[36] http://www.primas-project.eu/.

[37] http://assistme.ku.dk/.

becoming more informed about student learning and understanding was challenging. Overall, the full potential of FaSMEd activities and tools has not been realized in most partner countries at this stage.

For teachers involved in the project, there were several challenges. In addition to introducing one of a variety of technologies into their classrooms, they were also asked (in most cases) to adapt to both a pedagogy of engagement and a pedagogy of contingency (Wiliam 2006). The challenge to adopt a pedagogy of engagement caused tensions. For example, (as recognized in the cross-country report, Deliverable D5.3[38]), anxieties about performance in both mathematics and science are raised by the way in which governments and school management interpret international test results. Hence, *productivity* and *performance* come into conflict with a pedagogy of engagement where reflective periods for examining alternative meanings and methods are required. Indeed, it must be recognised that FA requires teachers to prioritise learning over teaching—and that learning takes time, whereas teachers have a limited amount of time in which to deliver the required curriculum content.

The adoption of a pedagogy of contingency challenged teachers to translate formative *intention* into formative *action*. Although all teachers appreciated the information about students' learning being made visible through the activities, assessment opportunities and technological support, some teachers found it difficult to use the information to adjust their teaching to better meet their students' learning needs. This was particularly true when the information was generated in the middle of an active lesson.

However, in schools where the leadership created time for teachers to come together regularly and frequently to plan, discuss and review, the teachers were generally much better equipped to engage with these challenges and tensions. Hence, it could be argued, time for professional development is a necessary (but not sufficient) prerequisite for successful innovation in the classroom.

Where the conditions for professional learning are good, there is some evidence that practices trialled through FaSMEd were being embedded into teachers' pedagogy and general classroom practice. For example, one teacher from the United Kingdom said:

> The information gleaned from the pre-assessment tasks has always proven to be invaluable in finding out where the stumbling blocks for the students are and where teacher intervention is required. While the barrier for completing the task is sometimes similar for all students – and where I would have probably expected – occasionally it has thrown up surprises. This is a highly transferrable strategy which I plan to use before all units of work to inform my planning for the group. (See Swan and Foster (this volume) for further discussion on this.)

Another teacher commented:

> The FaSMEd project has reinvigorated my every day teaching and made me think about how I approach lessons and their structure. I am already starting to use photographs of

[38]http://research.ncl.ac.uk/fasmed/deliverables/.

> students' work (displayed anonymously) to aid discussion and model working out/ explanation. I already do a lot of pair work, but I am thinking more carefully about which students are paired together and I'm trying to mix students up more.

Regarding the students, the investigations (and interventions) have shown a relatively positive picture: students seemed to welcome the formative assessment data provided by the technology (and the teacher/s) and they were ready to usefully build it into their learning strategies. Overall, promising patterns of engagement and motivation were identified, particularly for low-achieving students.

In conclusion, while acknowledging the complex and challenging environments in schools, the FaSMEd activities combined with the appropriate technological tools appear to have the potential to enhance the learning process. This can be achieved through active engagement with the FaSMEd Toolkit and rigorous professional learning, as exemplified through the FaSMEd Professional Development Package.

14.7 Recommendations for Policy and Research

The final deliverables for FaSMEd are two papers (D6.2 and D6.3)[39] with recommendations for policy and research in this field. We summarize the major elements here.

14.7.1 Policy

The FaSMEd project found that the introduction of innovative technology to create a digital environment (between students, peers and teachers) can assist teachers in making more timely formative interpretations. We recommend the use of such technologies within classrooms to further enhance formative assessment practices.

Through the case studies, there is evidence of teachers using technologies to gain information about their students' thinking, as well as to facilitate opportunities for students to learn from their peers. In interviews, students identified these practices as particularly beneficial in making their learning visible to the teacher, themselves, and their peers. We recommend that technologies are utilized within classrooms to facilitate making learning more visible to all 'in the moment'.

Our FaSMEd case studies show that most teachers opted for technology tools which were accessible and/or easy to learn how to use and apply in their classrooms. We would therefore recommend that when embarking on new technological innovations, the usability of tools is considered.

[39] http://research.ncl.ac.uk/fasmed/deliverables/.

FaSMEd found that where existing infrastructures supported the use of technology, schools could make considerable progress in their use of technology to support formative assessment practices. We would recommend investment in networking and wireless systems, together with technical support in schools. FaSMEd believes this is a priority and a prerequisite for the implementation of this technology on a larger scale.

Where teachers could work as professional learning communities, conditions were effective in enabling them to feel safe to experiment, to examine the impact of their innovations, to talk openly, and to establish principles about effective student learning. FaSMEd would therefore recommend that schools (wherever possible) facilitate time and space for teachers to plan and reflect on their practice. A commitment to this from school leaders is crucial.

14.7.2 Research

The FaSMEd framework provides a conceptual model for understanding the interactions within the classroom between agents (teachers, students, peers), formative assessment strategies, and functionalities of technology. This model enables researchers to analyse the role of technology within formative assessment and learning. FaSMEd partners are continuing to develop and apply the framework to their research.

In addressing the needs of lower achievers, several interventions used technologies that were more easily accessible and did not demand high levels of literacy. Using visual displays of students' work was shown to be useful in some circumstances. For example, it appears that the simple awareness on the part of students that their work *could* be displayed for their peers impacted on its quality because they reviewed their work from another point of view. This encouraged lower-achieving students to engage more fully in tasks and therefore *Activated students as owners of their own learning*. We recommend that future research explores further applications that support visual displays in mathematics and science.

Many of our teachers used technology with polling systems to gather evidence of student learning. Multiple-choice questions have become one of the ways teachers seek out feedback on the understanding of their students, but these need careful framing, interpretation, and response by the teacher. One problem is that single response multiple-choice questions may not give a very accurate indication of students' understanding if a significant number choose the right (or wrong) answer at random. A more accurate use of multi-choice would be to design questions where the correct answer is to select two (or more) choices simultaneously—thus reducing the probability of random choice being correct and a richer selection of information. Research is needed to develop these questions.

Another issue is that current assessment and polling software often aggregate the data from groups of students, but do not do any further processing. Interpreting and reacting to such data is one of the major challenges for teachers. We recommend that future research into technology that can work intelligently with student responses, recognize common errors, and suggest strategies and/or feedback is needed.

The research findings indicate that using technology to support teacher-mediated formative assessment calls for high levels of teacher expertise. In contrast, the most effective technology applications exemplified that: '*The first fundamental principle of effective classroom feedback is that feedback should be more work for the recipient than the donor*' Wiliam (2011, p. 138). In other words, where the task environment provides direct feedback to the student, such feedback is readily interpretable by the student and the necessity for the teacher to interpret and mediate is reduced. Research should thus support the development of such task environments.

The main objective for FaSMEd was the development of a Toolkit for teachers and a Professional Development package to support it. During the three-year project, a prototype toolkit was developed and evaluated, leading to the production of the final toolkit. However, this resource has not been evaluated and it remains an open question about the extent to which a website incorporating the resource will be used or valued by teachers. Hence, it is clear that, to ensure that the FaSMEd toolkit is fit for its purpose, a further iteration would be required, including feedback from teachers on the use of the resources.

The FaSMEd toolkit now sits alongside a range of research projects which were commissioned by the European Commission (EC), for example: SAILS—Strategies for Assessment of Inquiry Learning in Science[40]; MASCIL—Mathematics and Science for Life[41]; PRIMAS—Promoting Inquiry in Mathematics and Science Education across Europe[42], and ASSIST-ME—Assess Inquiry in Science, Technology and Mathematics Education.[43] In relation to mathematics and science education, then, there is clearly a great wealth of research and knowledge generated across Europe (and beyond). Although at project level there has been some knowledge exchange and collaboration, more needs to be done to ensure that cross-project findings are integrated and translated into research, policy, and practice. We recommend that such meta-analysis is essential for future research.

[40]http://www.sails-project.eu/.

[41]http://www.mascil-project.eu/.

[42]http://www.primas-project.eu/.

[43]http://assistme.ku.dk/.

References

Abrahamson, L., Davidian, A., & Lippai, A. (2002). *Wireless calculator networks—Why they work, where they came from, and where they're going*. Paper presented at the 13th Annual International Conference on Technology in Collegiate Mathematics, Atlanta, Georgia.

Ahmed, A. (1987). *Better mathematics*. London, England: Her Majesty's Stationery Office HMSO.

Ahmed, A., & Williams, H. (1991). *Raising achievement in mathematics project*. London, England: Her Majesty's Stationery Office HMSO.

Ares, N. (2008). Cultural practices in networked classroom learning environments. *Computer-Supported Collaborative Learning, 3,* 301–326.

Black, P., & Wiliam, D. (1996). *Inside the black box*. London, England: King's College School of Education.

Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experience of ability grouping—Disaffection, polarization and the construction of failure. *British Educational Research Journal, 26,* 631–649.

Cusi, A., Morselli, F., & Sabena, C. (this volume). The use of digital technologies to enhance formative assessment processes. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), ICME 13 Monographs *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 77–92). Cham, Switzerland: Springer International Publishing AG.

Dufresne, R. J., Gerace, W. J., Mestre, J. P., & Leonard, W. J. (2000). *ASK-IT/A2L: Assessing student knowledge with instructional technology* (technical report No. UMPERG-2000-09). Amherst, MA: University of Massachusetts Physics Education Research Group.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 Video Study*. Washington, DC.: U.S. Department of Education National Center for Education Statistics.

Ireson, J., & Hallam, S. (2001). *Ability grouping in education*. London, England: Sage.

Irving, K. I. (2006). The impact of educational technology on student achievement: Assessment of and for learning. *Science Educator, 15*(1), 13–20.

Looney, J. (2010). Making it happen: Formative assessment and educational technologies. *Thinking Deeper Research Paper n.1, Part 3*. Promethean Education Strategy Group.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Research Council, Washington, DC: National Academy Press

Rocard, M., Csermely, P., Jorde, D., Lenzen, D., Walberg-Henriksson, H., & Hemmo, V. (2007). *Science education NOW: A renewed pedagogy for the future of Europe*. Luxembourg: Office for Official Publications of the European Communities.

Roschelle, J., & Pea, R. (2002). A walk on the WILD side. How wireless handhelds may change computer-supported collaborative learning. *International Journal of Cognition and Technology, 1*(1), 145–168.

Roschelle, J., Penuel, W. R., & Abrahamson, L. (2004). The networked classroom. *Educational Leadership, 61*(5), 50–54.

Roschelle, J., Tatar, D., Chaudhury, S. R., Dimitriadis, Y., & Patton, C. (2007). Ink, improvisation, and interactive engagement: Learning with tablets. *Computer, 40*(9), 42–48.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(2), 1436–1460.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis research. *Review of Educational Research, 77*(4), 454–499.

Shayer, M., & Adey, P. (2002). *Learning intelligence: Cognitive acceleration across the curriculum from 5 to 15 years*. Buckingham, England: Open University Press.

Shirley, M., Irving, K. E., Sanalan, V. A., Pape, S. J., & Owens, D. (2011). The practicality of implementing connected classroom technology in secondary mathematics and science classrooms. *International Journal of Science and Mathematics Education, 9,* 459–481.
Shulman, L. S. (2002). Making differences: A table of learning. *Change: The Magazine of Higher Learning, 34*(6), 36–44.
Slavin, R., & Lake, C. (2008). Effective programs in elementary mathematics. *Review of Educational Research, 78*(3), 427–515.
Slavin, R., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics. *Review of Educational Research, 79*(2), 839–911.
Swan, M. (2014). Design research in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 148–152). Dordrecht, The Netherlands: Springer.
Swan, M., & Foster, C. (this volume). Formative assessment lessons. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), ICME 13 Monographs *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 11–24). Cham, Switzerland: Springer International Publishing AG.
Watson, A., De Geest, E., & Prestage, S. (2003). *Deep progress in mathematics*. Oxford: University of Oxford Department of Educational Studies.
Wiliam, D. (2006). Assessment for learning: Why, what and how. In *Excellence in assessment: Assessment for learning* (pp. 2–16). Cambridge, England: Cambridge Assessment Network. Retrieved June 19, 2017 from http://www.assessnet.org.uk/e-learning/file.php/1/Resources/Excellence_in_Assessment/Excellence_in_Assessment_-_Issue_1.pdf.
Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.
Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Author Biographies

David Wright is Senior Research Associate: Research Centre for Learning and Teaching, Newcastle University (United Kingdom) (now retired). David has fifteen years' experience teaching mathematics at secondary, further and higher education as an associate lecturer with the Open University. He was Subject Officer for Mathematics for the British Educational Communications and Technology Agency (Becta) for four years and ten years in initial teacher education and research at Newcastle University. He is the Scientific Director of the European Union research project: Formative Assessment in Science and Mathematics Education (FaSMEd).

Dr. Jill Clark is the Principal Research Associate and Executive Director of the Research Centre for Learning and Teaching at Newcastle University in the United Kingdom. Jill joined the School of Education, Communication and Language Sciences in November 1996, having worked within the Newcastle Centre for Family Studies during the previous four years. Although now working in the field of educational research, she has a strong background in Social Sciences research. She studied Behavioural Sciences (majoring in psychology) at Huddersfield Polytechnic and then completed her postgraduate degree in Criminology at Cambridge University.

Lucy Tiplady is CfLaT Research Associate. Lucy joined the Education section at Newcastle University (United Kingdom) in 2005 after completing a Graduate Diploma in Psychology and BA (hons) English Studies. As a researcher within the Research Centre for Learning and Teaching, Lucy has developed a wealth of experience in project management, research design, literature reviews, qualitative, quantitative, visual and participatory research methods, data analysis and reporting research findings. Lucy continues to develop her research interests through involvement in several current research projects.

Part VI
Conclusion

Chapter 15
Looking to the Future: Lessons Learned and Ideas for Further Research

David Wright, Megan Burton, Annalisa Cusi and Denisse R. Thompson

Abstract This concluding paper to the volume highlights some lessons learned from the various papers relative to the issue of formative assessment and draws them together as a range of attempts to make students' learning visible. In addition, possible avenues for further research related to this important topic are discussed, including formative assessment as an instrument or a process, the development of tools for assessment, and a more nuanced understanding of classroom assessment.

Keywords Formative assessment · Research issues related to assessment

15.1 Introduction

Mathematical education benefits from a rich set of research-based descriptions of how learners develop concepts, problem solving, and reasoning competencies. Despite this, the assessment of students' learning in mathematics is complex, subject to continuing debate and cannot be taken for granted.

D. Wright (✉)
Research Centre for Teaching and Learning, Newcastle University, Newcastle upon Tyne NE1 7RU, UK
e-mail: wrightdavidg@gmail.com

M. Burton
Auburn University, 5020 Haley Center, Auburn, AL 36849 USA
e-mail: megan.burton@auburn.edu

A. Cusi
Department of Philosophy and Education, University of Turin, Via Gaudenzio Ferrari 9, 10124 Turin, Italy
e-mail: annalo@tin.it

D. R. Thompson
University of South Florida, Tampa, FL 33620, USA
e-mail: denisse@usf.edu

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*, ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5_15

As Bennett comments: "Formative assessment, like all educational measurement, is an *inferential* process because we cannot know with certainty what understanding exists inside a student's head… We can only make conjectures based on what we observe…" (2011, p. 16). (See also Pellegrino et al. (2001, p. 42).) Significant attempts to address these difficulties have been made. For example, in 1998 the National Research Council (NRC) of the USA, with the support of the National Science Foundation (NSF), convened a committee on the foundations of assessment. Its report, *Knowing What Students Know* (Pellegrino et al. 2001), provides a comprehensive overview of educational assessment at the time and sets out goals for educational assessment for the 21st century. It is now some sixteen years since the publication of this report and it may be instructive to draw on its conclusions to compare the current issues in classroom assessment in mathematics.

The report provides a clear statement of principle for all assessments:

> Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students' performance, and an interpretation method for drawing inferences from the performance evidence thus obtained. (Pellegrino et al. 2001, p. 2, italics in original)

It adds that "*These three elements—cognition, observation, and interpretation—must be explicitly connected and designed as a coordinated whole.* If not, the meaningfulness of inferences drawn from the assessment will be compromised" (p. 2, italics in original). In its focus on classroom assessment in particular, it advises: "*assessments, especially those conducted in the context of classroom instruction, should focus on making students' thinking visible to both their teachers and themselves so that instructional strategies can be selected to support an appropriate course for future learning*" (Pellegrino et al. 2001, p. 4, italics in original).

It could be argued that, despite the wide range of approaches to classroom assessment exemplified in this volume, they are connected through the attempts to address part or all of the 'pillars of assessment' and in making 'students' thinking visible' in the classroom. Examples from this volume include: Bostic & Sondergeld on validating a measure to vertically align 7th and 8th grade problem-solving instruments; Straumberger on self-assessment for improving mathematical competence; Sia and Lim on using Cognitive Diagnostic Assessment to assess primary students' knowledge of time; and Burton et al. on how professional development related to formative assessment should be intricately woven into the fabric of all professional development on effective instruction. Further, Andrade-Molina and Diaz Moreno provide an analysis of Chilean curriculum guidelines which, they argue, fail to address the pillars of assessment in regard to students' spatial learning.

Most importantly, the National Research Council report reminds us that assessment is not simply a technical issue but one of equity (Pellegrino et al. 2001) and that fairer assessments are needed to ensure that all students can achieve their potential. (See also OECD (2010) for further discussion of achieving equity in education.)

15.2 Formative Assessment—An Instrument or Process?

In the introduction to this volume, we quote Black and Wiliam (2009) who define assessment as formative:

> to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded than the decisions they would have taken in the absence of the evidence that was elicited. (p. 9)

However, in his critical review of formative assessment, Bennett (2011) refers to a 'split' in the understanding of formative assessment between those who think of formative assessment as an instrument, such as a diagnostic test, and those who hold the view that formative assessment is a process whose distinguishing characteristic is the feedback used to adapt teaching to meet student needs. Bennett attempts to resolve this issue by suggesting that formative assessment should be conceived: "*as neither a test nor a process, but some thoughtful integration of process and purposefully designed methodology or instrumentation*" (Bennett 2011, p. 7). In addition, he proposes that a definition of formative assessment needs a theory of action and a concrete instantiation:

> the theory of action: (1) identifies the characteristics and components of the entity we are claiming is 'formative assessment', along with the rationale for each of those characteristics and components; and (2) postulates how these characteristics and components work together to create some desired set of outcomes (Bennett 2010). The concrete instantiation illustrates what formative assessment built to the theory looks like and how it might work in a real setting. (Bennett 2011, p. 8)

Swan and Foster (this volume) describe an ambitious attempt at such an instantiation through the Mathematics Assessment Project,[1] and their examples of 'formative assessment lessons' demonstrate an integration of process and instrumentation. Wright et al. (this volume) and Cusi et al. (this volume) provide further examples of instantiations which attempt to build on the exemplars of 'formative assessment lessons' through the support of classroom technologies and professional development.

There appears to be evidence from these initiatives that adoption of approaches which combine a plan of action and an instantiation of resources which make students' learning visible does change classroom assessment. However, Fullan and Donnelly (2013) argue that three elements need to come together in lock-step for there to be an effective impact on education on a large scale: digital innovation, pedagogy and change knowledge:

> a simpler approach might be more helpful — one that is quite close to the strategising that will be required. There need to be policies and strategies that will simultaneously i) conceptualise and operationalise the new pedagogy; ii) assess the quality and usability of specific digital innovations; and iii) promote systemness. (Fullan and Donnelly 2013, p. 12).

[1]http://map.mathshell.org/.

Further research into the adoption of such resources and approaches might gain from these principles.

15.3 Tools for Assessment

In 2001, the authors of *Knowing What Students Know* were optimistic about the possible impact of technology on assessment:

> computer and telecommunications technologies offer a rich array of opportunities for providing teachers with sophisticated assessment tools that will allow them to present more complex cognitive tasks, capture and reply to students' performances, share exemplars of competent performance, engage students in peer and self-reflection, and in the process gain critical information about student competence. (Pellegrino et al. 2001, p. 307)

Laurillard notes that there has always been a strong relationship between education and technology, but that whereas: "Tools and technologies, in their broadest sense, are important drivers of education, their development is rarely *driven by* education" (2012, p. 2). Hence, like many other predictions for the benefits of technology for education, the impact of these innovations in the classroom has largely failed to materialize in the ensuing interval and the adoption of tools supporting assessment has been very slow. Fullan (2013, p. 37) for example, reports, "8% of teachers fully integrate technology into the classroom and only 23% of teachers feel that they *could* integrate technology in the classroom" (Moeller and Reitzes 2011 cited in Fullan 2013). Further, even where technology is present in the classroom, effective pedagogical application of it is missing (Fullan 2013). To benefit from the opportunities and affordances of technology for education, Fullan advocates an approach which integrates technological tools, pedagogy, and knowledge of how change occurs in education; in this way, it is hoped, education will begin to drive its use of technology.

Examples in this volume address this challenge. For example, FaSMEd creates a model which maps how the affordances of technology for assessment can interact with the actors (students, peers, teacher) in the classroom and the variety of strategies for formative assessment; Cusi et al. (this volume) and Wright et al. (this volume) provide case studies of these in action in the classroom. Evidence from several papers in this volume (Cusi et al.; Haddif and Yerulshamy; Platz et al.; Wright el al.) shows that technology has the potential to enhance classroom assessment[2] in as much as it:

- provides immediate feedback about the students' solutions, thus providing the teacher with an overview of the class' achievement of the target competence;

[2]See http://research.ncl.ac.uk/fasmed/positionpapers/The+use+of+technology+in+FA+to+raise+achievement_Revision+UNITO-FINAL.pdf.

- enables teachers to monitor students' incremental progress and keep them oriented on the path to deep conceptual understanding, providing appropriate remediation to address students' needs;
- supports positive students' thinking habits, such as arguing for their point of view;
- creates immersive learning environments that highlight problem-solving processes;
- gives powerful clues to what students are doing, thinking, and understanding;
- enables most or all of the students to contribute to the activities and work toward the classroom performance, therefore taking a more active role in the discussions;
- provides students with immediate private feedback, encouraging them to reflect and monitor their own progress;
- provides multi-level analyses of patterns of interactions and outcomes thanks to their potential to instrument the learning space to collect the content of students' interaction over longer time-spans and over multiple sets of classroom participants.

The FaSMEd project (Cusi et al. and Wright et al. this volume) demonstrates how tools and technology, for example classroom aggregation systems using polls and applications for displaying students' work for the whole class to view, can support teachers' assessments in the classroom. Case studies showed that technology tools provided immediate feedback for teachers about pupils' difficulties and/or achievement with a task. Technology also potentially provides, and seems to encourage, ample opportunities for classroom discussions. Further, the teachers used the technological tools as opportunities for changing practices, in the sense that teachers expanded their repertoire of strategies with the technological tools and adapted their preferred strategies in new or different ways.

Another example of the teacher's use of technology as an opportunity to adapt their teaching according to students' responses is within Downton's paper (this volume), which analyses the use of digital technology by teachers as tools to collect assessment data to document student learning. The papers by Platz et al. (this volume) and Haddif and Yerushalmy (this volume) both deal with the design of e-resources that could support formative assessment processes. Haddif and Yerushalmy focus on the design of e-assessments to highlight students' response space, revealing different characteristics of students' thinking. Platz et al. focus on the creation of an educational e-proof environment that supports and guides students to prove a given theorem.

Interpreting and reacting to the data collected and aggregated through technological tools are major challenges for teachers. Indeed, many teachers use such systems for 'long-cycle' formative assessment, in other words they store the data for further analysis and may respond in the following lesson or even later, often because it is very difficult to decide what to do minute-by-minute when using these technologies.

Current classroom systems are what Wiliam (2007) refers to as ‘2nd generation’. They aggregate the data from groups of students, but will not do any further processing. Teachers need support through research into developing technology which would work intelligently with student responses by parsing them in some way to recognize common errors and suggest strategies or feedback.

15.4 Issues for Assessment, Feedback and Learning

Pellegrino et al. (2001) identified three *pillars of assessment*—a model of cognition in the domain, tasks which make students’ learning visible, and a process of interpretation of the observed performance. Bennett (2011) also identifies three factors for effective assessment by teachers: pedagogical understanding of assessment practices, deep domain understanding, and knowledge of measurement fundamentals to support inference from the evidence elicited by assessment. The clear intersection between these two approaches indicates both the challenge for classroom teachers of carrying out effective assessment and the challenge of providing adequate professional development to enhance teachers’ practice.

In this volume, Sayac provides an analysis of teachers’ assessment practices in France, which demonstrates the inadequacy of the assessments developed by teachers where, for example, domain understanding is lacking. Pai (this volume) also identifies and analyses the pedagogical skills needed by a classroom teacher in attending to learners’ responses, which highlights the demanding nature of classroom assessment. Burton et al. (this volume) discuss how professional development could support teachers in developing effective formative assessment; they advocate that formative assessment should be woven into the fabric of all professional development on effective instruction.

Some contributions in this volume describe how tools can measure performance for both teachers and students themselves. The challenge, however, for both students and teachers, is to *interpret* their performance in terms of learning—‘making learning visible’. Bostic and Sondergeld (this volume), for example, focus on validating classroom-based measures to vertically align 7th and 8th grade problem-solving instruments, suggesting that they could be considered as a foundation for assessments that can be employed to derive meaningful feedback and inform instructional practices. A second example is proposed by Sia and Lim (this volume), who investigate the use of Cognitive Diagnostic Assessment as an informative tool to teachers when trying to determine next steps for teaching. Finally, Straumberger (this volume) focuses on the creation of tools for self-assessment. He examines the use of self-diagnosis sheets as part of classroom assessment aimed at improving mathematical competence.

A fruitful area for further research and development is on the use of multiple-choice questions, which have become one of the ways teachers seek out feedback on the understanding of their students. But these need careful framing, interpretation, and response by the teacher. For example, single response multiple-choice questions may not give a very accurate indication of students' understanding because there may be a significant number choosing the right (or wrong) answer at random. A better use of multi-choice would be to design questions where the correct answer is to select two (or more) choices simultaneously—this reduces the probability of random choice being correct and a richer selection of information (Wiliam 2007). However, these sorts of questions are harder to design, and further research is needed on this sort of development. In addition, currently available polling systems may not allow more than one simultaneous response.

15.5 Discussion and Conclusion

The concepts of summative and formative assessment and the evidence for the power of feedback for enhancing student learning have been available to the education community for decades (Black and Wiliam 1996). The ubiquitous presence of these concepts for such an extended period has had two simultaneous effects: On the one hand, it has led to an oversimplification of the ideas (assessments are either summative or formative and all feedback is beneficial); on the other hand, it has also led to a more thoughtful pushback against some of these oversimplified ideas, leading to a more nuanced understanding of both assessment tasks and feedback.

Some of this confusion lies in the difference between learning and performance since performance is what can be observed but learning must be inferred from performance. Pai (this volume) analyses the challenges faced by teachers making 'in-the-moment' decisions in the classroom. Further, research shows that learning can occur when there is no change in performance and conversely, that apparent success in performance does not necessarily imply progress in learning (Soderstrom and Bjork 2015). Indeed, an intriguing finding from research in performance and learning is that, in some circumstances, tasks which produce the least success in performance (inducing numerous errors, for example) produce the greatest progress in learning (Bjork 2017).

Similarly, Hattie & Timperley make the point that, "Feedback is one of the most powerful influences on learning and achievement, **but this impact can be either positive or negative**" (2007, p. 81, bold added for emphasis). This may seem obvious for feedback which might be poorly framed or timed. However, psychologists note that:

> One common assumption has been that providing feedback… fosters long-term learning to the extent that feedback is given immediately, accurately, and frequently. However, … empirical evidence suggests that delaying, reducing, and summarizing feedback can be better for long-term learning than providing immediate, trial-by-trial feedback… the very feedback schedules that facilitate learning can have negligible (or even detrimental) performance effects. (Soderstrom and Bjork 2013, p. 18)

Developments in formative assessment may increase the workload on teachers and the demands made on their domain knowledge and decision making in the classroom. However, as Wiliam recommends: "The first fundamental principle of effective classroom feedback is that feedback should be more work for the recipient than the donor" Wiliam (2011, p. 138). In other words, where the task environment provides direct feedback to the student, such feedback is readily interpretable by the student and the necessity for the teacher to interpret and mediate is reduced. The FaSMEd project found, for example, that the simple awareness on the part of students that their work *could* be displayed for their peers (through the use of projective technology) impacted on its quality, because they began to review their work from another point of view. This encouraged lower-achieving students to engage more fully in tasks and therefore *Activated students as owners of their own learning*. Hence, research on developing such task environments may be the most effective strategy for incorporating formative feedback into the classroom.

Finally, it is often assumed that although summative assessments are necessary to identify the attainment of students, they have relatively minor impact on student learning. However, the psychological construct known as 'the testing effect' (van Eersel et al. 2016) demonstrates that testing can have an impact on learning. Indeed, it appears that it is more effective to re-test learners than to re-teach—research shows that studying once and testing three times is 80% more effective than studying three times and testing once (Butler 2010). Hence, used in appropriate ways such activities can clearly be seen to be examples of 'assessment *as* learning'.

All the evidence from school effectiveness research points to the interactions which occur in the classroom to be the key to making a difference for students (Hattie 2009; Kane et al. 2013). Consequently, all around the world, teachers, to a greater or lesser extent, are under pressure to demonstrate their impact on the learning of their students.

The authors of the National Research Council report challenged teachers to change:

> the power offered by assessments to enhance learning in large numbers of classrooms depends on changes in the relationship between teacher and student, the types of lessons teachers use, the pace and structure of instruction, and many other factors. To take advantage of the new tools, many teachers will have to change their conception of their role in the classroom. They will have to shift toward placing much greater emphasis on exploring students' understanding with the new tools and then undertaking a well-informed application of what has been revealed by use of the tools. This means teachers must be prepared to use feedback from classroom and external assessments to guide their students' learning more effectively by modifying the classroom and its activities. In the process, teachers must guide their students to be more engaged actively in monitoring and managing their own learning—to assume the role of student as self-directed learner. (Pellegrino et al. 2001, p. 302)

Given such a radical disruption of teaching practice, it is hardly surprising that nearly two decades have passed without a widespread adoption of these practices. Michael Fullan, the architect of Ontario's prized education system, argues that policy makers too often focus on the accountability system when they should be focused instead on building the collective capacity of the system to improve, with teachers being the key agents of change (Fullan 2010). Building the capacity of the teaching profession should therefore be the main emphasis of schools' policy (IPPR 2013).

Bennett, it could be argued, draws on a design research approach to teaching (Laurillard 2012) and recommends that teachers:

> need time to reflect upon their experiences with these materials. If we can get teachers to engage in iterative cycles of use, reflection, adaptation, and eventual creation – all firmly rooted in meaningful cognitive-domain models – we may have a potential mechanism for helping teachers better integrate the process and methodology of formative assessment with deep domain understanding. (2011, p. 19)

In addition, Wiliam (2016) reports that the following factors impact on changing practice:

- Choice (Allowing experienced teachers to choose the aspect (out of the formative assessment strategies) of their practice for development);
- Flexibility (Teachers should be allowed to adapt or modify the strategy to work for them);
- Small steps (An incremental approach to change);
- Accountability (A commitment to improve practice through a written action plan);
- Support (Leaders in schools should provide time, space, dispensation, and support for innovation).

However, systemic change requires a significant amount of time, for example, the Learning Community Project in Mexico spread from an initial thirty schools to over six thousand, but it took over eight years (Rincon-Gallardo et al. 2012). Wiliam (2016) warns that it is common to find that a significant impact on standardized test scores might only materialize after two to three years of implementation of an innovation. Policy makers, therefore, must be prepared to plan for and sustain change over an extended time scale and be patient.

References

Andrade-Molina, M., & Diaz Moreno, L. (this volume). Assessing visualization: An analysis of Chilean teachers' guidelines. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 179–191). Cham, Switzerland: Springer.

Bennett, R. E. (2010). Cognitively based assessment of, for, and as learning: A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research and Perspectives, 8*(2–3), 70–91.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678.

Bjork, R. A. (2017). *How we learn versus how we think we learn: Desirable difficulties in theory and practice*. https://bjorklab.psych.ucla.edu/research/#itemII. Accessed July 01, 2017.

Black, P., & Wiliam, D. (1996). *Inside the black box*. London, England: King's College School of Education.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Bostic, J., & Sondergeld, T. (this volume). Validating and vertically equating problem-solving measures. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 139–155). Cham, Switzerland: Springer.

Burton, M., Silver, E., Mills, V., Audrict, W., Strutchens, M., & Petit, M. (this volume). Formative assessment and mathematics teaching: Leveraging powerful linkages in the US context. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 193–205). Cham, Switzerland: Springer.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology, 5,* 1118–1133.

Cusi, A., Morselli, F., & Sabena, C. (this volume). The use of digital technologies to enhance formative assessment processes. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 77–92). Cham, Switzerland: Springer.

Downton, A. (this volume). Using a digital flip camera as an assessment tool in mathematics lessons. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 63–76). Cham, Switzerland: Springer.

Fullan, M. (2010). *All systems go: The change imperative for whole system reform*. Thousand Oaks, CA: Corwin.

Fullan, M. (2013). *Stratosphere: Integrating technology, pedagogy, and change knowledge*. Toronto, Canada: Pearson.

Fullan, M., & Donnelly, K. (2013). *Alive in the swamp: Assessing digital innovations in Education*. London, England: Nesta.

Haddif, G., & Yerulshamy, M. (this volume). Resources and constraints designed to support online e-assessment of problem solving. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 93–105). Cham, Switzerland: Springer.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77* (1), 81–112.

IPPR. (2013). *Excellence and equity: Tackling educational disadvantage in England's secondary schools*. http://www.ippr.org/publications/excellence-and-equity-tackling-educational-disadvantage-in-englands-secondary-schools. Accessed July 01, 2017.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill & Melinda Gates Foundation.

Laurillard, D. (2012). *Teaching as a design science*. London, England: Routledge.

Moeller, B., & Reitzes, T. (2011). *Integrating technology with student-centered learning*. Quincy, MA: Nellie Mae Education Foundation.

Organization for Economic Cooperation and Development. (OECD). (2010). *PISA 2009 volume II: Overcoming social background: Equity in learning opportunities and outcomes.* Paris, France. Retrieved from http://www.oecd.org/pisa/pisaproducts/48852584.pdf.

Pai, J. (this volume). Observations and conversations as assessment in secondary mathematics. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 25–44). Cham, Switzerland: Springer.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. National Research Council, Washington, DC: National Academy Press.

Platz, M., Krieger, M., Niehaus, E., & Winter, K. (this volume). Suggestion of an E-proof environment in mathematics education. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 107–120). Cham, Switzerland: Springer.

Rincon-Gallardo, S., & Elmore, R. (Winter 2012) Transforming teaching and learning through social movement in Mexican Public Middle Schools. *Harvard Educational Review, 82*(4), 471–490.

Sayac, N. (this volume). French primary teachers' assessment practices: Nature and complexity of assessment tasks. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 159–178). Cham, Switzerland: Springer.

Sia, C., & Lim, C. (this volume). Using cognitive diagnostic assessment (CDA) as an alternative mode of assessment for learning. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 123–137). Cham, Switzerland: Springer.

Soderstrom, N. C., & Bjork, R. A. (2013). Learning versus performance. In D. S. Dunn (Ed.), *Oxford bibliographies online: Psychology*. New York, NY: Oxford University Press.

Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance. *Perspectives on Psychological Science, 10*(2), 176–199.

Straumberger, W. (this volume). Using self-assessment for individual practice in math classes. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 45–60). Cham, Switzerland: Springer.

Swan, M., & Foster, C. (this volume). Formative assessment lessons. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 11–24). Cham, Switzerland: Springer.

van Eersel, G., Verkoeijen, P., Povilenaite, M., & Rikers, R. (2016). The testing effect and far transfer: The role of exposure to key information. *Frontiers in Psychology, 7,* 1977. https://doi.org/10.3389/fpsyg.2016.01977.

Wiliam, D. (2007). *Assessment, learning and technology: Prospects at the periphery of control.* Keynote speech at the 2007 Association for Learning Technology Conference. Nottingham http://www.alt.ac.uk/docs/altc2007_dylan_wiliam_keynote_transcript.pdf. Accessed July 01, 2017.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, MS: Solution Tree Press.

Wiliam, D. (2016). *Leadership for teacher learning*. West Palm Beach, FL: Learning Sciences International.

Wright, D., Clark, J., & Tiplady, L. (this volume). Designing for formative assessment: A toolkit for teachers. In D. R. Thompson, M. Burton, A. Cusi, & D. Wright (Eds.), *Classroom assessment in mathematics: Perspectives from around the globe*, (pp. 207–228). Cham, Switzerland: Springer.

Author Biographies

David Wright is Senior Research Associate: Research Centre for Learning and Teaching Newcastle University (United Kingdom) (now retired). David has fifteen years' experience teaching mathematics at secondary, further and higher education as an associate lecturer with the Open University. He was Subject Officer for Mathematics for the British Educational Communications and Technology Agency (Becta) for four years and ten years in initial teacher education and research at Newcastle University. He is the Scientific Director of the European Union research project: Formative Assessment in Science and Mathematics Education (FaSMEd).

Megan E. Burton is an Associate Professor and the elementary education program coordinator at Auburn University, Alabama (USA). She teaches and advises undergraduate and graduate students in elementary education and conducts research related to elementary mathematics education, with focus on elementary teacher change, inclusion, and rural education. As a former elementary teacher with experience in inclusion and English Language Learners, Burton is committed to classrooms that allow all students to encounter strong mathematics instruction in meaningful, differentiated ways.

Annalisa Cusi graduated in Mathematics at Modena and Reggio Emilia University in 2001, where she obtained a Ph.D. in Mathematics in 2009. She's been teaching mathematics and physics in upper secondary school since 2001. She worked as a research fellow at the University of Turin from 2014 to 2016 within the European Project FaSMEd. Her main research interests are innovation in the didactics of algebra; the analysis of teaching/learning processes, with a focus on the role played by the teacher; methods to promote early algebraic thinking in young students; teacher professional development; and formative assessment processes in mathematics.

Denisse R. Thompson is Professor Emeritus of Mathematics Education at the University of South Florida in the U.S., having retired in 2015 after 24.5 years on the faculty. Her research interests include curriculum development and evaluation, with over thirty years of involvement with the University of Chicago School Mathematics Project. She is also interested in mathematical literacy, the use of children's literature in the teaching of mathematics, and in issues related to assessment in mathematics education. She served as co-chair of Topic Study Group 40 on classroom assessment at ICME 13. In addition, she is a co-editor of the series *Research in Mathematics Education*, published by Information Age Publishing.

Author Index

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*,
ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5

Subject Index

D. R. Thompson et al. (eds.), *Classroom Assessment in Mathematics*, ICME-13 Monographs, https://doi.org/10.1007/978-3-319-73748-5