Mark deBoer
Dmitri Leontjev *Editors*

Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms

Approaches and Conceptualisations

Springer

Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms

Mark deBoer • Dmitri Leontjev
Editors

Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms

Approaches and Conceptualisations

Springer

Editors
Mark deBoer
English for Academic Purposes
Akita International University
Akita, Japan

Dmitri Leontjev
Department of Language and
Communication Studies
University of Jyväskylä
Jyväskylä, Finland

ISBN 978-3-030-54127-9 ISBN 978-3-030-54128-6 (eBook)
https://doi.org/10.1007/978-3-030-54128-6

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

To Shuku and Anna

Foreword

Effectively used, formative assessment stimulates, amplifies, accelerates, and importantly deepens learning. In fact, deep learning is dependent on robust systems of assessment, in particular the ongoing thoughtful use of formative assessment (Fullan et al. 2018). Not using assessment in a knowledgeable and thoughtful manner as a tool for supporting learning is not a defensible option. Hence, *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms: Approaches and Conceptualisations* is a welcome contribution to the discourse on classroom-based assessment in CLIL contexts. It will help frame conversations about assessment in CLIL whilst also offering practical ideas for using assessment to support the learning of both content and language. Importantly, it also guides readers to diverse national and international resources.

Much as assessment can be a powerful and positive driver of learning, it can also thwart learning and indeed cause harm to individuals and education systems. As Scott (2010) and Stobart (2008) argue, assessment, particularly in high-stakes contexts, can have serious unintended negative consequences often referred to as 'negative washback effects'. These can occur at the macro level affecting an entire system or at the micro level affecting the individual. On the macro level, over an extended period of time, high-stakes tests can lead to a reduction in the knowledge, skills and dispositional states that we value and teach, and this in turn can eventually lead to an impoverished learning environment. On the micro level, ineffective use of assessment can undermine students and teachers' beliefs in their own efficacy, agency and capacity to be successful.

Assessment constitutes a potent mix of possibilities and risks. Importantly, thoughtful evidence-based formative assessment holds great potential, and can act as a strong counterweight to the possible negative effects of high-stakes summative assessment. When teachers and students have discussed/negotiated intended learning outcomes; when the criteria being used to measure achievement have been jointly explored and agreed upon; when students know what success looks like; when students have substantial opportunities for well-structured self- and peer-assessment throughout the learning process; when there are several feedback/feedforward loops built into a meaningful learning progression, which allow students to

work at applying feedback/feedforward to improve their learning; when teachers seek to understand if and how feedback/feedforward has been taken on board; and, finally, when both teachers and students reflect on and discuss how teaching and learning practices need to change based on evidence, then formative assessment has the potential to be a particularly positive and powerful tool for deepening thinking and learning.

This is likely for several reasons. If we have a sense of where we want to go, we are more likely to get there. If we have clarity in what constitutes quality, we are more likely to be able to create it. Similarly, if we can measure our progress in learning, we are more likely to be able to manage it. If we have a grasp of what deep learning looks and feels like, we are more able to seek it out. In other words, the effective use of formative assessment imposes greater intellectual demands on all involved. Teachers and students cannot hide behind generalities imputing diverse meanings to vague intended learning outcomes, or operate with only a general notion of what quality looks like and how it will be measured, or provide each other with only general feedback without being challenged to explain further. The editors of this volume and several of the contributors rightly argue that intended content and language learning outcomes, criteria, exemplars, feedback/feedforward, as well as self- and peer-assessment all need to cohere as part of the integrated and dynamic whole that constitutes thoughtful assessment whose purpose is to promote learning.

As students and teachers experience several feedback loops, this creates additional opportunities to reflect on their work, to think more deeply about what they are doing and trying to achieve. This reduces the possibility of doing work quickly and superficially. It also means students are likely to become more able and active partners in the learning process and not simply accept judgements and unhelpful feedback from others. Formative assessment can help students gain greater control over the learning process. Feeling some level of independence, power and control over our lives is a fundamental psychological need, and if this is denied to students, they will seek ways of satisfying that need in a manner that may well impede their learning (Frey and Wilhite 2005).

This volume wisely points out the role of affect in the learning process. '[T]he aspects of cognition that are recruited most heavily in education, including learning, attention, memory, decision-making, motivation and social functioning, are both profoundly affected by emotion and in fact subsumed within the processes of emotion' (Immordino-Yang and Damasio 2007, p. 37). Assessment can be comfortably added to that list. According to a meta-analysis of 213 social-emotional learning programmes by Durlak et al. (2011), students who are taught social and emotional skills not only become more socio-emotionally competent but also see on average an 11% increase in their achievement on high-stakes tests. Sklad and colleagues (2012) also conducted a meta-analysis of 75 social and emotional learning programmes and found similar results. Formative assessment involving self-reflection and peer-reflection is an important vehicle for the development of self-regulation skills.

Formative assessment also has the added potential benefit of making visible achievements one step at a time. The effect of achievement on self-concept is primordial. It 'is stronger than the effect of self-concept on achievement' (Muijs and

Reynolds 2017, p. 163). This implies that well-planned learning progressions, where achievements are experienced and made visible one step at a time, help instil a sense of self-confidence in students.

Ultimately, the purpose of this book is to contribute to the improvement of teaching and learning practices through effective formative assessment. If there is no improvement in teaching and learning practices, there has been no *assessment for learning*. There is much to be gained from investing time collectively into getting formative assessment right. Wiliam (2018) reminds us that research evidence suggests that the greatest benefit to students' learning is likely to accrue from teachers using short-cycle formative assessment in the here and now. This must be grounded in evidence used to reflect on student learning and teaching in order to enhance both. Hattie and Zierer (2018) hold a similar view, but also stress that teacher passion and enthusiasm are essential driving forces. It is the belief in all students' capacity to learn and the desire to make that happen that are at the source of this passion and enthusiasm. They invite their readers to develop what may be called an assessment for learning culture where a series of mind frames help make visible what is to be learnt, why and what success looks like from the outset. Teachers and students are co-agents working to understand and agree on progress and the impact of teaching and learning practices. This can only take place through a dialogic process in an atmosphere of trust where it is safe to err. They also stress the importance of focussing on both content and language. *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms: Approaches and Conceptualisations* similarly argues for much of the above, but in CLIL contexts.

The above and the core messages contained in this book constitute high expectations. High expectations are a hallmark of successful CLIL programmes (Genesee and Hamayan 2017). If the above is going to be applied, it will take teacher collective efficacy – an exceptionally powerful influence on improving student achievement (Hattie 2015) – as well as stakeholder engagement, cooperation and support as people jointly sort out their thinking and try to bring greater clarity and precision to their practice (Mehisto and Genesee 2015). As a case in point, teachers in Helsinki have over the past 3 years worked collectively to flesh out what students are being expected to learn and be able to do both in reference to content and language in their primary level CLIL programme where up to 20% of the curriculum is to be delivered through an additional language. This has involved teacher initiative, school principal and vice-principal support, support from the local education authority, parent involvement, and guidance from external experts. This foundational work, which should facilitate the use of formative assessment, is a work in progress. Hopefully, it also represents an ongoing way of working.

Implicitly, this book calls for a change in work culture in CLIL contexts. May it contribute both to enhancing *assessment for learning* practices amongst CLIL teachers and students, and help students become assessment literate agents of their own learning with all the short- and long-term benefits evidenced in this book.

University College London Institute of Education, UK Peeter Mehisto

References

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school- based universal interventions. *Child Development*, *82*(1), 405–32.

Frey, L. M., & Wilhite, K., (2005). Our five basic needs: Application for understanding the function of behavior. *Intervention in School and Clinic, 40*(3), 156–160.

Fullan, M., Quinn, J., & McEachen, J. (2018). *Deep learning*: *Engage the world change the world.* London: Corwin.

Genesee, F., & Hamayan, E. (2017). *CLIL in context. Practical guidance for educator*s. Cambridge: Cambridge University Press.

Hattie, J. (2015). *What works best in education: The Politics of collaborative expertise.* London: Pearson.

Hattie, J., & Zierer, K. (2018). *Ten mindframes for visible learning*. Oxon: Routledge.

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education, 1*(1), 3–10.

Mehisto, P., & Genesee, F. (Eds.) (2015). *Building bilingual education systems: Forces, mechanisms and counterweights*. Cambridge: Cambridge University Press.

Muijs, D., & Reynolds, D. (2017). *Effective teaching: Evidence and practice*. London: Sage.

Scott, D. (2010). *Education, epistemology, and critical realism*. London: Routledge.

Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment?. *Psychology in the Schools*, *49*(9), 892–909.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.

Wiliam, D. (2018). *Creating the schools our children need*. West Palm Beach: Learning Sciences International.

Preface

In the present, rather short, section, we will share the inspiration that led us to start working on this volume. We will also introduce the contributors and their respective chapters.

The idea for this volume emerged from our discussions around both the topic of assessment and CLIL amongst the two of us and our separate discussions with our colleagues. Within these discussions, it became clear to us that to quote ourselves in the open call for contributions to the volume, there is little "emphasis on assessment in CLIL research, so much so that there is no clear understanding or systematization of the process of assessment in CLIL (Coyle, Hood, & Marsh, 2010)." Our early work on the volume proposal was guided by this understanding. The more we studied the field as a part of the preliminary work on the volume, the more it became clear that this was indeed the case. This is not to say that no work at all has been conducted in this regard. For example, Lin (2016) has an excellent chapter discussing a number of practical ways for how assessment in CLIL can be planned and conducted with a view of balancing content and language in assessment and gives examples of assessment tasks accompanied with a discussion of what is assessed in these tasks and how. We highly recommend not just this chapter but also the whole book to any CLIL educator. Quartapelle (2012) includes a more detailed theoretical argument and examination of the balance between content and language in CLIL assessment, taking the viewpoint of the primacy of assessing content in CLIL, and, too, offers practical applications of assessment in CLIL from primary to tertiary levels.

We do not consider the present volume as a competitor to the aforementioned titles, rather as a further step that both encompasses a wider number of contexts geographically and limits the discussion to classroom assessment promoting learning. The volume, therefore, simultaneously hosts a more varied number of standpoints to assessment in CLIL than in Lin (2016) and in Quartapelle (2012) and allows for more detailed discussions of assessment in CLIL by focussing these discussions on specifically assessment *for* learning and learning-oriented assessment.

Partially the decision to focus on assessment promoting learning stems from our own backgrounds, both of our research strongly based on sociocultural theory, as

will transpire in our corresponding chapters in the volume. More importantly, however, we would argue that the central premise of CLIL, that is, the understanding of knowledge as co-constructed with others, teaching and learning thus interacting, make assessment an essential part of this process. Last but not least, the discourse in the field of educational assessment, inspired by both constructivist and critical turns in education, shifted towards the major purpose of any assessment in the classroom being to promote learning. We will return to these points in the following Chap. 1, but would like to add that while the contributors to the present volume share this general understanding, there is a great variability in the approaches to assessment across the chapters in the volume.

As we will elaborate in Chap. 1, we arranged the chapters based on the way that content and language are integrated in teaching, learning, and assessment as discussed in the chapters.

In Chap. 2, Dr. Stuart D. Shaw proposes an academic proficiency scale that is based on the CEFR (Common European Framework of Reference) descriptors. He then, at the end of the chapter, discusses assessment in a CLIL history lesson where the descriptors are used.

In Chap. 3, Dr. Alexander Nanni and Dr. Chris Carl Hale examine how sharing the assessment responsibilities with learners promotes their awareness of learning goals and self-reliance in directing their own learning.

In Chap. 4, Dr. Claudia Kunschak proposes a framework for assessment based on the principles of the CEFR, task-based language teaching, and learning-oriented assessment.

In Chap. 5, Ana Xavier discusses a framework, building on the principles of learning-oriented assessment she developed as a part of her Master's studies. She proposes guidelines and practical ideas emerging from the framework as to how the process of assessment promoting learning can be organised in practice in primary-level CLIL classrooms.

In Chap. 6, Dr. Taina Wewer explores teachers' assessment practices in Finnish primary school CLIL lessons. Based on the outcomes of this investigation, Dr. Wewer proposes ways in which these practices can be developed. We positioned her chapter at this point in the volume as it provides an excellent overview of practices and challenges of assessing language in CLIL contexts where the emphasis is on content.

In Chap. 7, Associate Professor Hidetoshi Saito, based on his study of EFL teachers in Japan, proposes a framework for classifying different types of assessment in classrooms where content and language are integrated. He builds the argument for developing CLIL practices in Japan, moving beyond EFL to actually integrating content and language in such classes.

In Chap. 8, Dr. Rachael Basse and Dr. Irene Pascual Peña study how incorporating a number of assessment *for* learning techniques promoted the shift from teacher-centred interaction to co-construction of knowledge in which both learners and the teacher contributed. In the end, this resulted in the learners adopting an active role in the classroom.

In Chap. 9, Dr. Dmitri Leontjev, Dr. Teppo Jakonen and Dr. Kristiina Skinnari study CLIL teachers' perspectives on assessment and, inspired by those, analyse an interaction between a teacher and a learner. Based on the outcomes of their inquiry, they build an argument for one purpose of such assessment being understanding learners and promoting learners' understanding.

In Chap. 10, Mark deBoer studies the mediation that occurs between learners in an online forum. He explores how content and language mediate one another in learners' joint construction of a poster. He then meditates on how this information can inform classroom-based assessment cycles.

The contributors cover a wide range of geographical contexts, approaches to assessment, and ways that content and language are integrated in CLIL, and we appreciate the depth to which everyone has contributed to the process that has brought it to fruition.

Yuwa, Akita, Japan — Mark deBoer
Jyväskylä, Finland — Dmitri Leontjev

Acknowledgements

First and foremost, we would like to express our sincerest gratitude to the contributors of this volume. The volume houses different perspectives on CLIL and assessment promoting learning. These different angles were what made working on the volume such an interesting journey in which both of us definitely developed as scholars, writers and individuals. We thank the contributors for their dedication throughout the journey of making this volume happen.

We also owe our gratitude to our critical friend in this volume, Dr. Peeter Mehisto, who dedicated his time to read and comment on the volume drafts at different points of its creation. We were also fortunate to have the assistance of many scholars who shared their expertise commenting on the early versions of the chapters. What comes next is a list which we present in alphabetic order by the family name: Renata Agolli, Dr. Attila Egri-Nagy, Professor Ari Huhta, Professor Barry Kavanagh, Dr. Bärbel Kühn, Assoc. Professor Ana Llinares, Dr. Waldemar Martyniuk, Professor Tarja Nikula-Jäntti, Dr. María Luisa Pérez-Cañado and Professor Srikant Sarangi — we are grateful to each and every one of you! We are also grateful to Professor Chris Davison for kindly giving us her permission to reuse the figure for the assessment cycle in our volume.

Developing one's understanding is essential to any process. Our own understanding, too, developed notably in the process of working on the volume. The discussions that we had during the symposium that we organised in Tokyo in October 2018 proved to be especially useful in directing this development. We are grateful to the participants in the symposium. Everyone who participated must know that it was also due to your input that the final chapter in the volume has shaped in the present way. We thank the JALT CEFR and Language Portfolio SIG, particularly Professor Maria Gabriela Schmidt and Professor Naoyuki Naganuma, for helping us organise the symposium and Tokai University for hosting the symposium.

Much collaboration happens online nowadays. However, the work on the volume would not be as fruitful (and the above-mentioned symposium would not have happened) should Dmitri have not been given a chance to spend a month at Akita International University, Japan, to work together with Mark on the volume. We are thankful to Research Collegium for Language in Changing Society (RECLAS;

https://www.jyu.fi/reclas/en) for funding this mobility period for Dmitri. We are also grateful for the hospitality that Akita International University provided during this month. Former Vice President of AIU Dr. Peter McCagg and EAP Department Director Professor Patrick Dougherty provided the means for us to be able to collaborate and also to host a number of faculty development sessions that certainly allowed us to develop our thinking further.

Last but not least, we would like to thank the reviewers, whose comments and suggestions helped us to build a stronger and a more coherent argument in this volume, and staff at Springer, especially Annemarie Keur, Claudia Acuna, Marianna Pascale and Malini Arumugam, for making the process as smooth as possible.

We hope this volume will offer valuable insights to CLIL researchers and educators alike and will serve as an inspiration for further developments in CLIL assessment.

Akita, Japan — Mark deBoer
Jyväskylä, Finland — Dmitri Leontjev

Contents

1 Conceptualising Assessment and Learning in the CLIL Context. An Introduction 1
Dmitri Leontjev and Mark deBoer

2 Achieving in Content Through Language: Towards a CEFR Descriptor Scale for Academic Language Proficiency 29
Stuart D. Shaw

3 Academic Culture as Content: Self-Assessment in the CLIL Classroom in the International Liberal Arts University 57
Alexander Nanni and Chris Carl Hale

4 CEFR, CLIL, LOA, and TBLT – Synergising Goals, Methods and Assessment to Optimise Active Student Learning 85
Claudia Kunschak

5 Assessment for Learning in Bilingual Education/CLIL: A Learning-Oriented Approach to Assessing English Language Skills and Curriculum Content in Portuguese Primary Schools 109
Ana Xavier

6 Assessment of Young Learners' English Proficiency in Bilingual Content Instruction (CLIL) in Finland: Practices, Challenges, and Points for Development 137
Taina Wewer

7 Gearing Teaching and Assessment Towards CLIL: Theorizing Assessment *for* Learning in the Junior High School Soft CLIL Classrooms in Japan 167
Hidetoshi Saito

8 Does A*f*L Promote Discussion in CLIL Classrooms? Exploring A*f*L Techniques and Their Effect on Classroom Communication 189
Rachel Basse and Irene Pascual Peña

9 Assessing (for) Understanding in the CLIL Classroom 205
Dmitri Leontjev, Teppo Jakonen, and Kristiina Skinnari

10 Teacher-Based Assessment of Learner-led Interactions in CLIL: The Power of Cognitive Discourse Functions 229
Mark deBoer

11 Conclusion: Dialectics in CLIL Classrooms 253
Mark deBoer and Dmitri Leontjev

About the Editors and Contributors[1]

Rachel Basse Universidad Autónoma de Madrid, Madrid, Spain

Dr. Rachel Basse is an Assessment Developer at a North American language testing company. She earned her Ph.D. in Applied Linguistics from the Universidad Autónoma de Madrid and has taught English at the tertiary level in the USA and Spain. She has also taught CLIL classes in primary bilingual schools in Madrid. Her primary research interests lie in exploring the connection between assessment and motivation in CLIL classroom contexts. rachel.basse@gmail.com

Chris Carl Hale Akita International University, Akita, Japan

Dr. Chris Carl Hale is Associate Professor in the Graduate School of Global Communication and Language at Akita International University, Japan. He has coordinated and conducted numerous workshops on teacher education in Japan and internationally and teaches graduate courses in language education and teaching methodology. His research has appeared *in Language Testing in Asia, TESOL International and Columbia University Teachers College Journal of TESOL and Applied Linguistics*. His research interests include writing self-assessment, classroom discourse and conversation analysis. He is also an active DJ and techno music producer. chale@aiu.ac.jp

Teppo Jakonen Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland

Dr. Jakonen is a Postdoctoral Researcher in the Department of Language and Communication Studies at the University of Jyväskylä. His methodological expertise is in multimodality and microanalysis of social interaction, particularly in language learning contexts. His research interests lie in exploring issues related to materiality, knowledge and mobility in interaction in learning environments. teppo.jakonen@jyu.fi

[1] The contributors in the volume are the following (in alphabetical order by the family name).

Claudia Kunschak College of International Relations, Ritsumeikan University, Kyoto, Japan

Dr. Claudia Kunschak is an Associate Professor at the College of International Relations at Ritsumeikan University in Kyoto, Japan. She received her Ph.D. in Education from the University of Arizona and her M.A. in Interpreting from the University of Vienna. She has taught at universities in Austria, China, Scotland, Spain, Ukraine and the USA. As a practitioner, she emphasises language awareness, learner autonomy and content-based instruction. Her research interests include curriculum development, assessment, and English as Lingua Franca. alexander.nan@ckr12106ritsumei.ac.jp

Alexander Nanni Humanities and Language Division, Mahidol University International College, Salaya, Thailand

Dr. Alexander Nanni is the Associate Dean for International Affairs at Mahidol University International College (MUIC), which is located in Salaya, Thailand. He began teaching English in Thailand in 2005. From 2012 to 2018, he was Director of the Preparation Center for Languages and Mathematics at MUIC. Alexander holds an M.Ed. in Teaching English as a Second Language from Rhode Island College and an Ed.D. in Curriculum, Teaching, Learning, and Leadership from Northeastern University in Boston. alexander.nan@mahidol.ac.th

Irene Pascual Peña Universidad Autónoma de Madrid, Madrid, Spain

Dr. Irene Pascual Peña got both her university degree and her Ph.D. from Universidad Autónoma de Madrid. Her Ph.D. was framed in the Applied Linguistics Programme. She has taught at tertiary level in different universities in Madrid, but is currently developing her teaching career at the secondary level. Her research interests include CLIL, classroom discourse and Assessment for Learning. irenepascual84@hotmail.com

Hidetoshi Saito College of Education, Ibaraki University, Mito, Ibaraki, Japan

Dr. Hidetoshi Saito, (Ph.D., The Ohio State University) teaches pre- and in-service EFL teachers and graduate students at Ibaraki University, Japan. Some of his research studies appeared in international journals including *Language Assessment Quarterly, Language Testing and Language Teaching Research*. His research interests include CLIL, discussion instruction and assessment for learning. He is currently a Vice President of Japan Language Testing Association. hidetoshi.saito.cldwtr@vc.ibaraki.ac.jp

Stuart D. Stuart Shaw Cambridge Assessment International Education, Cambridge Assessment, Cambridge, UK

Dr. Stuart D. Shaw has worked for Cambridge Assessment since January 2001, where he is particularly interested in demonstrating how Cambridge Assessment

seeks to meet the demands of validity in its assessments. He has a wide range of publications on English second language assessment and in educational and psychological research journals, including the forthcoming book on fairness in assessment titled *Is Assessment Fair* co-written with Isabel Nisbet. Stuart is a Fellow of the Association for Educational Assessment in Europe (AEA–Europe) and Chair of the Scientific Programme Committee (AEA–Europe). He is Chair of the Publications Committee (International Association for Educational Assessment) and is also a Fellow of the Chartered Institute of Educational Assessors (CIEA). stuart.shaw@cambridgeinternational.org

Kristiina Skinnari Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland

Dr. Skinnari works as a University Teacher in the Department of Language and Communication Studies at the University of Jyväskylä. She has worked as an English, special education and CLIL teacher in elementary school. Her recent research interests concern teacher agency, policies and subject-specific languages in CLIL education. kristiina.skinnari@jyu.fi

Taina Wewer Finnish National Agency for Education: European School Luxembourg I, Kirchberg, Luxembourg

Dr. Wewer is currently employed by the Finnish National Agency for Education in multilingual settings as a Class Teacher, and she also has a long career as a CLIL Class Teacher, Foreign Language Teacher and Teacher Educator. Her research interests have revolved around assessments, particularly target language assessment in bilingual settings. The two Fulbright scholarships in the USA granted to her have greatly contributed to her expertise in this specific research area and didactics of bilingual education. taina.wewer@gmail.com

Ana Xavier Direção-Geral da Educação/Ministério da Educação, Lisbon, Portugal

Ana Xavier is a Teacher working at the Directorate-General for Education, Ministry of Education, where she has been involved in the implementation of bilingual education in Portuguese primary schools with the British Council Portugal since 2011. She participated in national and international teacher training in CLIL and bilingual education and has recently completed a Master's degree in English Language Teaching at FCSH/UNL – NOVA Lisbon, focusing on CLIL assessment at early primary level. We, the two editors, but also the (co-)authors of the chapters in this volume, would like to introduce us the last. ana.cristina.xavier@gmail.com

Mark deBoer English for Academic Purposes, Akita International University, Akita, Japan

Mark deBoer is a Ph.D. candidate at the University of Birmingham, UK, and a Lecturer in the EAP Department at Akita International University. His areas of expertise include syllabus development, CLIL, and flipped classroom teaching and

learning. His research interests lie in exploring how knowledge is co-constructed in a learner-centred classroom and learner–learner mediation. markdb@aiu.ac.jp

Dmitri Leontjev Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland

Dr. Leontjev is a Postdoctoral Researcher in the Department of Language and Communication Studies at the University of Jyväskylä. His areas of expertise include dynamic assessment and feedback in the language classroom. His recent research interest lies in exploring how knowledge is co-constructed in classroom interaction. dmitri.leontjev@jyu.fi

Abbreviations

AECLIL	Assessment and Evaluation in Content and Language Integrated Learning Project
A*f*L	Assessment *for* Learning
A*o*L	Assessment *of* Learning
APA	American Psychological Association
BICS	Basic Interpersonal Communication Skills
BSP	Bilingual Schools Project, Portugal
CALLA	Cognitive Academic Language Learning Approach
CALP	Cognitive Academic Linguistic Proficiency
CDF	Cognitive Discourse Function
CEFR	Common European Framework of Reference for Languages
CELA	Cambridge English Language Assessment, currently Cambridge Assessment English
CLIL	Content and Language Integrated Learning
CLILA	CLIL Learner Assessment Project
CLT	Communicative Language Teaching
COPS	Capitalisation, Order and Organisation, Punctuation and Spelling
EAL	English as Additional/Alternative Language
EFL	English as a Foreign Language
ELP	European Language Portfolio
EMI	English Medium Instruction
ESP	English for Specific Purposes
FNAE	Finnish National Agency for Education
FSCH	Faculty of Social and Human Sciences, University in Lisbon
HOTS	Higher Order Thinking Skills
ICT	Information and Communications Technology
IGCSE	Cambridge International General Certificate of Secondary Education
IO	Improvable Object
IRF/E	Initiation-Response-Feedback/Evaluation
ISCED	International Standard Classification of Education
ITP	Institutional Testing Program

JALT	Japanese Association of Language Teachers
L1	Mother Tongue
L2	Second or Foreign Language
LMS	Learning Management System
LOA	Learning-Oriented Assessment
LOTS	Lower Order Thinking Skills
LSP	Language for Specific Purposes
MEC	Ministerio de Educación y Ciencia (Ministry of Education [Spain])
METI	Japanese Ministry of Economy, Trade, and Industry
MEXT	Ministry of Education, Culture, Sports, Science, and Technology
MT	Mother Tongue
NALDIC	National Association for Language Development in the Curriculum
NCC	National Core Curriculum
SCT	Sociocultural Theory
SMART	Specific, Measurable, Achievable, Realistic, and Time-Bound (learning outcomes)
TBA	Teacher-Based Assessment
TBLT	Task-Based Language Teaching
TIB	This is Because
TL	Target Language
TOEFL	Test of English as a Foreign Language
TSST	Telephone-Based Standardised Speaking Test
WALT	What We are Going to Learn Today
WIDA	World-Class Instructional Design and Assessment
WILF	What I'm Looking For
ZPD	Zone of Proximal Development

List of Figures

Fig. 1.1 Classroom-based assessment cycle (Davison 2008) 7
Fig. 1.2 The *hard* and *soft* CLIL continuum. (Adapted from Ball et al. 2015) 13
Fig. 1.3 Four different orientations to integration in CLIL (Leung and Morton 2016, p. 237) 15
Fig. 1.4 Chapters placed on Leung and Morton's (2016) integration matrix 16

Fig. 3.1 Essay topic: bioethics 64
Fig. 3.2 Business ethics discussion materials 68

Fig. 4.1 Step-by-step development of curricular model 93
Fig. 4.2 Interlocking four-pillar model 98

Fig. 5.1 A rationale for supportive and pedagogical assessment 119
Fig. 5.2 A continuum of assessment methods/techniques/tools; Abridged and adapted from Xavier (2016) 121
Fig. 5.3 Planning, using and recording an assessment sequence. (Xavier 2016) 122
Fig. 5.4 Peer assessment card for individual presentation; adapted from Ioannou-Georgiou and Pavlou (2003, p. 97) and Xavier (2016) 128

Fig. 6.1 Means of gathering information of TL proficiency in CLIL contexts. (Adapted from Wewer 2014a) 148

Fig. 10.1 The poster file Pet bottles poster 7.docx (see Post 3–1) 241
Fig. 10.2 The poster file Pet bottles poster 8.docx (see Post 3–4) 242
Fig. 10.3 The poster file Pet bottles poster 10.docx (see Post 3–5) 243
Fig. 10.4 The poster file Pet bottles poster 14.docx (see Post 3–6) 243

Fig. 11.1 Dialectical relationships between teaching, learning, and assessment 255

List of Tables

Table 2.1 Paper 1 skills assessed using the levels-based mark scheme 35
Table 2.2 Analysis of history language (Shaw and Imam 2013) 38
Table 2.3 Tentative academic language proficiency scale (CEFR Level C2) 42
Table 2.4 Tentative academic language proficiency scale (CEFR Level B2) 43
Table 2.5 Comparison of IGCSE history grade descriptors (*coherent, logical arguments*) with CEFR level descriptors 45
Table 2.6 History lesson plan using academic language proficiency descriptors (based on Ellis 2003, p. 217) 49

Table 3.1 Student self ratings 65
Table 3.2 Student self-assessment scores on seven-point likert scale 69
Table 3.3 Teacher assessment scores on seven-point likert scale 70
Table 3.4 Differences between teacher and student assessments 70

Table 6.1 Frequency of teacher-given feedback on pupils' TL proficiency and its development 152
Table 6.2 Teachers' descriptions of feedback practices serving parents 153

Table 7.1 The four-factor CLIL A*f*L framework 170
Table 7.2 Results of teacher survey on students' evaluation of the main characters' behaviours 176
Table 7.3 A sample tick slip peer assessment 177
Table 7.4 Groups' responses on tabletop whiteboards from Aki's focal lesson 178
Table 7.5 Self and peer assessment form used in Sota's class 179
Table 7.6 Descriptors of content from Sota's reflection sheet 179
Table 7.7 Summary of the sample A*f*L observed in the data 182

Table 8.1 Data set 194

Table 9.1 Teachers in the interview excerpts 209

Chapter 1
Conceptualising Assessment and Learning in the CLIL Context. An Introduction

Dmitri Leontjev and Mark deBoer

Multiple CLIL writers (e.g. Ball et al. 2015) have emphasised the significant role of assessment in promoting learning in CLIL classrooms, where there is a dual focus on learning content and language (Genesee and Hamayan 2016). Because of this dual focus, the assessment process in CLIL becomes more complex. Assessment in CLIL should provide insights into learner content and linguistic knowledge as well as strategies used to learn both content and language in order to identify student progress and needs. This should inform both teachers and students about how to enhance learning. Despite excellent overviews, guidelines, and practical activities in CLIL assessment (Lin 2016; Mehisto and Ting 2017; Quartapelle 2012), teachers express concerns about adopting new assessment practices, principles, and techniques and are underusing the potential to support learning (deLuca and Bellara 2013; Hill 2017a; Tsagari and Vogt 2017).

This chapter introduces (a) classroom-based assessment promoting learning, focusing on assessment *for* learning and learning-oriented assessment, (b) CLIL, and (c) teaching, learning, and assessment in CLIL. We will then synthesise these to inform our conceptualisation of assessment promoting teaching and learning of content and language in CLIL classrooms. A further role of this chapter is to introduce essential terms, notions, and conceptualisations used in the volume.

We will next discuss classroom-based assessment promoting learning in general, without discussing CLIL. Our intention is to set the boundaries of how we conceptualise assessment promoting learning based both on a solid theoretical foundation. We will also detail how the chapters align with this foundation.

D. Leontjev (✉)
Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland
e-mail: dmitri.leontjev@jyu.fi

M. deBoer
English for Academic Purposes, Akita International University, Akita, Japan
e-mail: markdb@aiu.ac.jp

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_1

1.1 Classroom-Based Assessment Promoting Learning

The field of educational assessment has seen a proliferation of terms such as formative assessment, assessment *for* learning, alternative assessment, classroom-based assessment, teacher-based assessment, learning-oriented assessment, and dynamic assessment. These terms all share one characteristic—they are used to refer to "context-based, classroom-embedded assessment practice, explicitly or implicitly defined in opposition to traditional externally set and assessed large scale formal examinations used primarily for selection and/or accountability purposes." (Davison and Leung 2009, p. 395). While these conceptualisations are related, they are not the same and sometimes rest on different theoretical bases.

The issue is more complicated as, at times, various scholars define the same terms differently. To give an example, Davies et al. (1999, p. 11) defined assessment as "often used interchangeably with testing." However, Lynch (2001) insisted that assessment is a more general term and concept than testing is. We side with Lynch's (2001) understanding of assessment.

A classical definition of assessment adopted in education is *collecting information using clearly defined procedures based on a theory, methodology, and practice* (Bachman 1990). However, in our view, this definition needs to be modified to accommodate unplanned assessments (Hill 2017b). We define assessment as encompassing *various instruments and approaches aiming at yielding insights into learner abilities*. It captures the essence of why we assess—to *obtain information about learner abilities.* However, this information itself bears little meaning if it is not *acted upon*. We argue that the purpose of assessment defines what information is obtained, how it is obtained, how it is *interpreted*, and more importantly, how it is *used*.

We further argue that the context where assessment happens informs why we assess. Ours, to repeat, is classroom-based assessment. Hill and McNamara (2012, p. 396), based on McNamara (2001), defined classroom-based assessment as

> any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes.

While all of these purposes are relevant in the classroom, the main purpose of classroom-based assessment, we argue, is to promote learning, which should take precedence over reporting and management. One model of assessment in the classroom which is largely congruent with our thinking is that by Davison (2008) discussed in detail by Davison and Leung (2009).

This model conceptualises classroom-based assessment as a cycle of linked stages, in which assessment is used to change teaching, learning, and subsequent assessment. Still, some modifications to the model should be made to adopt it for our purposes. To start with, the authors discussed teacher-based assessment, limiting the assessing agent to the teacher, whereas classroom-based assessment also involves peer- and self-assessment. Second, assessment cycles can include unplanned assessments. Third, factors shaping the classroom-based assessment

cycle other than syllabi and curricula should be recognised. These include learner and teacher identities shaped by various communities and institutions as well as beliefs, ideologies, and values, all contributing to learners' and teachers' individual and shared histories. Finally, feedback and advice in the model should be limited to those promoting learning. This is not to claim that feedback to parents on learner achievement is not needed. Indeed, feedback to parents can, too, influence learning, albeit indirectly (Mehisto et al. 2008, p. 30). We, however, will not elaborate much on such feedback (but see Chap. 6, this volume). Regardless, the model is useful for visualising how various parts of the classroom-based assessment cycle are emphasised in the contributions to the volume.

Most of the contributions cover several parts of the cycle. Chapter 6, for example, while discussing particular assessments practices (or lack thereof) and giving feedback in primary CLIL classrooms, discusses these with reference to more macro levels of the Finnish context, such as teacher beliefs and the role of curriculum. Chapters 2 and 7 emphasise planning assessments promoting learning. Chapter 2 specifically focuses on establishing assessment criteria and standards through the use of a scale which integrates language and content goals. Chapter 7 proposes a framework for planning assessment promoting learning of content and language. Chapter 10 focuses on making inferences based on learner-learner interaction, emphasising the lower part of Davison's (2008) model. Chapter 3, too, focuses on the lower part of the cycle, but the main agents making the judgements about their performance are learners, as they compare and discuss their evaluation of their writing with that of the teacher'. Chapters 8 and 9 concentrate on how teaching, learning, and assessment occur in interaction with learners. These chapters can be positioned as discussing collecting information, making judgments about learner performance, and giving learner feedback, all at the micro level of the activity of classroom teacher-learner interaction. Therefore, they concentrate rather on the centre of the model. Finally, Chaps. 4 and 5 concern the whole cycle with reference to the framework of learning-oriented assessment (LOA), which we discuss later in this section.

Even though all chapters in this volume go beyond collecting information about learner abilities/performance, indicated on the right of Davison's (2008) cycle, we urge the reader not to misinterpret this as the contributors to the volume considering the whole cycle as assessment.

Feedback, for example, is not the same as the action of assessing learners. Rather it is about communicating to learners (and other agents) the interpretation of the information received during the assessment process. Yet, it is feedback to learners through which assessment realises its purpose—promoting learning. Teaching, learning, and assessment are, too, not the same. As Mehisto and Ting (2017, p. 263) stated, one cannot assume that learning and teaching progression are one and the same. Assessment helps teachers to infer what it is of what has been taught that their learners learned and how they can adjust their teaching to help learners learn what they have not yet. However, in the classroom-based assessment cycle, teaching, learning, and assessment are related. The exact way this connection is made varies depending on the information elicited by assessment approaches and the teaching/learning goals (see Sect. 1.7).

Considering the above, even when the emphasis is not on assessment proper in certain chapters, we deemed it necessary to include these to the present volume as it is important to consider the whole cycle, not just planning assessments, collecting information, and making inferences based on these.

In order to further conceptualise assessment in the volume, it is useful to differentiate between the function and purpose of classroom-based assessment. Following Wiliam (2017), we see 'purpose' at a higher level of abstraction (e.g. promoting learning) whereas 'function' referring to how this purpose is achieved in certain assessment activities (e.g. inferring which aspects need to be explained more).

In terms of their function, assessment activities are broadly classified into summative and formative (Scriven 1967). We met different understandings of concepts of 'formative' and 'summative'. We, based on a solid theoretical/conceptual basis, define here and elsewhere, how these terms, and others, are used *in the present volume*.

The function of summative assessment is to provide evidence for learner achievement at the end of a learning period. Such assessments are meant to evaluate learners' performance against a standard or a norm. The outcome is often giving a score or a grade. An assessment initially designed with the summative function in mind can also serve a formative function.

Assessment has formative function "to the extent that evidence about students' 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 that would have taken in the absence of the evidence that was elicited" (Black and Wiliam 2009, p. 9). Wiliam and Leahy (2015) list the following features of the formative function of assessment:

- teachers, learners themselves, or their peers can be agents in assessment;
- the focus is on decisions; the choice of assessment should be guided by the inferences to be made;
- the formative inferences improve the likelihood of learning to take place
- should the inferences from assessment deem it necessary, adjustments to teaching happen.

Following Wiliam (2017) and Wiliam and Leahy (2015), we do not consider formative or summative assessment a 'thing' but a *function*, so one and the same assessment tool can serve either of these functions or both. In other words, a test, for example, can be used both summatively—learner scores showing the group's and individual learners' achievement—and formatively—scores informing the teacher where additional support is needed (Wiliam 2017). An assessment activity has a formative function when the information obtained in this assessment *is used to bring together teaching and learning*. Certainly, though, there are assessments that serve either formative or summative function better. Furthermore, an assessment activity such as a test can serve a formative function to a teacher, but not to learners (see Black et al. 2003).

The above is just one interpretation of formative assessment. Numerous interpretations of formative assessment were the reason for the Assessment Reform Group

to adopt a different term—'assessment *for* learning' (Broadfoot et al. 2002). Wiliam (2017) still noted that in collaboration with Black, they consistently used the term 'formative', but also, as emerges elsewhere in their publications, used 'formative assessment' and 'assessment *for* learning' interchangeably. Wiliam (2017, p. 400), following such voices as Bennet (2011), and as also noted by Baird et al. (2017), stated that "a shift in terminology just moves the definitional burden."

Building on Black et al. (2004, p. 10), we define assessment *for* learning as any assessment for which the first priority in its design and practice is to promote students' learning. To elaborate, the purpose of assessment as discussed in the present volume, reiterating Black and Wiliam (1998) is to improve learning by yielding insights into learners' abilities, with the intention to promote learners' and assessors'/teachers' understanding of where learners need to go next and how they can be directed in their development. In many ways, this definition bears many a similarity with that of formative assessment that we gave previously. The difference, however, is in that assessment *for* learning emphasises the purpose of the process of classroom-based assessment, with specific assessment activities feeding into this process. Assessment *for* learning then becomes an iterative process where the previous assessment informs teaching and learning, which in turn, inform the following assessment, e.g. during a course, a semester, or a school year.

The question that arises with regard to assessment *for* learning is whether summative assessment can be *for* learning. Some voices, such as Broadfoot et al. (2002) argued for keeping formative and summative assessments apart, equating assessment *for* learning with formative assessment activities and not linking it in any way with summative. Black et al. (2003) argued for the impossibility of this, as the reality of the classroom is that summative assessment is a part of it, arguing instead for using assessments whose function is initially summative for formative purposes. That is, summative and formative functions of assessments co-exist in the classroom, summative and formative functions of assessment activities not seen as a dichotomy but rather as a continuum. Considering that 'summative' is a function, strictly speaking, it cannot be for learning, but in chapters on LOA in the volume, a somewhat different orientation is to the summative-formative continuum is used, as in LOA, both summative and formative assessments are there to promote learning.

Rather than continuing the discussion with reference to the formative/summative divide, we draw on Davison and Leung (2009) argument and propose to think about classroom-based assessment as either assessment *of* learning (A*o*L) or assessment *for* learning (A*f*L) *culture*. In *assessment of learning culture* summative and formative assessments are seen as having both different *form* and *function*. The roles of the teacher and the assessor are, too, separate. In *assessment for learning culture*, regardless of their form (e.g. a test) and function (e.g. planned for grading), classroom-based assessment activities should give learners feedback that guides learning. Considering the prior argument that summative and formative are to be seen as functions, we, building on the understanding classroom-based assessment as a culture, take the stance that assessment activities in the classroom should rather have either only the formative or both formative *and* summative functions.

To summarise, our conceptualisation of assessment *for* learning in the volume is that of *assessment for learning culture*. We further argue that in A*f*L culture, the emphasis is not on the *product of learning*, e.g. for the purpose of classifying learners based on how well they performed in assessment against a criterion or against one another; it is on the *process of learning* with the learner placed at the heart of the assessment process. This underscores ongoing monitoring and guidance of learning (Brookhart and Nitko 2008, p. 93) by both teachers and learners. We argue that it is the understanding that classroom-based assessment, regardless of its form, as contributing to the process of learning and teaching rather than overemphasising the completion, that makes it truly *for* learning.

Considering assessment *for* learning as both a process and a culture makes it more congruent with the classroom-based assessment cycle presented at the outset of the chapter, as it implies that any assessment in the classroom should be intrinsically linked to teaching and learning. One way this can be done is through feedback. Black and Wiliam (2010) argued that grades alone emphasise competition rather than learner development. For learning to occur, learners need feedback not just indicating what they are and are not able to do (where they are now), but also what needs to be done, and how (Alderson et al. 2015; Hattie and Timperley 2007; Wiliam and Leahy 2015; see Fig. 1.1). Hattie and Timperley (2007) discussed this with reference to feedback levels, arguing that whereas task feedback, telling learners how they are doing in relation to the task while is effective overall, does not help learners to transfer their knowledge to other tasks/contexts. Strategy level feedback, as termed by Alderson et al. (2015), who merged process and self-regulation feedback in Hattie and Timperley's (2007) model, is the most effective. Strategy feedback helps learners understand how they can advance in the direction to the goal. However, our stance is that effective feedback should simultaneously work on different levels, i.e. on task and strategy levels, but also should encourage learners (Hattie and Timperley 2007). Indeed, as Huhta et al. (in preparation) argued, used together with other feedback functions, it enhances opportunities for success that feedback to learners creates.

One issue with assessment *for* learning, as Christodoulou (2016) argued, is that it can be notoriously difficult to make formative inferences from some assessment activities initially designed to have a summative function (see Wiliam and Leahy 2015). Another challenge is in making the process of assessing learners coherent such that it leads to a desired outcome. Christodoulou (2016) argued that when the product, e.g. an essay, is to be graded, assessment *for* learning should not only be focused on the process of planning, drafting, and revising of the essay. It can also elicit aspects amounting to writing this essay, e.g. writing topic sentences, presenting different points of view, or coherence. It is essential that a connection among these activities is made clear to learners. Only then can learners gradually develop towards the desired outcome. The reality of the classroom, however, is that the focus of assessment often stays on learner performance on disconnected assessment activities.

It is due to the complexity of assessment *for* learning and the appearance of further conceptualisations such as assessment *as* learning, involving learners in

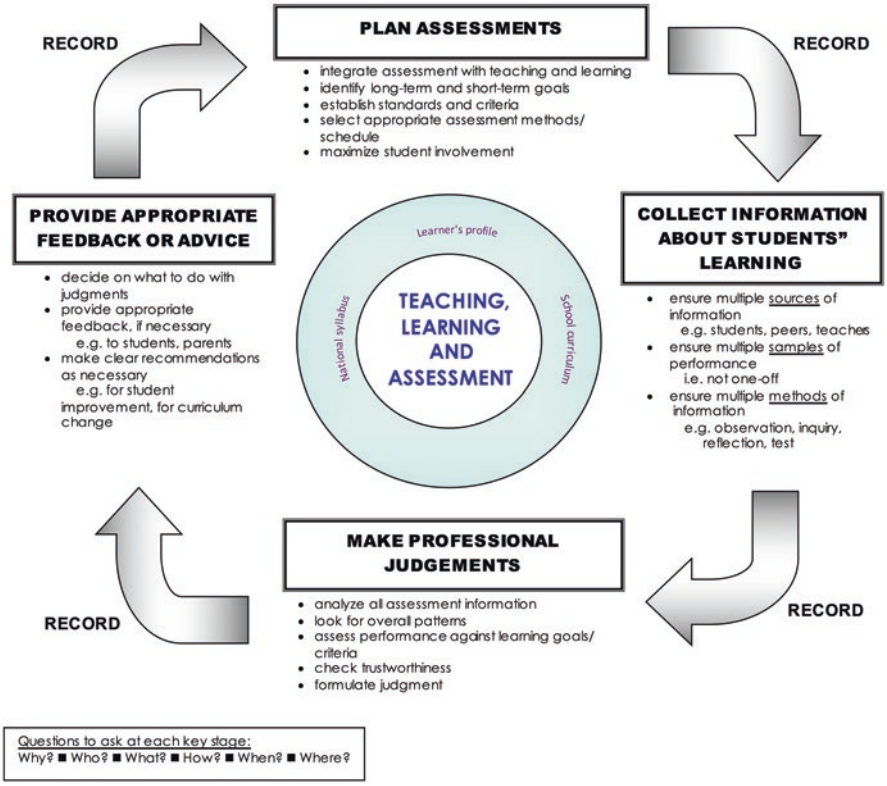


Fig. 1.1 Classroom-based assessment cycle (Davison 2008)

monitoring and regulating their own learners (Dann 2002), a clear framework for classroom-based assessment—learning-oriented assessment—was suggested (Carless 2015a; Purpura 2004; Turner and Purpura 2016). The idea behind LOA is that all classroom-based assessments, be it formative or summative, should be about developing learning (Carless 2015b). The work on LOA resulted in the development of a framework for it, which is also not only compatible with Davison's cycle but can inform its implementation, as it addresses all of its parts.

Chapters 4 and 5 of the volume explicitly build on the LOA framework. We refer the reader to these chapters for a detailed discussion of the framework and its applications. Next, we briefly elaborate on the interrelated dimensions of the framework (Turner and Purpura 2016, p. 261) with reference to the classroom-based assessment cycle:

- contextual dimension refers to factors, such as educational and political, shaping the context and promoting or hindering learning; in Fig. 1.1, it is the circle surrounding teaching, learning, and assessment in the middle;

- elicitation dimension refers to how learners' knowledge and abilities are elicited; in the classroom-based assessment cycle, this dimension refers to the planning and collecting information about learning;
- proficiency dimension is both about standards and how learner knowledge and abilities change over time with reference to these standards; it informs the planning and making of judgements about learning parts of the cycle;
- learning dimension refers to the conceptualisation of learning and, therefore, belongs to the very centre of the cycle;
- instructional dimension is about arranging assessments and how the information is used in the following teaching, stretching thus across the classroom-based assessment cycle;
- interactional dimension was discussed by Turner and Purpura mainly with reference to feedback and interaction between the teacher and learners in the teaching/learning process; however, we suggest it can also refer to assessments happening during the interaction between the teacher and learners or among learners; therefore, it informs both the centre of Davison's (2008) model and collecting information and providing feedback parts of the cycle;
- finally, affective dimension is the inferences emerging in assessment into affective aspects of the classroom teaching, learning, and assessment, such as learner beliefs and engagement.

We consider both assessment *for* learning and the LOA framework as possible and viable ways of informing the classroom-based assessment promoting learning. Both of these frameworks and conceptualisations are represented in the present volume and both align well with the classroom-based assessment cycle, informing it.

Ultimately, the use of assessment promoting learning is dependent on how clearly defined the goals are for the learners (what is learned) and the teacher (what is taught) (see Ball et al. 2015). Providing teachers with the skills to define these goals and to assess learning with reference to these goals in order to guide learners towards them should become the basis for planning assessments (the upper part of Davison's (2008) cycle). As learners need to possess an understanding of the goals aimed for (Sadler 1989), one aim of assessment promoting learning becomes making learners aware of the goals. A shared understanding of assessment criteria and goals becomes essential for the classroom-based assessment cycle, the more so in CLIL assessment. A further aim of assessment promoting learning is helping learners understand assessment as a part of the learning process, not just an opportunity to demonstrate their knowledge (Bransford et al. 2000). As Mehisto and Ting (2017, p. 224) put it, assessment should help "students to become knowledgeable partners in the learning process." with a shared understanding of what achieving these goals entails (the 'how' question in Davison's cycle). It should also help learners understand what constitutes quality.

In CLIL education, determining learning outcomes required to achieve both content and language goals and assessing this achievement is oftentimes challenging. Assessment variety is a *sine qua non* in CLIL, because (1) CLIL lessons have to take into account the wide range of knowledge and skills and (2) students with

different backgrounds need opportunities to demonstrate their abilities in multiple settings (Llinares et al. 2012). Above all, however, we argue that setting the goals should be guided by the recognition that CLIL, regardless of the specific approach it takes, is about *integration* of content and language. We will next elaborate on how we conceptualise CLIL in the present volume and what it means for assessment promoting learning.

1.2 Defining CLIL

Coyle, Hood, and Marsh succinctly defined CLIL as "a dual-focus educational approach in which an additional language is used for the learning and teaching of both content and language" (2010, p. 1). Similarly, Wolff (2007, p. 16) defined it as "any educational situation in which an additional language and therefore not the most widely used language of the environment is used for teaching and learning of subjects other than the language itself." In other words, CLIL is "a generic umbrella term which would encompass any activity in which … both language and the subject have a joint curricular role" (Marsh 2002, p. 58). CLIL is, then, a way of combining subject learning and learning a language that is not the mother tongue of the learners; 'two for the price of one' as Bonnet (2012) put it. That said, the goal of CLIL is to promote both content mastery and language proficiency by teaching a content subject, e.g. biology, through an additional language without hindering the first language. Yet, there is not always agreement about how CLIL is implemented (Clegg 2012) as teaching of content and language varies across contexts.

This is not to say that there is no common ground between the various approaches to CLIL. Nor should it be considered that CLIL is fundamentally different from other educational approaches where content and language are merged, such as immersion programmes. On the contrary, as Cenoz (2015) and Cenoz et al. (2014) argued, at its core, the pedagogical principles of various approaches to CLIL and other approaches to integrating content and language are not that different. Their goal is to provide enhanced opportunities to study an additional language, to master the content, and to acquire social and cognitive skills for successfully operating in the multilingual and multicultural world.

Broadly defined, goals of CLIL are to promote (a) learners' *academic competence,* (b) *proficiency in the L2* (second or foreign language; or additional language), and (c) *L1* (first and/or strongest language) *competence* (Mehisto et al. 2008; Mehisto and Ting 2017). A further goal of CLIL is also to promote the "understanding and appreciation of the *culture* of the L1 group, and of the L2 group(s)", "capacity for and interest in *inter-cultural communication*", and, as an overarching goal, "the *cognitive* and *social skills* and habits required for success in an ever-changing world" (Mehisto et al. 2008, p. 12; emphasis added). These general goals should, ideally, shape teaching, learning, and assessment in CLIL.

Cenoz et al. (2014) argued that it is best to conceptualise CLIL as an umbrella for all different approaches to integrating content and language at the same time calling

for a carefully constructed taxonomy of different approaches to CLIL. We follow their proposition for considering CLIL as an umbrella for a variety of classrooms where content and language are taught simultaneously, albeit (ideally) informed by the general goals of CLIL that we outlined above.

The contributors to this volume approach assessment promoting learning in CLIL from rather different angles and foci. However, what unites them is their focus on the *integrated nature* of content and language, which is the central premise of the volume. The instructional approaches discussed in the volume are not about suppressing other languages, i.e. for the most part, learners' mother tongues, but to enrich learner linguistic repertoire. More importantly, their premise is *simultaneous* teaching and learning (and assessment) of the language and the content—*a fusion of content and language* (Nikula et al. 2016).

Before discussing the integration in CLIL, we will outline the differences and similarities across the contexts discussed in the chapters of the present volume with reference to the geographical contexts they represent.

1.3 Geographical Contexts

Educational policy, state curricula, as well as educational history, ideologies, and beliefs strongly shape teaching, learning, and assessment in CLIL classrooms. The role of covering the educational contexts in the volume from a broader geographical perspective is to give a general understanding of how these factors inform approaches to integrating content and language in teaching, learning, and assessment. This volume brings together contributors from different regions of the world: Finland (Chaps. 6 and 9), Japan (Chaps. 3, 4, 7, and 10), Portugal (Chap. 5), Spain (Chap. 8), Thailand (Chap. 3), and the UK/international[1] (Chap. 2). Two foci become important in discussing geographical contexts: (1) educational history, beliefs, and ideologies, manifested in curricula and (2) the outcome of these with relation to CLIL—the relative emphasis on content or language in various geographical contexts. Next, we visualised how the countries are positioned west to east, which coincidentally roughly matches the change of emphasis from content to language in CLIL instruction.

Spain has become one of the European leaders in developing CLIL education (Coyle 2010). There is a variability in how CLIL is organised, as each of the 17 autonomous regions in Spain regulates education independently. Over several decades, bilingual CLIL-style education has been implemented at a progressively increasing rate within Spanish Autonomous Communities. To try to improve foreign language competence, Spain launched language-learning initiatives in the 1990s focusing on English, the most commonly studied language in Europe

[1]As a matter of fact, the context in Chap. 2 is rather international (Cambridge Assessment International Education). However, as the contributor is based in the UK, we suggest that both of these contexts should be acknowledged.

(Eurydice 2017). In 1990, an educational reform act was passed (*Ley Orgánica de Ordenación General del Sistema Educativo*), requiring the introduction of a first foreign language by the age of eight (Gobierno de España, Ministerio de la Presidencia 1990). As a result, in 1996, the British Council in conjunction with the Ministry of Education and Science launched the Bilingual Project, implementing CLIL in schools in Madrid. The project has been successful in promoting its educational purpose and preparing the younger generation to enter the increasingly multilingual European workforce in the future (Reilly and Medrano 2009). Teachers trained in content areas have been required to impart their subjects in the target language, requiring teachers to undergo in-service training to improve their language skills. Many teachers—trained in subject area expertise rather than language teaching—see themselves as content experts and regard the changes to the system as challenging (Muñoz and Navés 2007).

In *Portugal,* CLIL instruction is not widely spread. Still, it has been supported by the Ministry of Education through several projects and CLIL programmes implemented from the pre-primary through to higher education. While there are official guidelines for implementing CLIL there are still a variety of approaches to CLIL, with varying degrees of content and language integration. A process of curriculum redesign has been underway in Portugal since 2017/2018 (OECD 2018). This process has had implications in the launching of new curriculum documents aiming at providing students' profiles at the end of their compulsory schooling. This was done to enable deeper learning by identifying and developing core learning skills across different subject areas in the compulsory curriculum. This has significance for CLIL, as subject-related content areas are being used for the purpose of promoting learning depth and developing higher levels of language proficiency. The pilot project (Bilingual Schools Project; BSP) and the programme were jointly implemented by the Directorate-General for Education of the Portuguese Ministry of Education and the British Council Portugal. The programme is aligned with the curriculum, setting out what should be learned and thus shaping teaching and assessment. In the programme, language education has been integrated with the content lessons to encourage literacy development that stretches beyond the everyday foreign language to that required to convey content meaning and ensure "ongoing language growth (being alert to plateauing)" (Bertaux et al. 2010, p. 8).

In *Finland*, CLIL has a long history, being used as an educational approach since early 1990s. It is now acknowledged in both the national and local curricula, though under the more general umbrella of bilingual education. At the same time, it is still rather loosely defined in the National Curriculum. CLIL instruction generally appears to be content-driven. However, locally, on the municipality and school level, there is variation in the ways that CLIL is implemented. For example, there are a number of projects, applied for by municipalities, in which immersion and bilingual education programmes are developed (Peltoniemi et al. 2018; see City of Helsinki 2019). We should add that in Finland, in general, teachers enjoy freedom in how they implement the curriculum, which adds to the variation. Chapter 6 provides an excellent overview of CLIL in the Finnish context and gives a detailed

discussion of Finland's local educational policies and practices and their interrelationships.

In *Thailand*, CLIL is slowly gaining recognition at all levels of education. A pilot study in 2006–2007 was used to implement CLIL science courses at primary and secondary schools that followed the national curriculum (Keyuravong 2010; Marsh and Hood 2008). More recently, researchers have investigated the effectiveness of CLIL in the context of Thai universities (Lai and Aksornjarung 2017; Suwannoppharat and Chinokul 2015), arguing for moving away from the still ubiquitous grammar-translation method.

CLIL in *Japan* is still "a new-born" (Ikeda et al. 2013, p. 1), and the emphasis is on language, but CLIL continues to develop there, attempting to fracture the constraints of the traditional norms (Clavel 2014). English as a Foreign Language (EFL) pedagogy is generally emphasised, low levels of English proficiency of students and teachers alike is prevalent, and the textbook often becomes the authority in the classroom. This limits activities where content and language can be taught, learned, and assessed. There is a growing argument for the necessity of bilingual education; yet, CLIL is not systematically represented in the national curriculum and is largely implemented at the grassroots level (Ito 2018). Many CLIL (similarly to EFL) classes are taught primarily using the L1 Japanese. Exacerbated by a typological distance between Japanese and English (Cenoz 2017; Jackson and Kaplan 1999), this prevents learners from reaching higher levels of linguistic proficiency. To compensate, Japan has created a modified version (CEFR-J) of the Common European Framework of Reference, to include sub-levels (pre-A1) (Negishi 2011). Globalisation (METI 2010) has played a part in impacting Japan's educational policy and practice at the tertiary level, which has contributed to the rise of some pioneering programs, especially in the sciences, and a rise in English as a Medium of Instruction (EMI) programs.

The complexity of what shapes a particular approach to CLIL is not limited by the local educational context. Grin, (2005), for example, stated that up to 216 forms of CLIL can be identified, geographical context playing but a small part in this differentiation.

1.4 Hard to Soft

One difference with regard to various CLIL implementations is in their relative emphasis on content or language. It has been proposed to capture this varying emphasis of content-led to language-led approaches is a continuum from hard (or strong) CLIL to soft (or weak) CLIL (Ball et al. 2015; Met 1999). Hard CLIL "refers to subject-based aims and objectives, where subjects from the conventional curriculum are taught in an additional or foreign language" (Ball et al. 2015, p. 27). Soft CLIL is "used to describe the broad linguistic aims that a language teacher brings to the classroom" (Ball et al. 2015, p. 27) and "content is a useful tool for furthering

Fig. 1.2 The *hard* and *soft* CLIL continuum. (Adapted from Ball et al. 2015)

the aims of the language curriculum" (Met 1999, p. 5). The following Fig. 1.2 illustrates this continuum.

The definition of CLIL we adopt in the present volume sets boundaries to this continuum. English as a foreign language (EFL), while being language-driven, is outside the boundary of the soft-CLIL end of the continuum, as it does not perceive integration of content and language as a goal. However, Chap. 7, discussing a Japanese context that is essentially EFL with some emphasis on the content, is a part of the present volume, as it builds the argument for the value of integrating content with language.

Defining CLIL as an educational approach and the relative emphasis on content or language contextualised based on a region of the world is a good starting point to identify how one 'fits' into the CLIL community. Yet, the emphasis on content or language in CLIL depends on the local educational culture and tradition, if and by what means CLIL is reflected in the National Curricula, who teaches CLIL (the content or language teacher, or both), the starting age of CLIL instruction, and the subjects and number of hours taught in a foreign language. However, it also depends on teachers' understanding of the local educational policies and of their experience of and beliefs about teaching content and language simultaneously.

Based on the hard-to-soft continuum, CLIL approaches explored in the present volume can be divided into those studying and discussing contexts (not necessarily geographical) with a greater focus on the content (Chaps. 6, 9, and 10), those focusing more on language (Chaps. 7 and 8), and those scattered across the continuum while not approaching the ends of it (Chaps. 2, 3, 4, and 5). However, it does not help to fully understand what it is that is learned in CLIL and how to assess such that this learning is promoted. Assessment activities built on the principles of A*f*L and/or following the LOA framework can be useful for understanding where learners are, where they are heading, and how they can be guided in their learning. However, this focus is not enough to fully inform the classroom-based assessment cycle. We argue that it is the understanding of *integration* of content and language as "mosaics of different pedagogies and learning practices" (Leung and Morton 2016, p. 248) which can truly help conceptualise teaching, learning, and assessment in CLIL. We will discuss the basis for the conceptualisation of integration of content in language in Sect. 1.7. However, connected to the discussion of the hard-to-soft continuum, we feel we need first to add a brief discussion of the role of language in CLIL.

1.5 The Role of Language in CLIL

CLIL is concerned with the development of learners' academic language alongside content knowledge. Cummins (1999) proposed two dimensions of academic proficiency: Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Linguistic Proficiency (CALP).

BICS is what people use in their everyday conversations. This kind of language is easier to use and acquire, as it is context embedded. In discussions involving BICS, we can see gestures, facial expressions, and other clues about the details of the conversation. CALP on the other hand is context-reduced. When learners see a page in a textbook, they are confronted with more abstract language and less clues to the meaning (Lin 2016). The natural progression for learning language is through the development of BICS, which occurs much more quickly and easily. Even for L1 speakers, mastery of CALP requires explicit instruction.

Cognitive discourse functions (CDFs), e.g. classifying, defining, describing, evaluating, explaining, exploring, and reporting, is another way of looking at the role of language in CLIL (Dalton-Puffer 2007, 2016). Essentially, CDFs, based on Bloom's Taxonomy of Educational Objectives (Anderson et al. 2001; Bloom et al. 1956), provide a perspective that when teachers are "modelling/teaching how to verbalise subject-specific cognitive actions, they are not 'doing the language teachers' job', but actually teaching their subject in a very substantial way" (Dalton-Puffer 2016, p. 30).

Chapter 6 gives a detailed overview of the so far scarce existing research on language and language assessment in CLIL. Hence, we will not in this chapter provide these details. In the following section, we, rather, discuss how different ways content and language are integrated in CLIL change the conceptualisation of teaching, learning, and assessment and give examples of these.

1.6 The Integration Matrix

To understand what is learned, taught, and assessed in CLIL classrooms, integration of content and language should be considered. Our following discussion will be based on Leung and Morton's (2016) integration matrix, as their conceptualisation underscores the differences in teaching and learning across the four ways content and language integration can be understood.

Leung and Morton's (2016) discussion of learning from the perspective of *integration* in CLIL is informed above all by Bernstein's (2000) concepts of 'competence' and 'performance'. A competence approach refers to further developing learner's existing abilities. A performance approach is one where the learner is expected to demonstrate knowledge or skills based on some inventory of standards such as summative assessment criteria that can be used to share learner performance across contexts (see Christodoulou 2016, p. 56).

To discuss what is learned in CLIL and how, Leung and Morton build on the concepts of *classification* and *framing*. *Classification* refers to the separation of language based on subjects such as history or math and *framing* refers to the path in which acquisition can occur within that classification, or how the content is selected and sequenced (Leung and Morton 2016). Bernstein (2000) talks about a *visible* pedagogy when it is performance oriented and strongly classified and framed. In *invisible* pedagogy, on the other hand, "rules of organisation and criteria (are) implicit" (p. 109) and classification is weak.

The second conceptualisation that Leung and Morton draw on is Bakhtin's (1981) discussion of tensions between centripetal and centrifugal tendencies in language (and by extension, teaching and learning of content). Centripetal tendencies can be visualised as language development that takes 'subject-literacies', i.e. language of the subject is required to make meaning and the language of the classroom is a more unified language. These tendencies are more likely in visible pedagogy. Centrifugal tendencies on the other hand are a more diverse use of language in the classroom, where learners and the teacher could be discussing the same concepts using very different language.

Based on this conceptual understanding, Leung and Morton (2016) produced a matrix of four different orientations to integration in CLIL based on the degree of visible language pedagogy intersecting with the degrees of disciplinary orientations to language (Fig. 1.3).

We will next propose the *kinds of learning* that could occur in each of the quadrants, keeping in mind that "the boundaries demarcating the four quadrants are leaky" (p. 237). That is, we will use Leung and Morton's matrix as a lens to inform the understanding of conceptualise teaching, learning, and assessment in CLIL in terms of the integration of content and language in the chapters in the present volume.

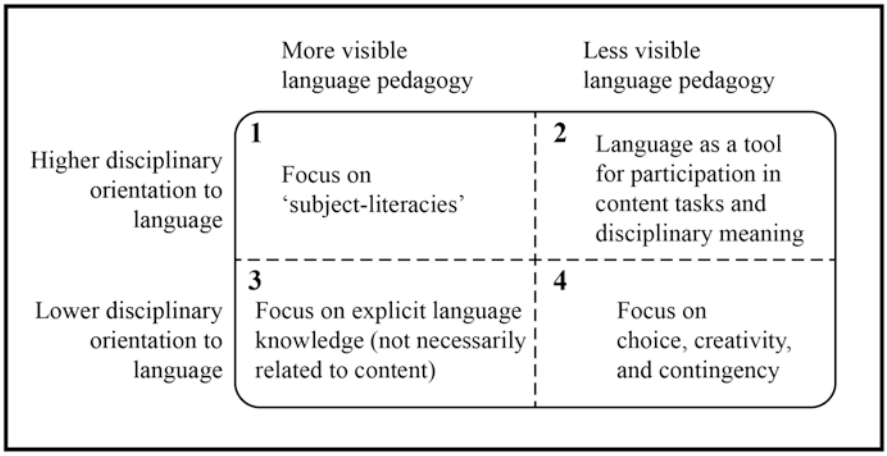


Fig. 1.3 Four different orientations to integration in CLIL (Leung and Morton 2016, p. 237)

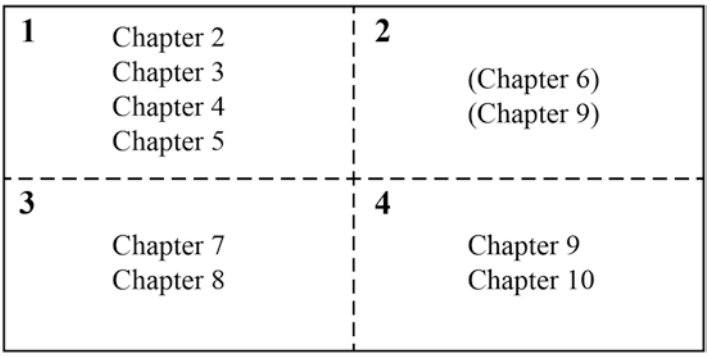


Fig. 1.4 Chapters placed on Leung and Morton's (2016) integration matrix

The matrix can help make useful generalisations with regard to describing a CLIL context, i.e. to which quadrant a CLIL context—geographical, school, or classroom—'belongs'. With this in mind, to help the reader navigate through the volume, we next tentatively 'position' each contribution on the matrix (Fig. 1.4).

Figure 1.4 served to help us organise the order of chapters in the volume. As we discussed previously, no chapter argues for the value of integration in Quadrant 2. Still, even though Chap. 6 argues for including systematic assessment of language to content-focussed CLIL, the picture of CLIL teacher practices that emerges in Chap. 6 refers to this type of integration (or lack of it). Hence, we positioned it in the volume to reflect this quadrant. We underscore it that at the micro level, content and language can be integrated in a variety of ways even within one and the same lesson, not to mention across assessment cycles. Hence more nuanced use of the matrix is needed, which we outline in the following section.

1.7 Teaching, Learning, and Assessment in the Matrix

We next elaborate on how we use the matrix to conceptualise teaching, learning, and assessment in CLIL. We will argue that different conceptualisations of integration in CLIL change how teaching, learning, and assessment are understood, as we elaborate next.

1.7.1 Quadrant 1: Visible Language Pedagogy and Higher Disciplinary Orientation to Language

First and foremost, this quadrant represents the development of language competence that is specific to meanings associated with content (Leung and Morton 2016). In the learning of a specific subject, such as mathematics or history, a syllabus will

have clear descriptions of what is to be acquired in language and content as strong classification criteria, and there will be strong framing, outlining clear stages of how the content and language are to be acquired. In this sense, linguistic competence is merged with the content competence—a visible pedagogy where learning is based on a performance orientation. CLIL is timetabled as content lessons in which the language of the subject is taught (Dalton-Puffer et al. 2014).

In this Quadrant, the achievement of learning goals of CLIL can be visualised through the content and language developing in tandem in stages. Reaching the following stage presupposes that the preceding stage has been achieved. Learning is demonstrated through performance on content tasks with a clear understanding of linguistic constraints impeding learners' mastering of the content. Thus, explicitly addressing learners' linguistic needs is required, and language specific descriptors for different stages should be designed and integrated with content goals. A major instructional goal is then to create links between lexical items and the ability to appropriate them to understand and discuss the content-specific knowledge (Dalton-Puffer 2016). The language of the classroom is subject oriented, making that content language necessary and an integral part of the construct taught and learned.

At a higher level of abstraction, in approaches to CLIL belonging to Quadrant 1, learning can be conceptualised as happening when learners are developmentally ready for it. The aim of assessment then becomes to find out whether learners are ready to move to a higher stage. However, other conceptualisations of learning are, too, possible.

Classroom-based assessment that can be associated with this quadrant (e.g. those discussed in Chaps. 2 and 3) is often informed by scales and descriptors such as those in the Common European Framework of Reference CEFR (Council of Europe 2001, 2018). These descriptors allow for establishing learner progression in broad terms, allowing for meaningfully establishing whether teaching and learning goals are reached.

The philosophy behind the CEFR is, above all, that of constructivism. Two main lines of conceptual inquiry characterise constructivism: the *cognitive constructivist* perspective and the *social-cultural constructivist* perspective. The concept of *cognitive constructivism* is associated with Swiss developmental psychologist Jean Piaget (1977). The understanding behind it is that learners cannot simply be given information which they immediately understand and use, but rather must construct their knowledge. Thus, learning occurs by active construction of meaning. Still, development is characterised by what individuals can and cannot do at different stages. Learning lags behind development. Teaching should target that what the learner is developmentally ready to learn, and assessment yields insights into this.

However, the development of the CEFR has also been informed by *social constructivism*—a philosophy championed by Lev Vygotsky (1986) in which the social and collaborative nature of learning in the development is stressed. Learners are active agents in the construction of meaning—knowledge is socially constructed. The process of knowing, therefore, is influenced by other people and is mediated by culture. Vygotsky's well-known concept of the *zone of proximal development* (ZPD) helps to understand how learning is conceptualised in social constructivist thinking.

ZPD denotes the range of what a learner is currently able to internalise with the assistance of a more knowledgeable interlocutor, such as a teacher (the potential development). According to Vygotsky (1978, p. 57), "Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological)." Learning, therefore, shapes development, as what a learner is able to do now with the assistance from others, the learner will be able to do independently in future. Development, in turn, directs teaching and learning, as what is taught and learned should build on the learner's development, teaching and learning happening in the learner's ZPD. Chapters 2 and 9 provide more discussion of this.

In the volume, Chaps. 2 and 3, and to a large extent, Chaps. 4 and 5, discuss this type of integration.

1.7.2 Quadrant 2: Less Visible Language Pedagogy and Higher Disciplinary Orientation to Language

In this quadrant, language is a semiotic tool to promote learning of "discipline-specific concepts and competences" (Leung and Morton 2016, p. 240), making the language not fully integrated but learned as a "side effect" (p. 240). Learners build their cognitive academic linguistic proficiency (CALP) through discussing scientific concepts in a specific discourse community. That is, the language is acquired through reaching the content-specific goals attained through *using* the language, the language pedagogy being more or less invisible.

There are no contributions in this volume advocating for assessment that can be conceptualised with reference to this quadrant. This is understandable, as in this quadrant, language is the least integrated into the process of acquisition of discipline-specific knowledge and competences. It goes counter to the perspective that the contributors to this volume adopt. That said, several contributions (Chaps. 6 and 9) observe how teachers struggle with integrating language into the content-specific knowledge, Wewer making it the focus of her Chap. 6. The authors use this as a starting point for their discussion and/or propose how assessment can be reconceptualised to make connections between the language and the content in assessment and teaching practices.

The goal of assessment in this quadrant is to collect information about learners' content knowledge. The language component in assessment is unsystematic at best, if present at all (Wewer 2014; see Chap. 6, this volume). Still, teachers scaffold language to make the content accessible to learners, but this scaffolding is based on implicit understanding of learner needs (see Lin 2016, for examples of assessment activities where language is scaffolded). Indeed, as Mehisto and Ting (2017, pp. 220–222) argued, if the focus of assessment is on content, it is essential to make sure that language does not hinder learners' capacity to demonstrate their content knowledge. Regardless, content teachers can be concerned whether or not teaching

and assessing language can be problematic, but at the same time fearing that they are not qualified to teach it or assess it (Skinnari and Bovellan 2016; Chaps. 6 and 9, this volume).

1.7.3 Quadrant 3: More Visible Language Pedagogy and Lower Disciplinary Orientation to Language

Next, there are CLIL approaches with a "highly visible language pedagogy without the linguistic elements being tightly linked to any specific discipline" (Leung and Morton 2016, p. 237). These approaches are usually performance- rather than competence-based, language being taught explicitly.

CLIL approaches in this quadrant are language-focused. There can be explicit emphasis on the four skills (listening, speaking, reading, and writing) and the focus can be on the language of lower order thinking skills (LOTS, Anderson et al. 2001). Teachers attempt to bring the language focus into the content areas, using authentic situations and building language learning around specific topics. Teachers, then, have to make an effort to systematically incorporate cognitive skills (e.g. integrating information) with language (e.g. functions, vocabulary, as well as grammar), fore fronting higher order thinking skills (HOTS) (Anderson et al. 2001), e.g. explaining, synthesising, or discussing concepts. This approach to language is functional, oftentimes, CDFs informing the understanding of how knowledge is constructed in CLIL lessons. In addition to focusing on form, lessons can be speech events, with teachers or more capable peers mediating the performance of novices—"an implied situated learning perspective" (Leung and Morton 2016, p. 241).

The distinction between the two general ways language teaching and learning can happen in this quadrant is best understood with reference to centripetal and centrifugal tendencies in language (Leung and Morton 2016; Bakhtin 1981; see Sect. 1.6). When *centripetal tendencies* are more likely, learners learn to speak the language of the subject, the language of the content embedded in the concepts themselves, with little room for 'saying it a different way'. *Centrifugal tendencies* in this quadrant have to do with the situatedness of the interaction, as learners' "competence is developed through participation in social practices" (Leung and Morton 2016, p. 242). These social practices are still mediated by the teacher (a more capable other), having an explicit focus on language. However, teachers build on what emerges in the interaction.

In CLIL classrooms, learners can use language very differently from the teacher to discuss the same concept; e.g. 'a round thing' *versus* a 'circle'; and the teacher can build on learners' contributions in order to guide learners to using academic language equipped with conceptual understanding. With centripetal tendencies, the language of the genre is the language the student uses. In centrifugal tendencies, the learners are able to use their everyday language to discuss academic language (Leontjev and deBoer 2020). CLIL teachers, in turn, mediate the use of the

language to verbalise content, through this, making the content alongside language accessible to the learners. To summarise, how exactly linguistic goals are achieved in this quadrant varies, but content knowledge is acquired through language instruction.

With regard to assessment, there is an emphasis on language, the purpose being to help learners become "socialised into rational academic discourse" (Dalton-Puffer 2016, p. 42) through promoting their linguistic knowledge. As mentioned previously, there is often strong framing, so progress can be measured through tests and with the help of scales. However, as Dalton-Puffer (2016) also mentioned, assessment of learners' abilities can also happen in dialogic interaction with more capable peers. Hence, teachers can use assessment as an inquiry (Birenbaum 2014) during the interaction with their learners in order to guide the learners as a part of the interaction. In the present volume, assessment as a part of interaction in Quadrant 3 is discussed in Chap. 8. Chapter 7 discusses assessment as part of both centrifugal and centripetal teaching and learning processes.

1.7.4 Quadrant 4: Less Visible Language Pedagogy and Lower Disciplinary Orientation to Language

In a less visible language pedagogy with weak classification and framing, the learner is often seen as "the author of the practice and even the authority" (Bernstein 2000, p. 110), eliciting centrifugal interaction, meaning that the dialogue between the teacher and the learners or among learners is "more open to contingent notions of communication" (Leung and Morton 2016, p. 236). This is a competence approach to learning, emphasising learner autonomy and agency in which the content and linguistic goals are more loosely defined. In this approach, learning is situated, meaning that it is a process of discovering content and language with learners co-constructing knowledge through dialogic interaction.

The underlying principle of assessment promoting learning in this quadrant is giving learners as much responsibility for their learning and development as possible, instruction becoming a dialogical learning-oriented process. Assistance from the teacher/assessor or peers emerges during the interaction. The availability of external support provides a fuller picture of learners' abilities, as it uncovers what learners can do with external support, simultaneously promoting their abilities. In language education, this conceptualisation of assessment has been pursued in dynamic assessment (Poehner 2008; Poehner and Infante 2015), which considers assessment as inseparable from teaching and learning, all three forming one development-oriented process. Learners' performance is not about what the learners can or cannot do but about what they can do when assistance from others is available.

At the level of the classroom-based assessment cycle, one can perceive these dialogical processes of teaching, learning, and assessment as one activity

('collecting information about students' learning' part of Fig. 1.1). This allows for making inferences about where the learner is in relation to their goals and what the sources of their struggles are, and how much effort on the part of the teacher is required to help individual learners or groups of learners develop. This serves as the basis for the subsequent feedback to learners and for adjusting the following teaching, learning, and assessment. Assessment activities in this quadrant can be both planned (e.g. a planned dynamic assessment activity) and unplanned (the teacher using opportunities for assessing as these are afforded in classroom interaction), but the way the assessment activity develops is always contingent on the interaction, assessment emerging in it. To summarise, in assessment promoting learning in this quadrant, performance is no longer individual but is co-constructed between the teacher and the learner (Chaps. 8 and 9) or among learners (Chap. 10).

1.8 Pulling the Threads Together

The contributors to this volume come from different angles with regard to their approaches to assessment. Yet, they all perceive assessment as a part of the process whose goal is to promote learning, assessment mediating the teaching and learning relationship. We argue that the synthesis of the integration matrix and the classroom-based assessment cycle informs this relationship. The central question with regard to classroom-based assessment, i.e. *What can you learn about your learners and what can learners learn as a result of using different assessment tools and approaches?* can be guided by the assessment cycle. The matrix, too, informs the answer to this question with regard to CLIL classrooms.

To repeat, classroom-based assessment should promote learner autonomy and help teachers to adjust teaching to address learner needs in relation to the goals of learning (Black et al. 2004). As Mehisto and Ting (2017, p. 224) put it, assessment should help "students to become knowledgeable partners in the learning process." In other words, the aim of assessment promoting learning in CLIL, regardless of whether the principles of A*f*L inform it or a framework such as LOA is used, is to encourage learners to take greater control of their learning of both content and language. Setting the goals becomes essential in this process. Here, the matrix can inform the thinking about the goals at the macro and the micro level. The macro level is shaped by the institution, curriculum and policy guidelines, and the educational tradition. We roughly operationalised this level as the geographical context. The micro level refers to how these goals are realised in specific classrooms as adjustments are made to teaching and assessment across assessment cycles.

To give an example, Japan, as we discussed earlier, has a long history of form-focused instruction in language education determined by the curriculum set by the Ministry of Education (MEXT 2015). Using the matrix lens, the macro CLIL context would be identified as Quadrant 3, the students and teachers focusing exclusively on the language. Likewise, the approach to CLIL in Spain, varies across the

country, though one generalisation that can be made is that many teachers consider themselves as content teachers, prioritising content goals (Quadrant 2).

At the micro level, however, the classroom-based assessment can show a markedly different picture. Chapter 8 gives an excellent example of how, at the micro level, instances of dialoguing with learners are language-oriented and best understood through the lens of Quadrant 3. The quadrant boundaries are, indeed, leaky and even in the context of the CLIL classroom which is content oriented, instances of language focus occur.

A teacher can use, for example, a rubric based on a proficiency scale (similar to those discussed in Chaps. 2, 3, and 4) to determine which problems learners have and gain an understanding of where learners are in terms of their content and language knowledge (Quadrant 1). Equipped with these insights, the teacher can then further probe and direct learners' development in the subsequent integration, using a more structured approach informed by the principles of A*f*L and focussing more on the language which implies a more structured approach even if centrifugal tendencies too, are visible in interaction (Quadrant 3; see Chap. 8, this volume). Alternatively, more centrifugal approaches to interaction between a teacher and learners (Chap. 9) or among learners (Chap. 10) can be used to guide the focus of teaching, learning, and assessment (Quadrant 4). The insights obtained in these previous assessment cycles can then help teachers to interpret learners' performance on essays focusing on content (Quadrant 2), helping teachers to understand whether learners display their own knowledge or repeat somebody else's words or understand the reasons for problems in their writing (Chap. 9). This will direct the following teaching, learning, and assessment and the cycle will, thus, continue. A framework classifying various approaches to assessment promoting learning in CLIL classrooms, such as that discussed in Chap. 7, can inform the understanding of these assessments with reference to the matrix, the agents assessing and the functions of these assessments (summative and formative), and the goals that are to be achieved. In other words, at the macro level, the goals stay the same, but the way these goals are reached can vary based on the unique ways that assessment cycles in CLIL classrooms inform the trajectory of these classrooms.

We realise that the reader might want to read the contributions to the volume with the focus on specific geographical contexts, different stages of education (primary, secondary, or tertiary), a specific discipline, particular parts of the classroom-based assessment cycle, or specific quadrants on Leung and Morton's (2016) integration matrix. However, as we elaborated, one of the main purposes of classroom-based assessment cycle is to elicit the relationship between teaching, learning, and assessment. Likewise, the matrix conceptualises more than the integration between content and language in CLIL classrooms. It also outlines how learning is conceptualised within each of the quadrants, informing the choice of assessment tools and approaches, what knowledge, understanding, and use are elicited, how learner performance is interpreted, and what feedback to learners is given. The two models together guide the understanding of how teaching, learning, and assessment change across assessment cycles. We hope that the reader goes through the volume with this understanding in mind.

References

Alderson, J. C., Haapakangas, E.-L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2015). *The diagnosis of reading in a second or foreign language*. New York: Routledge.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Allyn Bacon Longman.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Baird, J., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy and Practice, 24*, 317–350.

Bakhtin, M. M. (1981). *The dialogic imagination. Four essays*. Austin: University of Texas Press.

Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice*. Oxford: Oxford University Press.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles Policy and Practice, 18*(1), 5–25.

Bernstein, B. B. (2000). *Pedagogy, symbolic control, and identity: Theory, research, critique* (2nd ed.). Lanham: Rowman & Littlefield Publishers.

Bertaux, P., Coonan, C. M., Frígols, M. J., & Mehisto, P. (2010). *The CLIL teachers' competence grid*. http://tplusm.net/CLIL_Competences_Grid_31.12.09.pdf. Accessed 5 Nov 2018.

Birenbaum, M. (2014). Conceptualizing assessment culture in school. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 285–302). Dordrecht: Springer.

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 92*(1), 81–90. https://doi.org/10.1177/003172171009200119.

Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Berkshire: McGraw-Hill Education.

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), 8–21.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: The cognitive domain*. New York: David McKay.

Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal, 4*(1), 66–78.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Exp ed.). Washington, DC: National Academy Press.

Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. Cambridge: University of Cambridge School of Education.

Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River: Pearson.

Carless, D. (2015a, July). *Learning-oriented assessment in practice*. Urbana: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Carless, D. (2015b). Exploring learning-oriented assessment processes. *Higher Education, 69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z. Accessed 8 Dec 2018.

Cenoz, J. (2015). Content-based instruction and content and language integrated learning: The same or different? *Language, Culture and Curriculum, 28*(1), 8–24. https://doi.org/10.1080/07908318.2014.1000922.

Cenoz, J. (2017). The target language, the sociolinguistic and the educational context in CLIL programs. In A. Llinares & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 239–249). Amsterdam: John Benjamins.

Cenoz, J., Genesee, F. H., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics, 35*(3), 243–262. https://doi.org/10.1093/applin/amt011.

Christodoulou, D. (2016). *Making good progress? The future of assessment for learning*. Oxford: Oxford University Press.

City of Helsinki. (2019). *Bilingual education*. https://www.hel.fi/helsinki/en/childhood-and-education/comprehensive/what-how/bilingual-education/. Accessed 8 May 2020.

Clavel, T. (2014, January 6). English fluency hopes rest on educational overhaul. *The Japan Times Online*. https://www.japantimes.co.jp/community/2014/01/05/issues/english-fluency-hopes-rest-on-an-education-overhaul/#.WmWVOkuYMkg. Accessed 4 Sept 2019.

Clegg, J. (2012). Preface. In Quartapelle, F. (Ed.), *Assessment and evaluation in CLIL* (pp. 5–7). Aeclil.altervista.org. Accessed 12 May 2018.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Modern Languages Division, Council of Europe.

Council of Europe. (2018). *Common European Framework of references for languages: Learning, teaching, assessment*. Companion Volume with New Descriptors. Provisional Edition. Language Policy Programme. Education Policy Division. Education Department. Council of Europe. www.coe.int/lang-cefr. Accessed 19 Mar 2019.

Coyle, D. (2010). Assessment issues in CLIL. In D. Coyle, P. Hood, & D. Marsh (Eds.), *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Cummins, J. (1999). *BICS and CALP: Clarifying the distinction*. Eric. https://doi.org/ED419696.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. Amsterdam: John Benjamins Publishing Company.

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualizing integration in CLIL in multilingual education* (pp. 29–54). Bristol: Multilingual Matters.

Dalton-Puffer, C., Llinares, A., Lorenzo, F., & Nikula, T. (2014). "You can stand under my umbrella": Immersion, CLIL and bilingual education. A response to Cenoz, Genesee & Gorter (2013). *Applied Linguistics, 35*(2), 213–218. https://doi.org/10.1093/applin/amu010.

Dann, R. (2002). *Promoting assessment as learning: Improving the learning process*. London: Routledge/Falmer.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing: Studies in language testing 7*. Cambridge: Cambridge University Press.

Davison, C. (2008, March). *Assessment for learning: Building inquiry-oriented assessment communities*. Paper presented at 42nd annual TESOL convention and exhibit, New York, NY.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education, 64*(4), 356–372. https://doi.org/10.1177/0022487113488144.

European Commission/EACEA/Eurydice. (2017). *Key Data on teaching languages at school in Europe – 2017 edition* (Eurydice Report). Luxembourg: Publications Office of the European Union.

Genesee, F., & Hamayan, E. (2016). *CLIL in context: Practical guide for educators*. Cambridge: Cambridge University Press.

Gobierno de España, Ministerio de la Presidencia. (1990). *Ley Orgánica de Ordenación General del Sistema Educativo*. http://www.boe.es/buscar/doc.php?id=BOE-A-1990-24172. Accessed 29 June 2010.

Grin, F. (2005). Added value of CLIL. In *Paper presented at the conference the changing European classroom – The potential for plurilingual education* (pp. 9–11). Luxembourg.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487.

Hill, K. (2017a). Language teacher assessment literacy – Scoping the territory. *Papers in Language Testing and Assessment, 6*(1), iv–vii. http://www.altaanz.org/uploads/5/9/0/8/5908292/2.2017_1_introduction_final_final.pdf. Accessed May 14 2018.

Hill, K. (2017b). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment, 6*(1), 1–17. http://www.altaanz.org/uploads/5/9/0/8/5908292/3.si1hill_final_formatted_proofed.pdf. Accessed 26 Apr 2018.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420. https://doi.org/10.1177/0265532211428317.

Huhta, A., Harsch, C., Leontjev, D., & Nieminen, L. (monograph in preparation). *Diagnosing SFL writing*.

Ikeda, M., Pinner, R., Mehisto, P., & Marsh, D. (2013). Editorial. *International CLIL Research Journal, 2*, 1. http://www.icrj.eu/21/editorial.html. Accessed 25 Oct 2019.

Ito, Y. (2018). CLIL in practice in Japanese elementary classrooms: An analysis of the effectiveness of a CLIL lesson in Japanese traditional crafts. *English Language Teaching, 11*(9), 59. https://doi.org/10.5539/elt.v11n9p59.

Jackson, F. H., & Kaplan, M. A. (1999). Lessons learned from fifty years of theory and practice in government language teaching. *Georgetown University Round Table on Languages and Linguistics, 1999*, 71–87.

Keyuravong, S. (2010). Insights from Thailand. In *Learning through English: Policies, challenges and prospects: Insights from East Asia* (pp. 69–95).

Lai, Y., & Aksornjarung, P. (2017). The effects of content-based instruction on listening and speaking abilities of Thai EFL university students. *Journal of Creative Practices in Language Learning and Teaching, 5*(2), 1–20.

Leontjev, D., & deBoer, M. A. (2020). Multimodal mediational means in assessment of processes: an argument for a hard-CLIL approach. *International Journal of Bilingual Education and Bilingualism*, 1–17

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol/Buffalo: Multilingual Matters.

Lin, A. M. Y. (2016). *Language across the curriculum & CLIL in English as an additional language (EAL) contexts: Theory and practice*. Singapore: Springer.

Llinares, A., Morton, T., & Whittaker R. (2012). *The Roles of Language in CLIL*. Cambridge: Cambridge University Press.

Lynch, B. K. (2001). Rethinking assessment from a critical perspective. *Language Testing, 18*(4), 351–372. https://doi.org/10.1177/026553220101800403.

Marsh, D. (2002). *CLIL/EMILE – The European dimension: Actions, trends and foresight potential*. Brussels: European Commission.

Marsh, D., & Hood, P. (2008). *Content and language integrated learning in primary East Asia contexts (CLIL PEAC)*. In The proceeding of primary innovations regional seminar (pp. 43–50).

McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing, 18*(4), 333–350.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge: Cambridge University Press.

Mehisto, P., Marsh, D., & Frígols, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.

Met, M. (1999, January). *Content-based instruction: Defining terms, making decisions* (NFLC Reports). Washington, DC: The National Foreign Language Center.

METI. (2010). *Release of report by global human resource development committee of the industry-academia partnership for human resource development*. Tokyo: Ministry of Economy, Trade and Industry.

MEXT. (2015). *Seito no eigoryoku koujou suishin pulan* (The plans on the promotion of improvement of students' English abilities). Tokyo, Japan. http://www.mext.go.jp/en/news/topics/detail/__icsFiles/afieldfile/2016/10/19/1378469_001.pdf. Accessed 16 Mar 2019.

Muñoz, C., & Navés, T. (2007). CLIL in Spain. In D. Marsh & D. Wolff (Eds.), *Windows on CLIL. Content and language integrated learning in the European spotlight* (pp. 160–165). European Centre for Modern Languages.

Negishi, M. (2011). The development of the CEFR-J: Where we are, where we are going. In *New prospects for foreign language teaching in higher education* (pp. 105–116). World Language and Society Education Centre.

Nikula, T., Dalton-Puffer, C., Llinares, A., & Lorenzo, F. (2016). More than content and language: The complexity of integration in CLIL and bilingual education. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 1–25). Bristol/Buffalo: Multilingual Matters.

Organisation for Economic Co-operation and Development (2018). *Curriculum flexibility and autonomy in Portugal – an OECD review*. https://www.oecd.org/education/2030/Curriculum-Flexibility-and-Autonomy-in-Portugal-an-OECD-Review.pdf. Accessed 10 Oct 2018.

Peltoniemi, A., Skinnari, K., Mård-Miettinen, K., & Sjöberg, S. (2018). *Monella kielellä Suomen kunnissa 2017. Selvitys muun laajamittaisen ja suppeamman kaksikielisen varhaiskasvatuksen, esiopetuksen ja perusopetuksen tilanteesta* (Many languages in Finnish municipalities 2017. A report of other large-scale and small-scale bilingual ECEC, preschool and basic education). Jyväskylä: University of Jyväskylä. https://jyx.jyu.fi/bitstream/handle/123456789/57577/978-951-39-7391-9.pdf?sequence=5&isAllowed=y. Accessed 16 Mar 2019.

Piaget, J. (1977). *The development of thought: Equilibration of cognitive structures* (A. Rosin, Trans.). New York: The Viking Press.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Norwell: Springer.

Poehner, M. E., & Infante, P. (2015). Mediated development: Inter-psychological activity for L2 education. *Language and Sociocultural Theory, 2*(2), 161–183. https://doi.org/10.1558/lst.v2i2.26982.

Purpura, J. (2004). Learning-oriented assessments of grammatical ability. In *Assessing grammar* (pp. 212–250). Cambridge: Cambridge University Press.

Quartapelle, F. (Ed.). (2012). Assessment and evaluation in CLIL. *Ibis*. http://aeclil.altervista.org. Accessed 28 Mar 2019.

Reilly, T., & Medrano, P. (2009). MEC/British Council Bilingual Project: Twelve years of education and a smooth transition to secondary. In E. Dafouz & M. C. Guerrini (Eds.), *CLIL across educational levels* (pp. 59–70). Madrid: Santillana Education, S.L./Richmond Publishing.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. https://doi.org/10.1007/BF00117714.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (pp. 39–83). Chicago: Rand McNally.

Skinnari, K., & Bovellan, E. (2016). CLIL teachers' beliefs about integration and about their professional roles: Perspectives from a European context. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 145–167). Bristol/Buffalo: Multilingual Matters.

Suwannoppharat, K., & Chinokul, S. (2015). English communication ability development through the CLIL course. *NIDA Journal of Language and Communication, 20*(25), 63–97.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment, 6*(1), 41–63. http://www.altaanz.org/uploads/5/9/0/8/5908292/5.si3tsagarivogt_final_formatted_proofed.pdf. Accessed 27 Apr 2018.

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–274). Berlin/Boston: De Gruyter. https://doi.org/10.1515/9781614513827-018.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL*. Doctoral dissertation, University of Turku. http://urn.fi/URN:ISBN:978-951-29-5754-5. Accessed 12 May 2018.

Wiliam, D. (2017). Assessment and learning: Some reflections. *Assessment in Education: Principles, Policy & Practice, 24*(3), 394–403. https://doi.org/10.1080/0969594X.2017.1318108.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. West Palm Beach: Learning Sciences International.

Wolff, D. (2007). CLIL: Bridging the gap between school and working life. In D. Marsh & D. Wolff (Eds.), *Diverse contexts- converging goals. CLIL in Europe* (pp. 15–25). Frankfurt: Peter Lang.

Chapter 2
Achieving in Content Through Language: Towards a CEFR Descriptor Scale for Academic Language Proficiency

Stuart D. Shaw

2.1 Background

The work described in this chapter is positioned within a chain of research approaches to integration in CLIL. Conceptualising language competence in terms of subject-specific literacy, specifically second language (L2) historical literacy, the chapter focusses on the acquisition of history-specific concepts and competences, and planning assessment for supporting learning and teaching in an L2. The chapter is also firmly located within the 'Higher disciplinary orientation with more visible language pedagogy' quadrant of Leung and Morton's (2016) matrix "in which the notion of developing language competence is closely tied to the expression or construction of meanings associated with specific disciplinary content, skills and thinking processes." (Leung and Morton 2016, p. 238).

This investigation is presented against the backdrop of an international awarding body offering programmes of learning and assessment through the medium of English to schools in a variety of multilingual and educational contexts. A key function of Cambridge Assessment International Education (hereafter, 'Cambridge') is to prepare students whose first language is not necessarily English as candidates for international high-stakes assessments in a variety of bilingual education settings. Schools following Cambridge programmes choose to offer bilingual education approaches in order to develop both the first and second languages systematically. Whilst bilingual education programmes use and adapt numerous different models and approaches, of particular relevance to this investigation is Content and Language Integrated Learning (CLIL) which has been used highly successfully across Europe and beyond (Coyle et al. 2010; Thürmann et al. 2010).

S. D. Shaw (✉)
Cambridge Assessment International Education, Cambridge Assessment, Cambridge, UK
e-mail: stuart.shaw@cambridgeinternational.org

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_2

2.2 Proposing an Academic Language Scale

As reliance on the use of the Common European Framework of Reference for Languages (CEFR) in international school contexts becomes more widespread, there is a greater need for a complementary scale of academic language proficiency. With detailed descriptors of academic language, such a scale might be applicable to academic school contexts and to students in general, whether English is a first or additional language. I believe targeting this wider student group would not be inconsistent with the philosophy of the CEFR, which acknowledges that a learner's cognitive processes and skills develop through engagement with the communicative tasks that arise in social interaction. The development of an academic language proficiency scale would have the ultimate goal of helping school educators plot the progress of their students in the key academic language needed to achieve in content subjects, thereby performing a formative function.[1]

However, a descriptor scale for academic language proficiency is a complex and multidimensional notion, to the extent that a functional description of academic language use inevitably introduces a range of factors. These include: cognitive stage, general language proficiency (given that the language of schooling may not be the learner's first language), the processes and skills involved in mastering the specific curricular objectives of each subject area, as well as the processes and skills involved in learning in general. Neither can it be assumed that these processes and skills are the same across countries or cultures, given possibly different educational traditions and modes of discourse.

In this chapter, I propose an academic language proficiency scale that would draw together aspects of academic language ability found in other scales and, if needed, could add new skills not currently covered by the CEFR. The Cambridge International General Certificate of Secondary Education (IGCSE) History – a general education qualification for 14 to 16 year olds, constitutes the focus of study. The chapter will seek to provide answers to the following questions:

1. What are some of the generic features of an academic language proficiency scale that could be used in the student learning process in a variety of CLIL contexts?
2. How and to what effect can such a scale be used in assessments in CLIL classrooms?

One purpose of developing an academic language scale of the kind proposed here is to use it to a positive effect in specific contexts of learning of interest to Cambridge, where typically students will be non-native speakers of the language of schooling. This is critical: what is proposed is at once a description of language use in school classrooms, and a tool supporting intervention in how language is used in school classrooms. This implies that at its centre must be some theory of how learning happens, and what role language plays in the process.

[1] For an exploration of how learners' achievements can be assessed and reflected in the domains of linguistic performance and subject-content learning, see Massler et al. (2014).

Whilst the demarcations separating the four quadrants of Leung and Morton's (2016) matrix (described in detail in Chap. 1) are sometimes indistinct, the research described here can be located within the first quadrant with its emphasis on higher disciplinary orientation and more visible language pedagogy. In this quadrant, approaches to CLIL conceptualise linguistic competencies as subject-specific literacies, the overall aim of which is to promote learners' academic competence through proficiency in a second or foreign language as part of an instructionally-aligned, progressive and sequentially-phased curriculum. The chapter also builds on, and complements, the curriculum model described in Chap. 4 which explores how a set of four interlocking language teaching methodologies (CEFR, CLIL, Task-based Language Teaching—TBLT, and LOA—learning-oriented assessment) can work together to support student learning in a localised context.

2.3 Mediation in the Context of a Social Constructivist Model of Learning

The model of learning adopted herein is a social constructivist one (refer to Chap. 1 for more detail). Language is the fundamental medium through which learning happens. Learning is a social act, and becoming a better learner is to become a member of a community of practice (Lave and Wenger 1991). Thus, interaction is at the heart of learning.

Mediation, a key feature of interactions and one familiar in a variety of social concepts, is also fundamental to the sociocultural theory of Vygotsky (1981, 1986). In mediation, the learner (and teacher) functions as a social agent creating bridges and conveying meaning either within the same language or across languages (See Chap. 9, this volume). In Vygotskian thinking, mediation involves the use of culturally-derived psychological tools, such as utterances in natural language, in transforming the relations between psychological inputs and outputs. In general terms, mediation in cognition is considered important for cognitive development. Carpendale and Lewis (2004) note that "language mediates children's knowledge of reality" (p. 89), the role of the teacher being one of a mediator for the student's cognitive development.

The CEFR introduces the concept of mediation as follows:

> In both the receptive and productive modes, the written and/or oral activities of mediation make communication possible between persons who are unable, for whatever reason to communicate with each other directly. Translation or interpretation, a paraphrase, summary or record, provides for a third party a (re)formulation of a source text to which this third party does not have direct access. Mediation language activities, (re)processing an existing text, occupy an important place in the normal linguistic functioning of our societies. (CEFR Sect. 2.1.3) (Council of Europe 2018, p. 32)

The CLIL lesson plan suggested in this chapter focusses on mediation activities. The use of academic language descriptors based on CEFR mediation scales affords

a practical means of implementing content-based language learning. A scaffold of pedagogic and linguistic support allows learners to access curriculum content. A primary role of assessment is to support learning (whilst also acknowledging that assessments can serve a number of purposes including accountability and summative purposes). A key issue central to successful CLIL practice is, therefore, the achievement of intended content and language learning outcomes (as outlined in the lesson plan). Specific, Measurable, Achievable, Realistic, and Time-bound learning outcomes (SMART) (Doran 1981) provide a positive reference point for assessment *for* learning.

Within second language learning the centrality of interaction has also been stressed. Two main schools of thought prevail: the *interaction hypothesis* (Gass, Doughty, and Long 2007; Long 1996) and the *comprehensible output hypothesis* (Swain 1985). The *interaction hypothesis* claims that comprehensible input is important for language learning. Moreover, that implicit negative feedback, which can be obtained through negotiated interaction, facilitates second language acquisition. Long (1996) perceives the negotiation of meaning as the means by which learning takes place. The *comprehensible output hypothesis* maintains that learning occurs when learners encounter a 'gap' in their linguistic knowledge of the L2. The learner employs strategies to modify his or her output thereby learning something new about the language (Swain and Lapkin 1995). Swain (1985) argues that production and practice are necessary for the self-monitoring which enables the learner to test and modify hypotheses about the language.

2.3.1 Towards an Action-Oriented Model for Learning

The Common European Framework of Reference for Languages – or CEFR (Council of Europe 2001) rests on a conception of language as action and demonstrates remarkable coherence with a social-constructivist approach to learning. The CEFR offers an action-oriented model for language use and learning (even if not explicitly identified as such) and describes a learner's cognitive apparatus (general knowledge, language competences, and strategies), developing through engagement with communicative tasks that arise in social interaction. The CEFR places "*the co-construction of meaning* (through interaction) at the centre of the learning and teaching process. This has clear implications for the classroom. At times, this interaction will be between teacher and learner(s), but at times, it will take a collaborative nature between learners themselves." (emphasis in original) (Council of Europe 2018, p. 27).

The most recent published CEFR Companion Volume (Council of Europe 2018), intended as a complement to the CEFR, provides illustrative descriptors for new areas not in the original text. Mediation, an idea which is central to Vygotskian philosophy, is one such area and includes reactions to creative text/literature; mediating

communication, texts or concepts. The concept of mediation in the CEFR, which has assumed great importance with the increasing linguistic and cultural diversity of global societies, takes in a range of communicative tasks and strategies relating to collaborative team work, integrated skills, relaying and synthesising text and meanings, and fostering better understanding among others. The CEFR approach to mediation is broader than cross-linguistic mediation but also covers "mediation related to communication and learning as well as social and cultural mediation" (Council of Europe 2018, p. 33). This more expansive approach has particular relevance to CLIL because "mediation is increasingly seen as a part of all learning, but especially of all language learning" (Council of Europe 2018, p. 34).

In terms of the macro-functional basis of CEFR categories for communicative language activities, mediation includes scales for reception, production, and interaction for *mediating communication* – enhancing the effectiveness of cognitive mediation (Creative, interpersonal language use), *mediating a text* – relaying/synthesising information from spoken or written sources (Transactional language use), and *mediating concepts* – collaborating with others in order to come to a decision or solve a problem (Evaluative, problem-solving language use). The lesson plan proposed later in the chapter employs academic language descriptors grounded in CEFR 'mediating a text' scales.

The CEFR's model of language use and learning is a good starting point for defining proficiency in terms of specific assessable constructs: that is, a theory or model of what knowing a language entails. It stresses that whilst the notion of ability to learn is of general application, it is particularly relevant to language learning; also ability to learn mobilises a range of other skills, such that "attitudes and personality factors greatly affect not only the language users'/learners' roles in communicative acts but also their ability to learn" (Council of Europe 2001, p. 106).

2.3.2 *Scales of Proficiency: Including Cognitive Stage and Academic Language for Specific Content Areas*

The CEFR is designed with European adult foreign language learners in mind but was intended to be adaptable to individual contexts. However, whilst the CEFR's focus is on foreign language learning there are two foreign language contexts which are not best accommodated:

- "young learners, because there is no explicit treatment of cognitive stage
- CLIL because language for learning is not clearly distinguished from language for social use." (Jones 2014, p. 2017).

The two foreign language contexts are not unrelated: CLIL, which includes a cognitive dimension not explicitly considered in the CEFR, invariably entails young learners learning content subjects through the medium of a foreign language in a

wide variety of second language learning contexts. To take these two factors into account it is necessary to expand the familiar proficiency dimension by an additional two dimensions relating to age and academic content area. The result is a three-dimensional matrix where each cell distinguishes a learner at a specific proficiency level, at a specific age, studying a specific subject.

The most relevant work in this area appears to be that of the *World Class Instructional Design and Assessment* (WIDA) consortium in the US. The WIDA consortium has developed standards – *English Language Development Standards* (WIDA 2014), primarily aimed at ELLs (English language learners) with the 'No Child Left Behind' goal of ensuring that success in school is not hindered by poor language skills. The WIDA standards are stated to be based on a can-do philosophy. The 2012 amplification of the *English Language Development Standards* adds more detail, for example, by providing links to the US common core syllabus; and assigning a common skill construct (e.g. analysis) to the performance levels of a standard, underlining that such constructs can be made accessible to language learners at differing degrees of language proficiency or cognitive stage.

Other notable US and European initiatives include: the US *Cognitive Academic Language Learning Approach* (CALLA) – an instructional model developed to meet the academic needs of students learning English as a second language in American schools (Chamot and O'Malley 1994); the *National Association for Language Development in the Curriculum* (NALDIC)—the UK's National Association for Language Development in the Classroom, and targets EAL (English as Alternative Language) students (NALDIC 2009); and, the German *FörMig* programme ("Key-stage descriptors for German as a second language")—an instrument concerned with academic language learning by secondary students from migrant backgrounds in mainstream content classrooms (Lengyel 2010).

2.4 Cambridge International General Certificate of Secondary Education (IGCSE)

The Cambridge IGCSE assessment takes place at the end of the course and can include written, oral, coursework, and practical assessment. This broadens opportunities for students to demonstrate their learning, particularly when their L1 is not English. The Cambridge IGCSE syllabuses (subject documents that describe what learners need to know, what they must be able to do, and how they will be assessed) are international in outlook, but retain a local relevance. Syllabuses have been created specifically for an international student body and seek to avoid cultural bias. IGCSE assessment outcomes are benchmarked using eight internationally recognised grades, A* to G (A* being the highest) which have clear guidelines to explain the standard of achievement for each grade.

2.4.1 Cambridge IGCSE History

Cambridge IGCSE History, the focus of this research, explores some of the major international issues of the nineteenth and twentieth centuries, as well as covering the history of particular contexts in more depth. The emphasis is on both historical knowledge and on the skills required for historical research. It encourages learners to raise questions and to develop and deploy historical skills, knowledge, and understanding in order to provide historical explanations. Two of the syllabus aims additionally encourage the development of arguments and communication skills.

All candidates take three components. Papers 1 and 2 are compulsory and candidates must additionally take either Component 3 (Coursework) or Paper 4 (a written Alternative to Coursework paper). The questions in Papers 1 and 2 are differentiated by outcome. Student responses are marked using a levels-based mark scheme. The mark scheme (which provides the marker with all necessary marking criteria) is designed to represent the underlying constructs to be assessed. The questions in Paper 1, for example, are in the form of structured essays, split into three parts (a), (b), and (c) with maximum marks of 5, 7, and 8 respectively. The levels of performance in the mark scheme relate to a progression of skills (summarised for Paper 1 in Table 2.1).

Table 2.1 Paper 1 skills assessed using the levels-based mark scheme

Question part	Level and marks available	Skills rewarded at each level
Part a	Level 1 (0 marks)	
	Level 2 (1 mark)	Answer lacking specific contextual knowledge
	Level 3 (2–5 marks)	Description
Part b	Level 1 (0 marks)	
	Level 2 (1 mark)	Answer lacking specific contextual knowledge
	Level 3 (2–3 marks)	Description/identification
	Level 4 (4–7 marks)	Explanation
Part c	Level 1 (0 marks)	
	Level 2 (1 mark)	Answer lacking specific contextual knowledge
	Level 3 (2 mark)	Description/identification
	Level 4 (3–5 marks)	Explanation of one side of the argument
	Level 5 (5–7 marks)	Explanation of both sides of the argument
	Level 6 (8 marks)	Evaluation

2.5 History-Related Language Skills

An exploratory investigation of the CEFR scales would suggest:

- *Pragmatic Competences*, *Language Activities/Strategies* and *Processing Text* are most relevant to Common Underlying Proficiency (Cummins 1980) and Cognitive Academic Language Ability (CALP) (from B1)
- *Sociolinguistic Competence* is least relevant (from upper B2)
- *Linguistic Competence* is partly relevant (from B2).

Any language described below these levels tends (with the understanding that there is no 'normal' L1 learner) to be simply part of the foreign/L2 learning process. These judgments are crude based on surface level examination of the language used for the purpose of considering the relevance of the CEFR to CALP and to the approach to this investigation.

It is probably easier for History teachers to identify and teach 'content-obligatory' vocabulary (language the student "must know to master content", Mehisto et al. 2008, p. 101) than cognitive discourse functions (Dalton-Puffer 2016, pp. 29–54). However, it is probably less easy for teachers to teach 'content- compatible language' ("helpful but not absolutely necessary language", Mehisto et al. 2008, p.101) which aids communication and is probably more the domain of the language teacher. The heart of the problem is identified by Srole (1997): "This fusion of language and content requires students to understand non-history content and cultural references that linguistically and culturally diverse students do not yet possess." (p. 104). Srole (1997) advocates two approaches: building vocabulary and disclosing rhetorical structures. Building vocabulary involves "defining (supplying specific meanings) and embedding (situating the contexts)" (Srole 1997, p. 105). Disclosing rhetorical structures "exposes academic frameworks which help students master course content independent of the vocabulary." (p. 105). Srole also supports the explicit teaching of 'academic rhetoric' in order to organise, control and communicate ideas better, to understand how language cues guide a reader through an essay and "help students recognise conceptual themes and organise them appropriately" (1997, p. 116). This corresponds to the CEFR 'Pragmatic Competences'.

Beacco (2007) provides a prototype for a descriptive framework for communicative/linguistic competences involved in the teaching and learning of History. He lists and describes the educational values targeted by history teaching: the social situations of communication involving history; the expected historical knowledge; and the existing in-school communication situations for transmission of history. Referencing the CEFR, Beacco (2010) comments:

> It is plain that the specifications of the CEFR relate more to reading as comprehension than as interpretation or critical response. For languages of instruction, the comprehension strategies need to be re-interpreted as a function of the knowledge in the discipline (in this case, critical comprehension). (p. 10)

Of particular relevance here are subject competences and linguistic-semiotic skills. Beacco (2007) contends that a common inventory (for different languages and different subjects) could be made which would enable linguistic competences to be specified and actually defined i.e. a set of language ability descriptors required for learning and teaching history. Beacco (2010) explores links between historical knowledge and language descriptors in in-school discourse types in order to produce a construct called 'historical communication'. Beacco (2007) distinguishes between 'competences concerning knowledge of the subject' and 'epistemological competence' (p. 18). The latter he defines as 'expertise' in History ('historical literacy' or 'historiography'). He discusses different inventories to capture this expertise such as, 'cognitive skills' (e.g. arrive at balanced, responsible conclusions), 'historical concepts' (e.g. recognising similarities and differences), and 'core skills' (e.g. ability to examine potential sources of information and distinguish between primary and secondary sources).

Beacco (2007) argues that the linguistic-cognitive resources needed for subject competences could be based on the CEFR. His approach is to formulate sets of language competence descriptors that integrate CEFR can-do statements. One such example would be: 'place the occurrence under discussion in a broader context (chronological, cultural)' and 'distinguish objectified discourse from judgement' (Beacco 2010, p. 10).

Beacco's (2007) *strategic competence* (p. 17), for example, refers to the 'skills needed to manage texts' and are 'independent of the languages and discourses used and can therefore be used in a reference document on languages of schooling' (p. 21) and corresponds with the CEFR's category of 'communicative language activities and strategies'. Beacco's *discourse competence* corresponds to the CEFR's category of 'pragmatic competence'. Finally, Beacco's *formal competence* involves the resulting formal conventions of discourse types 'avoiding using complex terminology' and using categories that are more 'integrated and transversal', the 'discursive representation of the cognitive processes' (p. 24).

2.5.1 Linking IGCSE History English Proficiency Levels to the CEFR

Shaw and Imam (2013) investigated aspects of the CEFR relevant to academic language proficiency in IGCSE History. In their study, syllabuses, question papers, mark schemes and candidate performances were analysed. Findings suggest that history necessarily requires academic language (CALP) and key academic language skills were identified. Table 2.2 provides an overview analysis of the main findings.

Based on these findings, there is a case to suggest that CEFR level B2 could represent a critical CALP level for this age group (see also Imam 2010). However, as the lexical input of accompanying authentic stimulus source material is much

Table 2.2 Analysis of history language (Shaw and Imam 2013)

Language	Expected/Demonstrated
LEXICAL RESOURCES (content vocabulary)	Grade descriptions: • 'appropriate historical terminology' for higher grades (Grade C and above) • 'simple historical terminology' for lower grades (Grade F). 'Appropriate historical terminology' (*content-obligatory* vocabulary): phrases relating to specific periods/events, topics, and concepts in the curriculum – mainly nouns and proper nouns, e.g. *free states abolitionism, nationalism, the Great Powers, the League of Nations.* 'Simple historical terminology' (*content-compatible* vocabulary): general useful historical phrases not tied to a specific part of the curriculum, e.g. *collapse, defeat, democratic* and phrasal verbs such as *set up, step in*.
FUNCTIONAL RESOURCES (task: language demand)	Mainly open questions with constructed responses: describing, explaining, evaluating. In-depth source evaluation. e.g. • *How far…* • *How successful…*
STRUCTURAL RESOURCES	modal verbs and conditionals (speculation and deduction) passive forms (impersonal tone) language of comparison for sources cohesive devices (evaluation and conclusions) reason and result clauses language of exemplification.
CRITERIAL TASK FEATURES OF INPUT (cognitive reading demand)	High reading load – authentic sources, textual and visual (e.g. cartoons), requiring cultural familiarity. Reading skills: • Scanning for exemplification and reading for detail • Skimming to decide optional questions • Using referencing skills (exophoric referencing to link reading of a source text to other source texts, own contextual knowledge, and wider historical context) • Reading for explicit and implicit meaning to interpret historical concepts / source reliability • Comparing and contrasting sources.

(continued)

Table 2.2 (continued)

CRITERIAL TASK FEATURES OF OUTPUT (writing)	• short answers at word / phrase / sentence level; • extended reasoning, with the ability to organise ideas clearly and to present effective, balanced arguments with exemplification.
CEFR LANGUAGE LEVEL (reading)	B2 (sources C1*)
CEFR LANGUAGE LEVEL (writing)	B2 (C1* useful for higher marks requiring coherent, reasoned explanations)
OTHER FINDINGS	* Although CEFR level C1 level is useful, candidates with B2 level, or grade C in IGCSE English as a Second Language, can achieve high IGCSE History grades if they: • Demonstrate a high level of CALP (e.g. reading strategies, essay cohesion) • Use 'achievement strategies' (e.g. paraphrase to cover gaps in vocabulary) • Do not use 'avoidance strategies' (e.g. reduce the cognitive demand of the task). (Imam, 2010)

higher, students would need to be at least CEFR C1 level to be able to process the text fully. CEFR C1 level in certain scales influences higher grades for a subject such as History. Such influential scales could be those that are not focused on the form and accuracy of language but are more cognitive-academic in nature (e.g. 'Pragmatic' scale, which includes thematic development, propositional coherence, coherence and cohesion). However, IGCSE History is not an English comprehension test and candidates do not need to understand all the text to perform well.

What follows is an attempt to advance this work and help develop a supplementary scale focusing on academic language proficiency, with descriptors for each CEFR level.

2.6 A CEFR History Descriptor Scale for Academic Language Proficiency

The development of a supplementary scale focusing on academic language proficiency, with detailed descriptors of academic language for each CEFR level, could constitute the basis of CLIL lesson planning by facilitating, for example, the creation of conditions for communication and cooperation through mediation activities.

The CEFR provides a wealth of specialised scales—for example, the text processing scale, the pragmatic scales, the strategic scale of compensating, as well as tables that combine scales—for example, the CEFR's Table A5, which draws together 'relevant qualitative factors for production' (Council of Europe 2009, p. 149). However, aspects of academic language are found across various scales, which can make it hard to locate and apply them to school contexts. Furthermore, it is often assumed that academic language proficiency features in the upper parts (higher proficiency) of the CEFR scale, where there are high expectations for foreign language learners that would not always be met even by learners for whom English is an L1. Conversely, the lower end of the CEFR scales might not always capture the academic language that learners may be developing in their early stages. The proposed new scale would draw together aspects of academic language ability found in other scales and, if needed, could add new skills not currently covered by the CEFR.

Research into Cambridge IGCSE History, involving reading (sources) and writing, led to the following beginnings of a scale representing CEFR levels C2 (Table 2.3) and B2 (Table 2.4).[2]

By way of illustration, the first column in Table 2.3 (*CEFR history level*) shows the CEFR and history mark scheme levels. The second column (*Quality*) relates to a construct of history instantiated in the mark scheme. The third column (*Descriptor*) depicts history descriptors that appear to relate to a CEFR description. The final column (*CEFR scales*) attempts to align the information in column 3 to the most relevant CEFR scale.

The skill 'Evaluate'—shown in Column 2 of Table 2.3, is partly defined in the syllabus under the Grade A grade description as 'Compare and contrast a range of sources and draw clear, logical conclusion'. On the question paper it manifests itself in the form of asking candidates to 'use the sources to explain your answer' or 'Explain your answer using the details of the sources and your knowledge'. 'Evaluation' can be understood as entailing five steps:

[2] Similar analyses for each of the other CEFR levels are available on request.

1. Before being able to compare and contrast a source, the student first has to be able to *comprehend* the source, in terms of its explicit meaning, to make surface-level comparisons.
2. Then, in order to compare and contrast, at a deeper level (involving historical concepts e.g. motives, beliefs, cause, consequence, continuity, change, reliability), the student needs to be able to *interpret* the source, to infer conceptual comparisons.
3. Next, in order to evaluate the sources, the student needs to be able to *compare and contrast* the sources with each other, or with a given proposition, to see 'how far they agree'.
4. By assessing the extent of, and weighing up, the evidence in step 3, the student can give an opinion and reach a conclusion about 'how far the sources agree', by assessing the extent of, and weighing up, the evidence in step 3.
5. Finally, all the above needs to be *communicated* clearly and logically, exemplifying and justifying.

The five steps could be said to be the linguistic competences required for the historical cognitive process, or subject competence, *evaluate*.

'Evaluate' relates to the *Coherent and cohesive* descriptor (the first descriptor in the third column) and aligns with the 'Pragmatic' CEFR scale which addresses the ways in which context contributes to meaning:

Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices.

The next descriptor in Column 3—Reconstructs arguments from different sources—aligns with the 'Text Processing' CEFR scale:

Can summarise information from different sources, reconstructing arguments and accounts in a coherent presentation of the overall result.

The word 'evaluate' in the History mark scheme attracts the highest number of marks (level 6) whereas the word 'evaluate' appears only in the upper B2 level descriptors of the CEFR. This means one cannot rely simply on a surface comparison of words in the two scales as they might mean different things when unpacked. Clearly, given the non-linear correspondence between the history scale and the CEFR scale, it is not a simple task to compare descriptors between scales.

Table 2.3 Tentative academic language proficiency scale (CEFR Level C2)

CEFR history level	**Quality**	**Descriptor**	**CEFR scales**
CEFR: C2 history: bonus marks	'Evaluate & create'	*CEFR* Coherent and cohesive. Reconstructs arguments from different sources. Clear, complex, logical. Smooth substitution for specialist words. *IGCSE history mark scheme:* Bonus marks: evaluation of sources	*Pragmatic* Can create coherent and cohesive discourse making full and appropriate use of a variety of organisational patterns and a wide range of connectors and other cohesive devices *Text Processing* Can summarise information from different sources, reconstructing arguments and accounts in a coherent presentation of the overall result *Overall Written Production* Can write clear, smoothly flowing, complex texts in an appropriate and effective style and a logical structure which helps the reader to find significant points *Reading for Inform ation & Argument* No descriptor available *Strategic* Can substitute an equivalent term for a word he/she can't recall so smoothly that it is scarcely noticeable *Socio -linguistic* Appreciates fully the socio-linguistic and sociocultural implications of language used by native speakers and can react accordingly

Table 2.4 Tentative academic language proficiency scale (CEFR Level B2)

CEFR history level	Quality	Descriptor	CEFR scales
CEFR: B2 history: levels 3 & 4	'Detail & analyse'	*CEFR* Clear, limited cohesion Relevant expanding and supporting detail Comments on contrasting points of view and main themes Synthesise and evaluates number of sources Paraphrases Appropriate *IGCSE history mark scheme:* Level 3: surface match or mismatch: Explicit understanding of one side of an argument ↓ Level 4: surface match and mismatch: Explicit understanding of both sides of an argument	*Pragmatic* Can develop a clear description or narrative, expanding and supporting his/her main points with relevant supporting detail and examples Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some "jumpiness" in a long contribution. *Text Processing* Can summarise a wide range of factual and imaginative texts, commenting on and discussing contrasting points of view and the main themes Can summarise extracts from news items, interviews or documentaries containing opinions, argument and discussion Can summarise the plot and sequence of events in a film or play *Reading for Information & Argument* Can obtain information, ideas and opinions from highly specialised sources within his/her field... *Overall Written Production* Can write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources *Strategic* Can use circumlocution and paraphrase to cover gaps in vocabulary and structure Can make a note of favourite mistakes and consciously monitor speech for them *Socio-linguistic* Can express him or herself appropriately in situations and avoid crass errors of formulation *General linguistic range* Upper B2: broad range of language to express clearly Lower B2: express viewpoints and develop arguments

The analyses in Tables 2.3 and 2.4 (together with similar analyses for the other CEFR levels) illustrate that a key CEFR level for IGCSE History could be B2, which moves language beyond the descriptive realm (B1) into the analytic realm. A crucial jump in the history mark scheme is from explicit understanding to implicit understanding of texts. Understanding implied opinions appear in the CEFR from level C1 (and C2). However, there may not be a consistent linear correspondence between language and history at the individual level:

- a student who is at CEFR B1 level (or lower), using simple, descriptive language, would not have the language to be able to access, analyse, and evaluate source material
- a student may have the sophistication of language at CEFR C2 level but lack sufficient cognitive ability, historical knowledge, or examination technique to evaluate history source
- a student with less sophisticated language at CEFR C1 or B2 level still may be able to grasp the content and effectively communicate their evaluation to examiners.

This can also be seen in Table 2.5, which suggests that the IGCSE history grade descriptor references to language could correspond to more than one CEFR language level. The structure for the table is identical: the first column (Grade Descriptor) shows the assessment grade. Accompanying the grade is a syllabus description of expected performance. For brevity, only IGCSE history grades A and C are shown. In addition to the assessment objectives, the grade descriptors also make reference to coherent, logical arguments, to clear and coherent communication using appropriate historical terminology and to interpretation and evaluation of historical sources. The second column in each of the three tables relates to the CEFR scale and level.

Table 2.5 shows the Grade A and Grade C grade descriptors. Grade A relates to the descriptor "Recall, select and deploy relevant historical knowledge accurately to support a coherent and logical argument". This history grade descriptor aligns with the upper CEFR level B2 for the General Linguistic Range descriptor:

Has sufficient range of language to be able to…develop arguments…using some complex sentence forms.

However, elements of CEFR levels C1 and C2 for the same CEFR scale as well as elements of both upper and lower B2 and instances of C1 and C2 for the Writing Reports and Essays descriptor can also be observed.

In the same table, it can be seen that the history grade C descriptor relates to the middle of CEFR level B2 on the General Linguistic Range descriptor:

Expresses viewpoints and develops arguments.

and
to the lower CEFR level B2 on the Writing Reports and Essays descriptor:

Develops an argument, giving reasons in support of or against a particular point of view.

This variation was borne out in the examination performance data of individual candidates who had achieved higher or lower grades in IGCSE English as a Second Language compared to their grades in IGCSE history.

Table 2.5 Comparison of IGCSE history grade descriptors (*coherent, logical arguments*) with CEFR level descriptors

Grade Descriptor	CEFR Scale and Level
Grade A: Recall, select and deploy relevant historical knowledge accurately to support a **coherent and logical argument**	*General Linguistic Range* Upper B2: can express him/herself clearly and without much sign of having to restrict what he/she wants to say. Has sufficient range of language to be able to... **develop arguments**...using some complex sentence forms. C1: ...broad range of language to express him/herself clearly, without having to restrict what he/she wants to say. C2: can exploit a comprehensive and reliable mastery of a very wide range of language to formulate thoughts precisely, give emphasis, differentiate and eliminate ambiguity. *Writing Report and Essays* Lower B2: develops **an argument**, giving reasons in support of or against a particular point of view … can synthesise information and arguments from a number of sources. Upper B2: develops an argument systematically with appropriate highlighting of significant point and relevant supporting detail. Can evaluate different ideas or solutions to a problem. C1: clear, well-structured expositions of complex subjects, underlining the relevant salient issues. Can expand and support points of view at some length with subsidiary points, reasons and relevant examples. C2: smoothly flowing, complex … essays which present a case, or give critical appreciation of proposal or literary works … appropriate and effective logical structure which helps the reader to find significant points.
Grade C: Recall, select and deploy relevant historical knowledge **in support of a logical argument**	*General Linguistic range* Mid-B2: expresses viewpoints and **develops arguments** … using some complex sentence forms to do so. *Writing Report and Essays* Lower B2: **develops an argument** … giving reasons in support of or against a particular point of view … can synthesise information and arguments from a number of sources.

2.7 Using Academic Language Proficiency Descriptors in the CLIL Classroom

The construction of an academic language scale whose model of reference is the CEFR has clear implications for CLIL pedagogy: it is believed that CLIL learning can be improved through a CEFR task-based oriented approach (See Chap. 4, this volume). Meaningful content resides at the very heart of CLIL, and it is crucial that content teachers identify and teach core content concepts in a rigorous manner. Language constitutes an opening to the knowledge and skills of any given content subject. CLIL is one didactic way to support learners who are learning through an L2 and who need to learn the academic language in order to achieve in content subjects. Given the somewhat ambitious linguistic and cognitive objectives of CLIL classes, an academic language scale would support CLIL teachers and schools in implementing bilingual education by making content teachers teaching through the L2 aware of students' L2 language knowledge and needs. CLIL's multi-faceted approach serves to motivate students through a range of diversified teaching methods. An academic language scale could be used to make a foreign language programme more motivating by teaching real content through the target language.

Whilst the CEFR descriptor scales may allude to social and academic contexts of use and what a language user can do at various levels of competence, they do not elevate the BICS/CALP distinction to a level of prominence. A disjunction exists across the six levels of the CEFR descriptors: the first three levels include the 'Basic User' scales and the lower end of the 'Independent User' scale (A1, A2; and B1) which tend to relate to familiar, conversational language without reference to complex subject matter. From B2 (upper end of the 'Independent User' scale) language becomes more specialised (e.g. the reading and comprehension of academic texts). Only at higher levels (C1 and C2—'Proficient User' scales) is the learner expected to understand specialised articles and present and write detailed descriptions of complex subjects. The management of topic range and context within the CEFR assumes a progression from the personal, through the familiar and routine, to the transactional, moving towards increasingly complex and abstract topics at the highest levels. Though such an approach coheres with that which characterises sequencing in mainstream language teaching, it may not reflect the approach charted in a CLIL context or in language courses for professional adults. In general, students are expected to reach a B2 level at the end of their secondary schooling. However, CLIL classes may begin earlier when the language proficiency of many students is at A2 or B1. Thus, awareness of the BICS/CALP distinction can helpfully assist in the characterisation of the different profiles of competence demonstrated by adults and younger students.

Given that student communication through the L2 may be limited to the CLIL classroom, content teaching through the L2 needs to be as effective as possible. The use of academic language descriptors can both build on, and encourage, students' linguistic repertoires (academic proficiency in L1 and L2) and knowledge

(epistemological) repertoires in both spoken and written contexts. Such a scale, when used purposefully, can enrich bilingual programmes including CLIL.[3]

An illustration of how an academic language scale may be employed in the CLIL classroom is in the application and use of *Learning Outcomes*. Both the content subject and the language used as the medium of instruction are similarly involved in defining the learning outcomes. Achievement of intended content and language outcomes is a key point central to successful CLIL practice (Mehisto and Ting 2017, p. 214).

Learning outcomes define what a learner 'can do' (what they know and be able to do) by the end of a course of study (e.g. a lesson or groups of lessons). Pedagogic exploitation of academic, communicative can-do statements has the potential to inform planning and delivery of lessons, negotiation of syllabus content with learners and, more generally, build an effective learning environment. The clarity of content and academic learning outcomes can be enhanced with references to academic CEFR descriptors. Academic can-do descriptors, if clear and specific, not only guide students more effectively in their learning but also provide measurable outputs for teachers (thereby performing an *assessment for learning* function). Students would need to be presented with exemplars of the types of language-use in order to achieve such outcomes. (For instance, exemplars could be based, in part, on authentic student responses.) Teachers will also need to provide students with evaluative criteria and appropriate scaffolding. Scaffolding, which parallels Vygotsky's (1986) concept of zone of proximal development, affords a very practical means of implementing authentic language and content-area learning, and provides a framework of pedagogic and linguistic support which permits the learner access to the curriculum content through the teacher provision of templates, guides, frames and verbal supports. The efficacy of scaffolding is contingent upon learners building on what they already know in order to make sense of, and take on board, new knowledge. Thus, the scaffold acts as a bridge between prior knowledge and new material. Effective lesson planning entailing the use of academic language proficiency descriptors (such as the one described in this chapter) enables: the teacher to set clear targets for content-area learning; explicit teaching of the language needed to participate in content-area learning; acknowledgement of the needs of CLIL learners; learner participation in classroom activity based on an understanding of their language development; the use of cognitively challenging tasks that require learners to engage with cognitive academic language; and, the provision of models of authentic language in use and opportunities to practice it.

When introducing students to specific language skills associated with the topic of the learning outcome, the teacher may draw upon CALP 'can-do' statements as a means for signposting when students need to learn how to use language in a new way; using authentic material to illustrate the kind of language that is used for an

[3] The CLIL content will follow the curriculum of the subject matter but will not constitute the core teaching of that subject.

activity; and leading learners step by step through the different stages of practising the skills.

The teacher must also consider the abilities of students within a CLIL class. CLIL classrooms tend to consist of mixed ability language students. Instructional *differentiation* provides an effective framework for classrooms that includes diverse students. As a discipline, history is ultimately expressed through advanced language structures and cognitive discourse functions which necessarily demands cognitive maturity and language competence amongst learners. Effective differentiation (by CEFR level, and mark scheme level, say) is dependent upon an awareness of each learner's current level of understanding and achievement and their individual learning needs. Armed with this knowledge, teachers would be able to provide appropriately differentiated learning tasks and activities to mediate learners based on a range of CEFR level cognitive descriptors. It must be stressed, however, that implementing the CEFR in the CLIL classroom is no easy accomplishment (for reasons given in the introductory chapter). Notwithstanding the challenges, the value of the CEFR in assessment *for* learning is not in determining learning outcomes but in how knowing what learners can do can direct further teaching and learning (see Chap. 1, this volume). Ultimately, assessment *for* learning and feedback improve teaching and learning practices which is why CLIL needs to assume a formative orientation.

2.8 A Worked Example Using the Academic Language Proficiency Scale

By way of illustration, consider the proposed lesson plan reproduced as Table 2.6. The lesson adopts a task-based approach which, as Kunschak (Chap. 4, this volume) points out, "lend themselves to interactive classrooms, the development of learner autonomy and collaborative work" (p. 96). As mediation is increasingly seen as a part of "all language learning" (Council of Europe 2018, p. 34), the lesson focusses on CEFR descriptors relating to 'mediating a text' (See Sect. 2.3). Given that the proposed lesson entails small group, collaborative tasks, interactions between learners that occur will have mediating functions (e.g. organising collective work, and facilitating access to, and the construction of, knowledge).

This kind of activity affords opportunities for content and language teachers to both plan and implement their own assessment instruments in an attempt to monitor and evaluate the progress of learners in their classrooms – thereby enhancing teacher professional judgement and agency (Davison and Leung 2009). Additionally, it gives learners the opportunity to develop their own ideas and understanding with direct input from the CLIL teacher. It also allows learners to employ mediation strategies to explain new concepts (by breaking down complicated information into constituent parts, and demonstrating how these parts fit together to give the whole picture) as well as strategies to simplify a text (through streamlining a written text to its essential messages).

Table 2.6 History lesson plan using academic language proficiency descriptors (based on Ellis 2003, p. 217)

Focus Points

- Why was there opposition to Soviet control in Hungary in 1956 and Czechoslovakia in 1968?
- How did the USSR react to this opposition?

Learning Outcome(s):

- Learners understand the issues underpinning opposition to Soviet control
- Learners are aware of why and how the USSR reacted the way they did

Cognitive Academic Language Proficiency Descriptors:

- Can understand in detail a wide range of lengthy, complex authentic historic texts.
- Can summarise in writing and speech long and complex historical source texts, respecting the style and register of the original, interpreting the content appropriately through the meanings of content-compatible language.
- Can use high -level phrases, idiomatic and colloquial language in response to historic stimulus material
- Can use appropriate content-obligatory terminology which could include phrases relating to specific historic periods/events, topics, and concepts in the curriculum (mainly nouns and proper nouns).
- Can facilitate understanding of a complex historical issue by highlighting and categorising the main points, presenting them in a logically connected pattern and reinforcing the message by repeating the key aspects in different ways.
- Can recogn ise a complex historical source text in order to focus on the points of most historic relevance to target audience.

Final Task	**Type of Input (Scaffolding)**	**Instructional Differentiation**	**Processes**	**Micro-Tasks (focus on one aspect of language)**	**Assessment**
Present a version of historical facts	L1 and L2 Textbooks, authentic documents	CEFR Level C1: *Reading for information and Argument* History Mark Scheme Levels 4 & 5 C1: Mediation – Conveying clearly and fluently in well structured language the significant ideas in long, complex historical texts.	Individual work Group work Oral and written production	*Vocabulary*: according to topic *Lexico-grammar*: structures that present an interaction of time and causes and the expression of temporal markers.	Students providing sticky notes with reasons to a whole class diagram for discussion. A small group exercise involving ICT and asking the groups to produce a short script for a radio news bulletin to be broadcast to the West immediately after the Soviet response. Where possible details could be based on authentic material from the time. Following presentations the different approaches could be discussed.

(continued)

Table 2.6 (continued)

Example use of cognitive academic language descriptors based on the CEFR 'Mediating a text' descriptors for *Relaying specific information in speech and writing* (from long, complex historical text) and *Processing text in speech and writing* (from long, complex historical text) (See Cognitive Academic Language Proficiency Descriptors shown above):

- *Understanding content-compatible language from co-text*
  - "From very early in 1968, other Communist leaders in Eastern Europe were alarmed by developments in Czechoslovakia. It was clear to them that the growing freedom could be highly infectious."
- *Identifying non-essential language to know in order to understand the text*
  - "Indeed, it was not long before demonstrating Polish students shouted, 'We want a Polish Dubcek!' The first sustained pressure put on the Czechoslovak leadership came at a meeting with five member states of the Warsaw Pact in March 1968 ."
- *Identifying language that needs to be translated*
  - "The meeting in early August between the Czechoslovak leaders and the Soviet and East European leaders produced a compromise document. At the very time when this agreement was being reached, the Soviet leadership were sent a letter they had been asking for to justify an invasion."
- *Identifying essential to know yet difficult to translate language*
  - "It was a request from the hard-line members of the Czechoslovak leadership calling for intervention. The final decision to launch an invasion was taken between 15 and 17 August."

Through lesson tasks (such as the kind described here) learners are able to engage in ways which require complex language derived from curricular complex relations. Language is the mediating tool through which content and language are co-constructed in the learning environment: language is used to mediate content knowledge and content is used to mediate language. Mediation takes place through teacher and learner talk in interaction. Academic content and language descriptors scales offer a direct means for assessing learning—how much (or how little) takes place (Chap. 3).

Learner mediation here helps to develop historical concepts and ideas by talking ideas through and articulating thoughts, thereby facilitating understanding and communication. As such, the use of CEFR cognitive mediation scales is specifically relevant for the CLIL context where small group, collaborative tasks constitute the focus of lesson activities. Such tasks afford class participants opportunities to share

disparate input whilst allowing learners to exchange information and work collaboratively to accomplish a common objective.

The CLIL activity described here provides opportunities for learners to experiment with the disciplinary language of the given subject and in so doing promote 'a level of talking and interaction that is different from that of the traditional language classroom' (Coyle 2006, p. 11). The activity also affords learners the chance to use language for various purposes in order to meet curriculum expectations (such as engaging with historical academic texts, writing, and presenting historical events); to employ a balanced and proficient use of the four macro skills of language; and, to use authentic literary-specific materials (such as the use of authentic source materials from the media).

A scale of academic language proficiency framed in terms of CEFR 'can-do' statements can significantly improve the depth and scope of content learning for the student. Such a scale would provide the support needed by content teachers in navigating and teaching the L2 academic language of the given subject. Core content topics and concepts can be taught in manageable bites of content and language thereby affording the learners freedom to use their cognitive resources to learn new content and language.[4]

The use of such a learning tool has additional implications for assessment. Assessment serves as a tool for the learning of content and language. A fundamental role of assessment is to support student learning. There exists a direct relationship between a learners' exposure to the curriculum, how the curriculum is taught, and their performance on assessments (Schmidt et al. 2001). It is generally understood that an 'aligned' instructional system is one of the characteristics of a good educational programme (Pellegrino et al. 2001; Achtenhagen 2012). Alignment occurs when curriculum, pedagogy and assessment are mutually supportive, that is, they identify and support the learning of agreed knowledge, skills, and competencies.

Assessment *for* learning (including feedback) is a core driver of learning. Assessment outcomes can be used to identify teaching and learning needs and subsequent actions (Black et al. 2002). A distinction needs to be made, however, between the assessment of content learning and the learning of academic language. Mehisto and Ting (2017) contend that assessment literacy is inextricably linked to making explicit intended learning outcomes for both content and language, and for students to work with and practice assessing exemplars of poor, satisfactory and excellent work.

The primary aim of assessment *for* learning in CLIL "is to support the learning of content and language, as well as to foster critical thinking about both, and *ultimately to improve teaching and learning practices*" (Mehisto and Ting 2017,

[4] Lorenzo and Dalton-Puffer (2016) discuss 'L2 historical literacy' in the context of CLIL. They outline an integrated model of historical competence and language competence. The model presents the representation of history in language units of different levels with genres including predictable discourse functions and lexico-grammar. The authors explore the learning conditions for historical literacy in real CLIL classrooms with examples drawn from the ConCLIL project database (with history lessons from Austria, Finland, and Spain).

p. 213; emphasis in the original). The lesson plan described here provides opportunities for learners in the CLIL classroom to reflect on their own work as well as on the work of others. For reflection to be efficacious, however, learners need to be aware of the learning outcomes (in terms of content and language) as well as having an appreciation of how their learning will be measured. At the same time, teachers must have clear evidence of learning in the CLIL classroom in order to make informed decisions about future teaching and learning of both content and language, whilst providing learners with appropriate feedback on how to progress in terms of achieving the intended learning outcomes.

2.9 Reflections on the Challenges of Constructing a Scale for Academic Language Proficiency Based on the CEFR

The construction of a scale for academic language proficiency based on the CEFR is a complex endeavour. Clearly, the work described here is in its infancy and needs significant further development. Attempts at developing other proficiency scales besides the CEFR, such as the US WIDA's *English Language Development Standards* and the FörMig key-stage descriptors for German as a second language, are already advancing. Additionally, the Council of Europe *Platform* and the *European Centre for Modern Languages* are engaged in a related development that is much broader in scope, possibly involving plurilingual and intercultural competences. Plurilingualism is at the heart of the Council of Europe's conception of how language policy should be developed within Europe, and the same concerns are relevant to much of the work of Cambridge.

However, the mere identification of new bodies of descriptive material, more or less similar to the approach taken by the CEFR, does not get fully to the heart of the issue, which is, by describing how language operates as the object and medium of learning in the classroom, to find ways of directing its use more effectively. This is what the Council of Europe *Platform* has attempted to address albeit with varying degrees of success.

At any rate, the multidimensional nature of the subject is clear, and it would take a drastic degree of abstraction to entirely reduce it to a single dimension describing something called 'academic language'. By accepting this, the interesting challenge becomes to identify the minimal set of constructs and parameters that would address the complexity of the task. If successful what would emerge would be a more complex, composite picture of an individual's language profile in relation to dealing with academic subject matter.

The precise purpose of an academic language proficiency scale will need to be determined. For example, would it be:

- a series of descriptors of language use in the content classroom (a history lesson, perhaps) and assessments?

- target specific programmes, subjects, and students' ages (such as IGCSE history typically at the age of 16)?
- more generic?

It also needs to be determined whether it would be a tool supporting intervention in language use in CLIL classrooms. (Interestingly, CLIL receives some attention in the Council of Europe *Platform* documents, Thürmann et al. [2010]). This implies that a solid understanding of how learning happens, and what role language plays in the process must be at its centre.

Additionally, there is great potential for tools such as Beacco's *Descriptive Framework* and the CEFR scales to be utilised to identify the linguistic features that influence performance in History (for example). Longer term Beacco's *Descriptive Framework* (2007) could be extended to other cognitive discourse functions, besides 'evaluation', within History and across other subjects. The results could feed into a high-level inventory of the language of schooling, which could be published and made available to CLIL teachers.

Quite apart from in-depth learning of the curricula subject and development of language competencies, CLIL provides abundant opportunities for the development of transversal or 'soft' skills. Some of the skills associated with History underpin the manifestation of competencies found in other subject areas. The approach to constructing history CALP descriptors can be replicated elsewhere for other subjects representing the sciences, social sciences, and humanities. Such an approach enables consideration of other subject-specific literacies and the specification of what is to be assimilated language-wise and content-wise at specific phases in the development of the integrated content/language competences.

Used in the ways suggested, an academic language proficiency scale might potentially inform the construction of syllabus level 'language awareness' guides. Such guides could include:

- subject-specific terminology
- language skills derived from the Assessment Objectives and marking criteria
- definitions of command words,[5] and
- scaffolding to help plan answers to questions (e.g. student organisers).

It is believed that the approach described here will provide insights which will prioritise professional development of CLIL skills and heighten language awareness in the assessment process. The challenge of constructing a scale for academic language proficiency—no matter how great, should not impede its continuing consideration and development. The work continues.

Acknowledgements I am deeply indebted to the work and inspiration of Helen Imam without whom this research would not have been remotely possible.

[5] Command words shape the scope, nature, and depth of treatment expected in a candidate response.

Further Reading

Jones, N. (2014). *Multilingual Frameworks: The construction and use of multilingual proficiency frameworks* (Studies in Language Testing 40). Cambridge: CUP.

Mehisto, P., & Genesee, F. (Eds.). (2015). *Building bilingual education systems: Forces, mechanisms and counterweights* (The Cambridge Education Research Series). Cambridge: Cambridge University Press.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge: Cambridge University Press.

Nikula, T., Dafouz, E., Moore, P., & Smit, U. (Eds.). (2016). *Conceptualising integration in CLIL and multilingual education*. Bristol: Multilingual Matters.

Shaw, S. D., & Imam, H. C. (2013). Assessment of international students through the Medium of English: Ensuring validity and fairness in content-based examinations. *Language Assessment Quarterly, 10*, 452–447. https://doi.org/10.1080/15434303.2013.866117.

References

Achtenhagen, F. (2012). The curriculum-instruction-assessment-triad. *Empirical Research in Vocational Education and Training, 4*, 5–25.

Beacco, J. C. (2007). *A descriptive framework for communicative/linguistic competences involved in the teaching and learning of history*. Strasbourg: Language Policy Division.

Beacco, J. C. (2010). *Items for a description of linguistic competence in the language of schooling necessary for learning/teaching history (end of obligatory education): An approach with reference points. Language and school subjects. Linguistic dimensions of knowledge building in school curricula*. No.1. Document prepared for the Policy Forum 'The right of learners to quality and equity in education – The role of linguistic and intercultural competences' Geneva, Switzerland, 2–4 November 2010. Language Policy Division, Strasbourg.

Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box: Assessment for learning in the classroom*. London: King's College London School of Education.

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *Behavioral and Brain Sciences, 27*, 79–151. https://doi.org/10.1017/S0140525X04000032.

Chamot, A. U., & O'Malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Reading: Addison-Wesley.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Modern Languages Division, Council of Europe.

Council of Europe. (2009). *Relating language examinations to the Common European framework of reference for languages: Learning, teaching, assessment (CEFR): A manual*. Strasbourg: Language Policy Division, Council of Europe.

Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors. Provisional edition*. Language Policy Programme. Education Policy Division. Education Department. Council of Europe. www.coe.int/lang-cefr. Accessed 19 Mar 2019.

Coyle, D. (2006). Content and language integrated learning. Motivating learners and teachers. *The Scottish Language Review, 13*, 1–8.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: content and language integrated learning*. Cambridge: Cambridge University Press.

Cummins, J. (1980). The entry and exit fallacy in bilingual education, reproduced in Baker, C. and Hornberger, N. (Eds.). *NABE Journal, 4*, 25–60.

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualizing integration in CLIL in multilingual education* (pp. 29–54). Bristol: Multilingual Matters.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly., 43*(3), 393–415.

Doran, G. T. (1981). There's a S. M. A. R. T. way to write management's goals and objectives. *Management Review, 70*, 35–36.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Gass, S. M., Doughty, C. J., & Long, M. H. (2007). *Input and interaction. The handbook of second language acquisition*. Malden: Blackwell Publishing.

Imam, H. (2010). *Calpability: Achieving in content through language: Teacher perceptions, examiner expectations, and student performance in IGCSE history*. Unpublished MA dissertation, Anglia Ruskin University, sponsored by Cambridge Assessment.

Jones, N. (2014). *Multilingual frameworks: The construction and use of multilingual proficiency frameworks* (Studies in Language Testing 40). Cambridge: CUP.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lengyel, D. (2010). *Language diagnostics in multilingual settings with respect to continuous procedures as accompaniment of individualized learning and teaching*. Strasbourg: Language Policy Division. Directorate of Education and Languages, DGIV Council of Europe.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & S. Ute (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol: CPI Books Group Ltd.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bahtia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic.

Lorenzo, F., & Dalton-Puffer, C. (2016). Historical literacy in CLIL: Telling the past in a second language. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 55–72). Bristol: Multilingual Matters.

Massler, U., Stotz, D., & Queißer, C. (2014). Assessment instruments for primary CLIL: The conceptualisation and evaluation of test tasks. *Language Learning Journal*, *42*(2), 137–150. https://doi.org/10.1080/09571736.2014.891371

Mehisto, P., Marsh, D., & Frigols, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.

NALDIC. (2009). *Introduction to EAL formative assessment*. http://www.naldic.org.uk/Resources/NALDIC/Teaching%20and%20Learning/NALDICEALFormativeAssessmentIntroductionfinal.pdf. Accessed 18 Dec 2018.

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

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H.-C., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco: Jossey-Bass.

Shaw, S. D., & Imam, H. C. (2013). Assessment of international students through the medium of English: Ensuring validity and fairness in content-based examinations. *Language Assessment Quarterly, 10*(4), 452–447. https://doi.org/10.1080/15434303.2013.866117.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Rowley: Newbury House.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(3), 371–391. https://doi.org/10.1093/applin/16.3.371.

Srole, C. (1997). Pedagogical responses from content faculty: Teaching content and language in history. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom: Perspective on integrating language and content* (pp. 104–116). White Plains: Longman.

Thürmann, E., Vollmer, H., & Pieper, I. (2010). *Language(s) of schooling: focusing on vulnerable learners*. Strasbourg: Language Policy Division, Council of Europe.

Vygotsky, L. S. (1981). The genesis of higher mental functions. In J. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 144–188). White Plains: Sharpe.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

WIDA. (2014). *The cornerstone of WIDA's standards: Guiding principles of language development*. http://www.cesdp.nmhu.edu/ccss/pdf/The-Cornerstones-of-WIDA-Standards-Guiding-Principles-of-Language-Development.pdf. Accessed 18 Dec 2018.

Chapter 3
Academic Culture as Content: Self-Assessment in the CLIL Classroom in the International Liberal Arts University

Alexander Nanni and Chris Carl Hale

3.1 Introduction

This chapter conceptualises CLIL within the greater context of the English medium liberal arts university in Asia and presents practical approaches to student self-assessment by examining data from academic writing classes in Japan and academic discussion classes in Thailand. Both examples are drawn from contexts that could be understood as "lower disciplinary orientation with more visible language pedagogy" (Leung and Morton 2016, p. 241), or Quadrant 3 of the matrix described in Chap. 1 of this volume. However, the two activities presented here were not typical of this context as learners were not typical either—they were advanced-proficiency students and enrolled at universities where English was the medium of instruction (EMI). Therefore, the activities are best understood with reference to Quadrant 1 of the matrix, in that there was more focus on "subject literacies" in addition to language control. Both examples share the objective of acculturating students into the liberal arts academic community as an explicit course goal. As an additional means of achieving this objective, the educators in both contexts make use of student self-assessment.

The value of self-assessment in CLIL is consistent with the idea of CLIL as "two for the price of one" (Bonnet 2012; see also Chap. 1, this volume). Self-assessment in language alone would not welcome students as emerging members of the community of experts, nor would it push them to improve their metacognition as it

A. Nanni (✉)
Humanities and Language Division, Mahidol University International College, Salaya, Thailand
e-mail: alexander.nan@mahidol.ac.th

C. C. Hale
Akita International University, Akita, Japan
e-mail: chale@aiu.ac.jp

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_3

relates to the content area. Similarly, self-assessment in content alone would not share ownership of the language with students in the same way, nor would it push them as strongly to develop their metalinguistic awareness. The content objectives are inextricable from the language objectives, and therefore the two must be assessed together. The communicative approach to language teaching, which emphasises the ability to engage in meaningful exchanges much more than it does grammatical accuracy, is fundamentally aligned with CLIL. Students' language ability is gauged largely by their ability to communicate meaningfully about content, and the content inherent in CLIL provides substantial issues about which learners can communicate. The content that students are engaging with is not merely a tool for the development of their language skills; it is significant in itself.

Self-assessment in CLIL provides opportunities to learn. This fits the discussion of assessment promoting learning, namely assessment *for* learning (A*f*L) culture discussed in Chap. 1 of this volume: it is the *process* (not the product) of learning. Rather than emphasising grades, it focuses on direction for further development. Self-assessment achieves this by developing learners' metacognition and metacognitive awareness. Self-assessment adds an additional dimension that is relevant to their liberal arts education as a whole: the development of their voice and authority as a member of the academic community. By participating in self-assessment, they are beginning to establish themselves as members not only of a community of learners, but also of a community of assessors.

Self-assessment has emerged as a possible solution for educators who exhaustively devise ways to make students more active in the learning process. Compelling research suggests that the ability of students to accurately self-assess their own learning is perhaps the single most salient factor in determining the effectiveness of their education (Hattie 2008). In addition to enhancing student autonomy, agency, and metacognitive awareness of learning, it is also a way to encourage critical thinking and responsibility—traits that will serve students well in their lives after graduation. However, assessment in higher education is often at odds with the purported values espoused by universities (Boud 2000), and in particular the liberal arts university, where there is meant to be a focus on fostering students' abilities "to reason clearly, to think independently, to solve problems elegantly, and to communicate effectively" (Kim 2011). Compounding this challenge, then, is how to incorporate these lofty goals with the other assessment needs inherent to the CLIL context, namely measuring content knowledge and language proficiency. In addition, when students are attending an English medium instruction (EMI) institution in Asia based on the North American model of higher education, they are also experiencing a concomitant "paradigmatic and cognitive shift from what could be loosely called the Confucian-style transmission model to the Socratic argumentative model" (Hale and Wadden 2013). In the context of international liberal arts colleges based on the North American model, language and content in the CLIL classroom cannot be separated from the academic culture inherent to the institution itself; therefore, assessment practices must in some way reflect the overall institutional philosophy of the liberal arts. A question, thus, arises: how can students be assessed in such a way that all of the competing needs (language, subject content, *and* promotion of

the overarching academic culture) are met? We believe that self-assessment, when conceived as its own learning outcome, can achieve this goal.

The qualitative findings indicate that by carefully curating content, thoughtfully designing assessment prompts, and, most critically, *incorporating students themselves into the assessment process* (self-assessment), teachers can satisfy the immediate summative assessment needs of the course (that of evaluating content comprehension and language control), while strongly promoting learning as espoused in the liberal arts. After reviewing the literature on self-assessment, this chapter describes two contexts in which self-assessment has been applied and provides details of studies carried out in these contexts. It then discusses the commonalities of these studies and possible implications for self-assessment in CLIL.

3.2 Literature Review on Self-Assessment

With the rapid increase of CLIL programs at Japanese universities and elsewhere in East Asia (Brown and Bradford 2019), the limited understanding of CLIL has led to questions regarding how to conduct assessment. At the same time, student self-assessment has been growing more common as a means of increasing student motivation, autonomy, and enhancing metacognitive awareness of the learning progress (Mehisto and Ting 2017). A purported goal of university learning is the promotion of critical thinking and recognition of one's place and responsibility in an academic community, and it would seem that self-assessment could promote the overall pedagogical philosophy of the university. If we consider it possible to perform summative assessment of an entire university experience, we might say that the ultimate outcome should be the ability of students to evaluate their own academic abilities thoughtfully and objectively, and be in possession of the skills to identify areas to improve. At the liberal arts university, this ability would seem to speak directly to the values and belief systems inherent to the university's ideals, not the least of which is the preparation of young minds in the pursuit of recognising and expanding their intellectual abilities, and then applying that knowledge to solving the world's problems for the betterment of humankind.

These goals aside, the literature seems, nonetheless, inconsistent in terms of the actual utility of self-assessment in various learning contexts. Some authors praise the practice (Blanche and Merino 1989; Gardner 2000; McDonald and Boud 2003; Patri 2002), while others report serious concerns in terms of actual reliability and validity (Blue 1994; Huang 2010; Matsuno 2009; Oldfield and MacAlpine 1995; Sullivan and Hall 1997). Much literature on self-assessment suggests that student self-assessments have a low correlation with ratings of professional raters (most commonly their teachers). In the East Asian context, Matsuno (2009) used multifaceted Rasch measurement to find that while students tended to rate their peers more highly than teachers, they consistently underrated themselves in comparison to teachers. It was speculated that this was perhaps a result of their East Asian cultural conditioning to appear individually modest while reverential to peers. This led the author to conclude, like many others, that self-assessment was less accurate and

therefore less valuable than other assessments, such as teacher-only grading. It should be stated that most of the research looking at self-assessment is from the perspective of summative assessment, that is, evaluation of discrete learning tasks, and in nearly all cases the students' ratings were for reflective purposes only and not considered in the formal grading process.

Researchers have proposed several benefits of self-assessment that are unrelated to accuracy. These include sharing authority in the classroom, increasing motivation, developing metacognition, and enhancing learning. Bedore and O'Sullivan (2011) promote self-assessment as a means of "removing the instructor from the position of sole authority" (p. 13). This sharing of authority is conducive to student-centred learning and the development of critical thinking. In a review of the literature on student self-assessment, Blanch and Merino (1989) identified many studies associating students' active involvement in assessment and increased motivation. In more recent studies, Harris (1997) and Gardner (2000) concluded that self-assessment was associated with higher levels of both motivation and learner autonomy. While also mentioning the sharing of authority, Sadler (1989) highlights the impact of self-assessment on metacognition, enabling students to develop their own evaluative knowledge. By engaging in the process of assessment, students become 'insiders' rather than 'outsiders' as they begin to understand and internalise the difficulties teachers face when making assessments. Further support for the benefits of self-assessment was provided by Hattie (2008), whose large-scale quantitative study investigated more than 100 factors influencing learning and identified accurate self-assessment as the single most salient indicator of learning.

Despite the demonstrated benefits of self-assessment and its alignment with the student-centred approach to education, it remains underutilised. Hilgers et al. (2000) observe that while many teachers accept in theory that self-assessment is beneficial, relatively few implement it, possibly due to concerns about accuracy. This gap between theory and practice should be remedied, particularly as many current practices in assessment undermine some of the purported goals of higher education. Boud (1990, 2000) argues that self-assessment fosters the growth of personal responsibility and critical thinking, qualities that will be of great value both at university and beyond. In contrast, assessment solely by teachers creates a situation of "unhealthy dominance" (p. 106) in which the development of student responsibility is stunted. Self-assessment is compatible with the stated goals of higher education in general and of liberal arts programs in particular, it can develop metacognition and responsibility in ways that assessment by teachers cannot, and it is aligned with the student-centred approach to education.

It must be noted that such a student-centred philosophy to assessment is not prominent in the education models of East Asia, including Japan and Thailand (Hallinger 2010; Marginson 2011). In these contexts, the emphasis is on rote learning and reproduction of learned forms, much less on displaying individual, creative and autonomous thought. Therefore, introducing student-liberating pedagogies to East Asian contexts may be met with confusion or resistance, at least initially, even at liberal arts universities in Asia that are based on the Western model.

3.3 Context

What follows are two practical presentations of how this approach to content and language learning is conceived of and approached at two international liberal arts universities in Asia. The first example describes a CLIL-based writing program from a university in Japan, and the second describes a CLIL-based academic discussion program in a Thai university. Before examining each case in detail, it is useful to consider the common background of these two applications of CLIL. Both courses were created for students who are just beginning their studies at a liberal arts university. Many of the students enrolled in these courses are in transition from a more traditional style of education to one that is more student-centred. Navigating this transition well is essential in ensuring their success in a liberal arts environment. In both cases, learners are self-assessing productive skills—writing and speaking—which can be assessed directly.

The first example presents the introduction of self-assessment in a freshman academic writing course at a Japanese liberal arts university. Over the 15-week term, at approximately four-week intervals, the 23 students in the course were given three timed-essay tests in which they were given a writing prompt related to the theme covered in the section of the course. The three four-week themes in the course were (1) educational values, (2) intercultural communication, and (3) bioethics. In addition to the timed-essay tests, student assessments included multiple-draft essays on each theme, presentations, and language-focused assessments (vocabulary, sentence structure quizzes, etc.) The data and discussion here are from the third timed-essay test covering the section on bioethics. Below are the CLIL objectives for the course as a whole:

Language Learning Outcomes.

1. Essay structure and format conventions (central claims, specific claims, etc.)
2. Ability to paraphrase information
3. Use of coherent and grammatical academic register in writing
4. Oral expression accuracy and proper academic register in presentations

Content Knowledge Outcomes.

1. Show understanding of the core concepts from the course readings and lectures
2. Show ability to synthesise information from course readings and class lectures while exhibiting original thinking on the topic
3. Display well developed, logical, and cogent opinions on the topic in writing
4. Display well developed, logical, and cogent opinions on the topic in oral presentations

The second example explores the use of student self-assessment in formative assessment of small group discussions in an upper-intermediate intensive English course at a Thai university. The CLIL approach is evident in the course objectives of this program. In addition to objectives related to language proficiency and study skills, the objectives of the course are as follows:

1. Construct an essay about a complex issue
2. Discuss an issue relating to global current events
3. Demonstrate capacity to evaluate crucial global issues connected to society, the economy, and the environment

The first two goals on this list have both linguistic and content components. Writing an essay about a complex issue requires learners to engage with texts and lectures, using content from these sources to respond meaningfully to issues that are beyond their direct experience. The reference to academic readings and lectures as part of the writing process provides opportunities for students to write about cognitively challenging topics similar to those that they will engage with at university. The second objective, discussing an issue relating to global current events, is similar to the first in that it requires learners to integrate language skills to engage in realistic communication in an academic context and to learn content as well as language. The third objective, demonstrating capacity to evaluate crucial issues, has a strong content focus. This is related to Kim's (2011) statement that liberal arts universities are meant to foster clear and independent reasoning and problem solving as well as communicative competence. The linguistic and content goals of the course are described in greater detail below.

Despite their differences, the two cases described below share significant commonalities. Both courses share similar CLIL objectives, and both aim to empower students and enhance their metacognition through the application of self-assessment in CLIL. In each example, language, content, and critical thinking are integrated into a single activity, demonstrating the added value of CLIL.

3.4 Case 1: Self-Assessment in a CLIL Writing Course in Japan

The process normally followed for high stakes assessment at the university involves each student essay being blind rated by two instructors using the rubric included in Appendix A. In this rubric, content knowledge (and students' ability to show their understanding of it) is represented on the left side and assigned 10 points. The language control is represented on the right side of the rubric and is assigned five points. The essays can be awarded a total of 15 points by each instructor, and each score is summed for a maximum score of 30 points. The two total scores of the raters are required to be within three points of each other. If the two instructors' scores differ by four or more points, a third instructor serves as an additional rater, and any two of the three scores that differ by less than four points are used in the assessment. The above process is preceded by a 'norming session' in which instructors meet and practice rating example essays to establish a standard. After each practice rating is done, instructors compare scores and discuss differences in scores greater than three. Through negotiation, instructors adjust the scores falling outside the range of the three points difference or less until all agree on how the piece of writing should be rated. Then the process is repeated for all remaining example essays. In this way, consistency is maintained among instructors doing the rating with data showing

inter-rater reliability of 80% or more commonly attained. Following the norming session, instructors receive the first of two packets of essays, the scores for which are recorded on a separate document to maintain the blind rating. Upon completion of the first packet of essays, each instructor exchanges packets with another instructor and rates those essays, recording scores on the essay form itself. After completion of the second packet, a testing coordinator enters the first scores on the test forms and determines which essays require a third rater.

In the process under discussion here, the students' classroom teacher acted as the first rater, and the students who wrote the essays *served as the second rater for themselves*. In the case of a discrepancy of more than three points between the teacher and student raters, just as is normal procedure when two teachers are blindly rating the test, a third rater would be called—in this case another teacher. This self-assessment was done for three essay tests that were taken over a fifteen-week term. The rubric for all three tests was identical (see Appendix A) and created based on a department-wide rubric used for assessing essay writing at the university. Criteria for this rubric were agreed upon by the department for internal rating of essays where an emphasis was placed on content (10 points) over writing convention and control (5 points). As there were two raters, the total points available was 30 (15-point scale × two raters = 30 points). Prior to each self-rating, students and their classroom teacher engaged in a norming session using sample essays from previous terms answering the same writing prompt the students also answered. Three essays were used per norming session. Following the norming session, students were given back their own essays with the first rater's (teacher's) score removed, and given approximately 15–20 minutes to read and evaluate their own essay. After assigning themselves a score on the rubric, the essay and the rubric were returned to their teacher, who then added their score to the rubric sheet and either summed the scores if the scores were within three points of each other or gave the essay to a third rater (another teacher) if more than three points apart. The rubric was slightly modified to reflect the fact that in this activity, each essay was given a value of five points (rather than 30), requiring the 30-point rubric to be divided by six in order to generate the earned score (30/6 = 5) (see Appendix C). This alteration was made because the course included three essay tests, and the maximum value for all tests was decided by the university department (a total of 15 points of the students' course grade was assigned to the three essay tests). In order for students to become familiar with the rubric and internalise the grading criteria, it was deemed necessary to maintain the 15-point scale rather than create a different 5-point rubric for these essay tests. Incidentally, three student term papers were rated at 15 points each (by teachers only) using the same rubric, which also reinforced the grading criteria.

At the end of the term-long process, the students were given a short questionnaire in order to ascertain their perceptions of self-assessment. While the writing test itself represented the summative acquirement of classroom content and language control, perceptions of the act of self-assessment were seen as important in determining the extent to which the activity acted as another kind of assessment, that is, the long-term impact it may have had on their learning and the broader implications to their overall education at the university, as well as how it might affect other areas of their lives, having thus formative function for the students. While not

the focus of this chapter, student reflections on the process did in fact produce salient data in earlier treatments (See Hale 2015 for a more comprehensive examination of student perceptions of having participated in self-assessment on graded tests). The participants in this treatment shared an overall language proficiency level of "advanced," or CEFR C1.

The segment of the course dealing with bioethics included authentic readings and class lectures related to the topic in areas such as eugenics, CRISPR technology (a technology used for editing human DNA sequences), and human bioengineering. The primary reading for the section was *The Case Against Perfection: What's Wrong with Designer Children, Bionic Athletes and Genetic Engineering* by Michael J. Sandel (2012). Below, in Fig. 3.1, is the writing prompt on which students wrote

Quiz #3: Argumentation Timed Writing

Consider the case of a deaf, lesbian couple who would like to have a child. Because they can only conceive through in vitro fertilization, they will have the opportunity to choose some of the traits through bioengineering before the fertilised egg is implanted into one of the women. The women were both born deaf, and have strong identities as deaf people. They only know the world as deaf people, and they think that they could best raise a child who is also deaf. They believe deafness is a distinct trait, like any other, such as hair or eye colour, height and even culture or race. They do not think of themselves as "disabled," but instead take great pride in their deafness and being part of the active deaf community. To them, deafness is an "enhancement" and hearing is a "disability."

They believe that if their child is also deaf, they can share their strong identities as deaf people with their child, and communicate more deeply with him or her. Also, they want to avoid the many communication difficulties other deaf parents have raising a "hearing" child.
For all these reasons, they ask their doctor to bioengineer their child to be deaf, and the doctor agrees. However, before they could proceed, a human-rights group hears of the plan and sues the couple to stop them. This group believes that using bioengineering to purposely make a child deaf is immoral and wrong.

Imagine you are the lawyers for *either* the lesbian deaf couple OR the human-rights group. Before the trial begins, the judge asked you to write an argumentative essay defending your clients, and stating why they are correct in this case. In your argument, use at least three ideas from the Sandel (2012) reading (they can be anything — examples, ethical defences, key concepts, etc.).
You have 45 minutes.

Fig. 3.1 Essay topic: bioethics

for 45 min. The writing prompt was adapted from an ethical dilemma also introduced by Sandel (2007).

The overall number of students overrating themselves remained consistent on all tests. In only five cases did the same students overrate themselves on all three tests, however. The spike in students underrating themselves on Test 3 represents an overreaction to having previously overrated themselves. In general, in each subsequent treatment, students fluctuated in their ratings (higher or lower than teacher raters), yet largely did so within the 80% interrater reliability threshold. As the number of third raters needed declined with each treatment, it can be said that students gradually improved in their self-assessment abilities, which is to say, their assessments drew more consistent with professional, teacher raters.

Students tended to overrate themselves in the content knowledge portion of the rubric. This may be due to an overconfidence in their abilities to display this knowledge, as well as there being a larger point spread in that area—thus making it more difficult to precisely rate their abilities. Conversely, students more accurately rated their language abilities, perhaps a result of their prior learning focus on language items in classes taught in English, or again, perhaps owing to the smaller point spread of that area of the rubric. Because of the potential for spread variance, it is perhaps more constructive to look at the total points possible (15) rather than at the individual sections. Table 3.1 presents the variance between student and teacher raters over all three treatments. 1 presents the variance between student and teacher raters over all three treatments. As the table illustrates, the need for third raters declined over time, while the number of students rating themselves more highly than teacher raters remained consistent over the three treatments, and students "under-rating" themselves increased over the three treatments, which correlates with the number of students rating themselves identically to teacher raters decreased.

The example essay in Appendix B, which is an example of the third treatment, was more positively rated by the teacher than the student who wrote it (see Appendix C), which follows the trend of more students underrating themselves on the third treatment than the first two. Understanding of course content is well represented in this response as the writer took the more difficult position of arguing for the deaf couple to have a deaf child. The student's arguments and support are logical, well-reasoned, and display original thinking on the topic. The language is also fairly

Table 3.1 Student self ratings

n=23						
	3rd raters needed (student higher)	3rd raters needed (student lower)	Student rating higher than teacher (but within three points)	Student rating lower than teacher (but within three points)	Teacher and student rating identical	Totals
Test 1	3	1	8	5	6	23
Test 2	2	1	8	7	5	23
Test 3	1	0	8	11	3	23
Totals	6	2	24	23	14	69

accurate, with only minor word choice and grammar mistakes. However, while the teacher awarded this essay 9 + 4 on the two parts of the rubric, the student underrated themselves by two points on the content knowledge part of the rubric and one point on the language control portion (7 + 3).

As students were not experienced in self-assessment prior to this activity, therefore their initial ratings were more "educated guesses" than anything else. While the trend was for students to overrate themselves in the first two treatments, it can be said that many of their first ratings, even those that matched with teacher raters, were influenced by this upward trend. This trend was largely corrected by the third treatment when student self-assessments were more informed and, we believe, accurate. After only three treatments, it is reasonable to see students "overshooting" and "undershooting" their performance, with the ultimate goal of student and teacher assessments getting closer together over each treatment, which was the case here. This fluctuation is normal when attempting to achieve interrater reliability among human raters, where any score within the 80% range is considered "statistically the same score."

While the three-point difference is within the range of interrater reliability, when looking at the student essay in Appendix B, it is clear that the student would have gained more points overall on the test had they overrated themselves. It is difficult to predict, however, how another teacher-rater would have evaluated this essay, though one would assume that a professional rater, such as a teacher, would be more closely aligned with other experienced teacher-raters. Whenever introducing self-assessment in ways that are reflected on final course grades, there is always the chance that students' grades will be negatively affected by their own assessments, and there may be little comfort in the notion that 10 points is "statistically the same score" as 13 points. Following this term-long exercise, students are asked to reflect on the process, and one of the questions asked is if they feel the grades they receive "fairly and accurately represent their learning." Historically, students have answered this question in the 70–80% range, indicating that there is still some room for improvement in terms of how we can achieve 'buy-in' for self-assessment by the very people it is intended to empower.

3.5 Case 2: Self-Assessment in a CLIL Speaking Course in Thailand

This section describes the application of CLIL self-assessment in an intensive English for Academic Purposes (EAP) course in Thailand. During the 10-week term, students regularly participate in small group academic discussions. These discussions are a major component of the course, and they are used in both summative and formative assessment. Often, the primary learning objectives for students' academic speaking development are accuracy and fluency (Folse 2006); however, accuracy and fluency alone are not sufficient to facilitate students' engagement in

meaningful interaction. In addition to these factors, the ability to convey meaningful content, to interact with others, and think critically are key.

Throughout the course, students are reminded of the purpose of the small group discussions: to prepare them to participate confidently and competently in the English-medium liberal arts program at the affiliated international college. Such participation requires more than linguistic skill—students must also be prepared to engage with cognitively challenging real-world content. For this reason, the language centre has adopted a CLIL approach in its upper-level courses. This chapter focuses on the upper-intermediate level course, the final course that students take before matriculating to the college. This 10-week course is comprised of a 10-hour-per-week writing class, an 8-hour-per-week integrated skills class, and a 6-hour-per-week reading class. In each of these classes, teachers use a wide range of materials—websites, texts, lectures, videos, etc.—related to a set of three themes. The students are responsible for being able to understand, discuss, and write about (using APA citations and references) issues related to the themes. There are two sets of themes, and each set of three themes is linked to a term paper that builds on key ideas covered during the term. In terms where the three themes are urbanization, climate change, and nuclear power, students write a term paper on innovation and technology. In terms where the three themes are business ethics, globalization, and food, students write a term paper about business sustainability using Elkington's (1997) Triple Bottom Line framework. Students engage in at least one small group discussion about each theme. The small group discussions described in this chapter are central to developing students' linguistic skill as well as their ability to engage critically with unfamiliar topics.

The rubric used for the discussion reflects the learning objectives of the course and of the activity. It includes three dimensions: *original input and use of sources, ability to interact, and academic language skills*. To pass the summative assessment, students must achieve a certain minimum standard in each of these three areas. The complete discussion rubric is included in Appendix D.

The first dimension of the rubric, *original input and use of sources*, focuses on the quality and quantity of students' contributions to the discussions. To achieve in this area, participants in the discussion should make multiple contributions, demonstrate that they understand the source material, integrate their use of source material into their contributions effectively, and provide spoken citations (i.e. indicate the source of their information) for some key points. The second dimension of the rubric, *ability to interact*, assesses students' ability to interact meaningfully with the other members of their group. Interaction could take the form of agreeing/disagreeing, asking for clarification, questioning, or expanding on the ideas of others by providing further data or examples. Students' interaction should evince an understanding of the overall direction of the discussion, and their interaction should help to progress the discussion. The third dimension of the rubric, *academic language skills*, encompasses participants' ability to communicate about a complex topic in an academic register of English. They should be able to summarise and paraphrase effectively and use key vocabulary from their various sources, all while minimising strain for their listeners. To pass the discussion, students must achieve a passing

score in each of the three dimensions. That is, both the content and linguistic criteria must be met.

The rubric has been designed to align with the objectives of the course. To achieve a passing score in the discussion, students are expected to draw on their knowledge of the issue and similar issues that they have discussed and written about over the course of the term, evaluating the issue critically and engaging meaningfully with the content and with the ideas of their group members. Both the first and second dimensions of the rubric require significant comprehension and knowledge of the topic at hand as well as certain language skills; only the third dimension of the rubric primarily assesses language.

In September of 2017, a group of five students engaged in a discussion on business ethics based on materials in Fig. 3.2 below. The five participants were randomly selected from a group of 21 students. Their entire class took notes on each of the two sources provided by the teacher, discussed these two sources, and engaged in short pair discussions of questions 2, 3, and 4. They also completed individual research, preparing notes on at least three additional sources. This discussion on business ethics took place as part of a larger unit on business ethics, and it was the second discussion on this topic. Successful participation in the discussion would require students to draw on the content knowledge that they had developed since the beginning of the term, not only on the sources immediately at hand. They were also expected to incorporate content on the Triple Bottom Line, on which their term paper is based. Questions 3 and 4 ask explicitly about Thailand; however, the two sources provided by the teacher discuss the meaning of business ethics (in the case

Business Ethics Discussion

We will have a discussion centred on these two sources:

- *Five Ways to Think Ethically* — Markula Center for Applied Ethics, Santa Clara University (2011)
- "Corruption Causes Business Inefficiency" — Rigoglioso (2007)

The discussion prompts are as follows:

1. Summarise the video and article.
2. How do corruption and unethical business practices affect companies? Discuss this in terms of the Triple Bottom Line.
3. What kinds of corruption harm Thailand's businesses?
4. How can corruption in Thailand be reduced?

Prepare notes (one sheet of paper, front and back) about the two main sources. In addition, you must include information from at least ***three*** further sources.

Fig. 3.2 Business ethics discussion materials

of *Five Ways to Think Ethically*) and the consequences of corruption for businesses (in the case of "Corruption Causes Business Inefficiency"). The participants needed to base their responses to these questions not only on the two sources provided, but also on their individual research and their personal experiences.

The discussion took place in the students' integrated skills class. The five participants sat around a table at the front of the room; the teacher and other students sat towards the back of the room facing the students engaged in the discussion. The 20-minute discussion was video recorded and transcribed. After the discussion, students completed a self-assessment based on the rubric. They then received comments from their peers and teacher. A partial transcript of the discussion is included in Appendix E. The transcript contains the majority of the group members' responses to the question of how corruption in Thailand can be reduced.

After the discussion, the five students were asked to participate in self-assessment and to reflect on their performance. They submitted responses to a series of questions in a Google Form, which asked students to self-assess using a series of Likert-scale questions based on the rubric. The rubric had been shared and discussed at the beginning of the term, so the participants were familiar with the dimensions of the rubric and the terminology. To help students differentiate their scores, the Likert-scale questions used a seven-point scale. A score of three or below indicated failure of the component being assessed (64% or below), a score of four was equivalent to a bare pass (65%), scores of five or six indicated a high pass (65% to 77%), and a score of 7 indicated distinction (78% or above). To further assist in differentiating the scores, the content dimension of the rubric (*original input and use of sources*) was subdivided into three areas—amount and relevance of contributions, demonstration of content knowledge and understanding of sources, and use of specific evidence from sources. The self-assessment scores from these three areas were averaged to obtain the overall self-assessment score for the content dimension of the rubric. Other two dimensions of the rubric—*ability to interact* and *academic language skills*—were each self-assessed using a single seven-point Likert-scale item.

Table 3.2 below shows the self-assessment data provided by the students. None of the students recorded a failing self-assessment score (3 or below) in any category.

The assessment scores given by the teacher following the discussion are contained in Table 3.3. The teacher recorded three failing scores (S1 in *original input and use of sources* and *ability to interact*; S5 in *academic language skills*). The range of the scores is also wider. The difference between the highest and lowest

Table 3.2 Student self-assessment scores on seven-point likert scale

Student	Original input and use of sources	Ability to interact	Academic language skills	Average
S1	3.67	5	4	3.56
S2	4.33	5	5	4.78
S3	6.33	5	5	5.44
S4	5.67	6	6	5.89
S5	4.33	5	4	4.44

Table 3.3 Teacher assessment scores on seven-point likert scale

Student	Original input and use of sources	Ability to interact	Academic language skills	Average
S1	3.33[a]	3[a]	4	3.44
S2	4.67	5	5	4.89
S3	6.00	5	5	5.33
S4	5.33	6	5	5.44
S5	5.00	5	3[a]	4.67

[a]A score of 3 or below in any of the three dimensions of the rubric would result in a failing score overall

Table 3.4 Differences between teacher and student assessments[a]

Student	Original input and use of sources	Ability to interact	Academic language skills	Average
S1	1.33	2	0	1.11
S2	−0.33	0	0	−0.11
S3	0.33	0	0	0.11
S4	0.33	0	1	0.44
S5	−0.67	−1	1	−0.22

[a]These were calculated by subtracting the teacher score from the student score

scores in the student self-assessments is 1.44, while the difference between the highest and lowest scores assigned by the teacher is 2.00.

Table 3.4 shows the difference between the teacher and students' scores. The results in this table were calculated by subtracting the teacher's scores from the students' self-assessment scores. A positive result indicates that a student's self-assessment was higher than the score assigned by the teacher, while a negative score indicates that the teacher's score was higher. With the exception of S1, the student and teacher average scores were similar.

The level of content knowledge demonstrated by the students generally indicated a well-developed understanding of business ethics, and they were able to integrate knowledge from a previous course, personal experience, the two provided sources, and their own research. Sources cited in the transcript include GAN Integrity (a firm based in Denmark) and the Cornell University website; one student also offered a personal example from her family business. One minor weakness was the student's lack of explanation of what GAN Integrity Solutions is—while this is a sufficiently credible source, not all members of the group would necessarily be familiar with it.

Differences between teacher and student assessment were minor, and the self-assessments of three of the participants—S2, S3, and S4—substantially matched the

teacher's assessments. The self-assessments and teacher assessments of S1 and S5, however, indicate disagreement about whether S1 and S5 would pass the assessment. According to the teacher's scores, S1 would fail due to both the overall average score and failing the dimensions of the rubric assessing content and interaction. In contrast, S1's self-assessment scores are passing overall and in each dimension of the rubric. In this student's case, the main issue was lack of participation as the student spoke approximately 1.5 min in total over the course of the entire 20-minute discussion. The relative lack of participation is also evident in the transcript above, where the student makes only one short comment. While the quality of the student's contributions was adequate, the limited amount of participation resulted in an inadequate demonstration of content knowledge and inadequate interactions with others. For this reason, the teacher assigned failing scores in *ability to interact* and in two of the three areas contributing to *original input and use of sources*—amount and relevance of contributions, and demonstration of content knowledge and understanding of sources.

According to the teacher's scores, S5 would fail due to not meeting the passing criteria in the *academic language skills* dimension of the rubric; however, S5's self-assessment in this dimension of the rubric is four, indicating a bare-minimum pass. These scores indicate differing interpretations of the amount of strain caused by the participant's grammar and pronunciation. The score of four awarded in the self-assessment corresponds to "Causes some strain for the listener, but the meaning is generally clear," whereas the score of three awarded by the teacher corresponds to 'Causes considerable strain for listeners through poor pronunciation and/or grammar.' As indicated by the participant's asking 'Do you understand me?' in the transcript and by other interactions in the full recording of the discussion, feedback from other members of the group provided evidence of considerable strain.

Comparing the results of the self-assessment and the teacher assessment after the discussion was an opportunity to raise the students' self-awareness and evaluative knowledge. This comparison was explicitly conducted in English. As none of the students gave themselves a failing score (3 or below) in any dimension of the rubric, the possibility exists that they deliberately inflated their scores to the minimum level required to pass. If that is the case, the issue is not one of raising their self-awareness and evaluative knowledge but perhaps one of destigmatising low self-assessment scores. When asked to estimate the total duration of her contributions, S1 guessed that she had spoken for a total of 2.5 minutes. This is a full minute longer than the actual duration. The student is now aware that she generally speaks less than she thinks she is speaking, and she will most likely make an effort to contribute more in future discussions. Quantifying the degree of strain caused by S5 is more challenging. Talking about this with the student helps to highlight the difficulties inherent in assessment, some of the "difficulties that even teachers face" (Sadler 1989, p. 135).

Arguments could be made for either passing or failing the student in this dimension of the rubric, but in either case engaging closely with the rubric and the assessment process will help to raise the student's awareness.

One method of drawing students' attention to any discrepancy between the teacher's assessment and the self-assessment would be to have the students watch the videos of their group discussions after receiving the scores and comments from their teachers. Watching their video would allow the students to see their performance more objectively. Afterward, they could complete a short reflection on the relative accuracy of the teacher's assessment and their self-assessment. This would allow the teacher to address any remaining questions or uncertainties that the students may have and would prepare the students to self-assess more effectively in future discussions.

3.6 Discussion

In both cases above, students' self-assessment was fairly consistent with the teachers' assessment. The study of self-assessment of essays by students at a liberal arts university in Japan found that students' ability to self-assess improved over time, and the study of self-assessment in a small-group discussion at a liberal arts university in Thailand found that the self-assessment scores of the majority of participants substantially matched the teacher's assessment. Both studies also identified the importance that students place on the score itself as a possible distorting influence on students' self-evaluation. Students in the Japanese study tended to overshoot the scores awarded by the teacher and reported only a moderate (70–80%) satisfaction rate with the accuracy of their final score. The study in Thailand found that students' self-assessment and the teacher's assessment tended to converge when the teacher awarded passing scores but diverge when the teacher awarded failing scores as no students in the study self-assessed as failing any criteria. While the relative accuracy of self- and teacher assessment is of consequence to educators who wish to implement this type of assessment, other aspects of self-assessment are arguably more valuable in the context of CLIL in the liberal arts university.

The examples explored in this chapter highlight the essential alignment of assessment *for* learning culture, self-assessment, the goals of liberal arts education, and CLIL. Assessment *for* learning emphasises the process rather than the product of learning, focusing on direction and avenues for development rather than scores. It also aims to make students "knowledgeable partners in the learning process" (Mehisto and Ting 2017, p. 224). This reflects the primary aims of self-assessment, which is to raise students' metacognitive awareness (i.e. to make them knowledgeable) and to invite them to be 'insiders' rather than 'outsiders' in assessment (i.e. to make them partners) (Sadler 1989). Similarly, liberal arts education emphasises clear reasoning and independent thinking (Kim 2011), which are fostered both by assessment *for* learning and by self-assessment.

As defined in Chap. 1 of this volume, the goals of CLIL are promoting academic competence, proficiency in L2, and competence in L1. Assessment whose purpose is promoting learning gives direction to the development of academic competence, while self-assessment enhances metacognition, a key element in academic competence. Self-assessment also strengthens student autonomy, another key element of academic competence (Harris 1997; Gardner 2000). While the benefits of self-assessment specifically for L2 development are contested, researchers such as Hattie (2008) have found significant connections between self-assessment and educational effectiveness. This finding suggests the effectiveness of self-assessment in learners' L2 development. As for the goal of L1 competence, sharing authority with learners contributes to the additive nature of L2 development. The teachers' sharing of authority with students explicitly values them as members of the learning community and implicitly values their L1.

Self-assessment seems especially relevant to the approach to CLIL evident in the programs represented here. Leung and Morton (2016) described four orientations to CLIL based on the possible combinations of disciplinary orientation to language and visible subject literacies. These four broad classifications have also been represented as a matrix (Leung and Morton 2016) and were summarised in Chap. 1 of this volume. The geographical contexts from which the two examples in this chapter were drawn would be associated with the Quadrant of the matrix representing lower disciplinary orientation with highly visible language pedagogy. Within this particular approach to CLIL, learning is sometimes situated and mediated (see Chap. 1, this volume). Situated learning is a social process involving co-participation (Lave and Wenger 1991); the sharing of authority inherent in self-assessment aligns well with this. Mediated learning values independence and self-regulation, often with the teacher taking the role of facilitator rather than sole authority (Kozulin and Presseisen 1995). Situated and mediated learning requires teachers to be more than transmitters of knowledge; similarly, it requires assessment to do more than merely obtain information about learners' progress and abilities. However, the rubrics used in both of the studied contexts also elicited content-goals, informing process and making the contexts, thus, also a higher-disciplinary orientation (Quadrant 1) and making students aware of the importance of the disciplinary knowledge alongside language.

Self-assessment in CLIL promotes the larger learning goals of the international liberal arts university in Asia by moving from a traditional teacher-centred process to more student-centred processes with greater emphasis on communication, use of integrated skills and "tests that also teach," i.e. that have pedagogical uses beyond simply measuring test-takers' performance (Richards and Renandya 2002, p. 335). Involvement of students in assessment of their own work may raise issues of validity (i.e. whether the assessment will measure what it is intended to measure), reliability, and objectivity. Indeed, many students feel unqualified to assess their own learning and exhibit lack technical knowledge about how tests and testing are conducted and designed, making it difficult to have confidence in evaluating not only their learning but also the testing and rating instruments themselves. Many teachers also feel there is little value in having students perform something for which they

are not professionally trained. However, these issues are not adequately addressed by excluding students from active involvement in their own assessment; the ability to self-assess is an instrumental part of the liberal arts approach to learning, one that asks students to engage in "a broader, more holistic, more intellectual, and more inquiry-based framework" of language and content learning (Wadden et al. 2012, p. 221).

Because of the resonance between CLIL and A*f*L, CLIL is more than just the context in which A*f*L is practiced. In CLIL, students learn content through language and language through content—i.e. the learning of language and content is integrated. In A*f*L, students' learning is assessed as they progress through the assessment process—i.e. learning and assessment are interwoven. When A*f*L is applied in CLIL, language, content, and assessment are meshed as part of the same pedagogical experience.

3.7 Conclusion

This chapter presented two examples of how self-assessment in the CLIL classroom can be introduced to both effectively assess student learning and achieve other goals of the liberal arts university. In the two studies, students were able to self-assess with a fair degree of accuracy. One of the studies tracked the accuracy of students' self-assessment over time, and found that they were becoming more accurate. In the other study, students' self-assessment was comparable to the teacher's assessment except in cases where the teacher awarded a failing score. While self-assessment in CLIL may be relatively accurate, accuracy is far from the only consideration. Self-assessment in CLIL is consistent with the concept A*f*L as a culture, i.e. with the idea that assessment should be part of the process of learning rather than an event having a summative function. Just as CLIL provides opportunities for learners to develop language skills as well as content knowledge, self-assessment in CLIL provides two major benefits, allowing for the development of metacognitive skills and student voice as a member of a community of experts. The development of metacognitive skills and student voice being goals of the liberal arts university, application of self-assessment is particularly appropriate in this context.

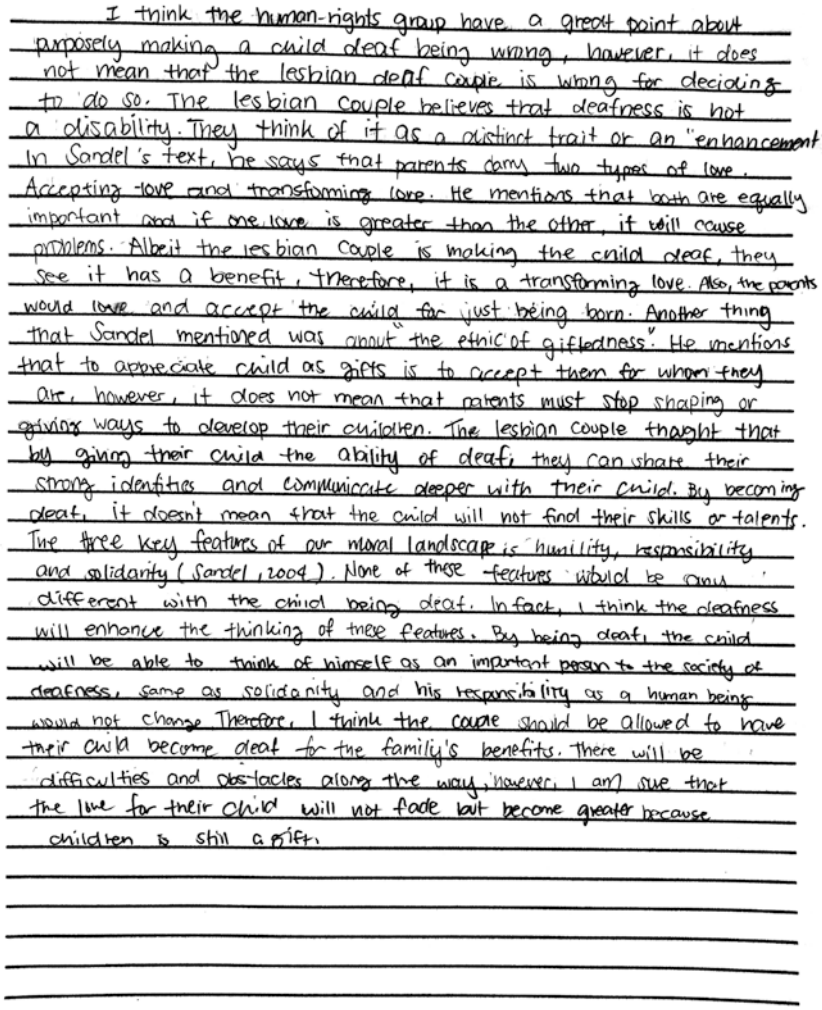
Further studies could build on the research presented in this chapter by investigating the development of self-assessment skills over a longer period; by comparing the relative pedagogical implications of self-assessment when it has an impact on grades and when it has none; or by exploring the relative utility and accuracy of self-, peer-, and teacher assessment of the same texts, and how using and discussing a rubric emphasising both content and language mediates this development.

Appendices

Appendix A: Standard Essay Rating Rubric Used by Teacher Raters

Quality of understanding & response	Writing	Total
9–10 Strong understanding of core concepts, and strong, well-developed support for your opinion.	**4–5** Strong paragraph unity and coherence with topic sentences (including key words from prompt) and transitions.	
7–8 satisfactory understanding of core concepts, with good **d**evelopment and support for your opinion.	**3** Minor problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors.	
5–6 Basic understanding of core concepts and basic development and support for your opinion.	**2** Some problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors.	
3–4 Only partial understanding of core concepts and much more development and support needed for your opinion.	**1** Major problems with paragraph unity and coherence such as lack of topic sentences or transitions and many grammar or word choice errors.	
0–2 Little or no understanding of core concepts with little or no development or support for your opinion.	**0** Little or no paragraph unity or coherence and pervasive grammar or word choice errors.	
First Rater /10	First Rater /5	/15
Second Rater /10	Second Rater /5	/15
Third Rater (if needed) /10	Third Rater (if needed) /5	/15
Final Score /20	Final Score /10	/30

Appendix B: Student Writing Sample, Third Treatment



I think the human-rights group have a great point about purposely making a child deaf being wrong, however, it does not mean that the lesbian deaf couple is wrong for deciding to do so. The lesbian couple believes that deafness is not a disability. They think of it as a distinct trait or an "enhancement". In Sandel's text, he says that parents carry two types of love. Accepting love and transforming love. He mentions that both are equally important and if one love is greater than the other, it will cause problems. Albeit the lesbian couple is making the child deaf, they see it has a benefit, therefore, it is a transforming love. Also, the parents would love and accept the child for just being born. Another thing that Sandel mentioned was about the ethic of giftedness. He mentions that to appreciate child as gifts is to accept them for whom they are, however, it does not mean that parents must stop shaping or giving ways to develop their children. The lesbian couple thought that by giving their child the ability of deaf, they can share their strong identities and communicate deeper with their child. By becoming deaf, it doesn't mean that the child will not find their skills or talents. The three key features of our moral landscape is humility, responsibility and solidarity (Sandel, 2004). None of these features would be any different with the child being deaf. In fact, I think the deafness will enhance the thinking of these features. By being deaf, the child will be able to think of himself as an important person to the society of deafness, same as solidarity and his responsibility as a human being would not change. Therefore, I think the couple should be allowed to have their child become deaf for the family's benefits. There will be difficulties and obstacles along the way, however, I am sure that the love for their child will not fade but become greater because children is still a gift.

Appendix C: Student and Teacher Assessment of Sample in Appendix B

Quality of Understanding & Response	**Writing**	**Total**
9-10 Strong understanding of core concepts, and strong, *well-developed support* for your opinion. **7-8** Satisfactory understanding of core concepts, with good development and support for your opinion. **5-6** Basic understanding of core concepts and basic development and support for your opinion. **3-4** Only partial understanding of core concepts and much more development and support needed for your opinion. **0-2** Little or no understanding of core concepts with little or no development or support for your opinion.	**4-5** Strong paragraph *unity* and *coherence* with topic sentences (including key words from prompt) and transitions. **3** Minor problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors. **2** Some problems with paragraph unity or coherence such as topic sentences or transitions or minor grammar or word choice errors. **1** Major problems with paragraph unity and coherence, lack of topic sentences or transitions and many grammar or word choice errors. **0** Little or no paragraph unity or coherence and pervasive grammar or word choice errors.	
First Rater (Teacher) 9 /10	First Rater (Teacher) 4 /5	13 /15
Second Rater (Student) 7 /10	Second Rater (Student) 3 /5	/15 10
Third Rater (If needed) /10	Third Rater (If needed) /5	/15 23
/20	**/10**	**/30 (/6)**
		3,83/5 **Points**

Appendix D: Rubric for Small Group Discussions

	Original input and use of sources	Ability to interact	Academic language skills	Score
Distinction	Makes a range of relevant and informed contributions Demonstrates clear understanding of topic and sources Integrates and cites provided sources and independent research effectively	Responds to the ideas of others (e.g. agreeing, requesting clarification) Asks relevant questions to initiate and maintain topic Expands on the ideas of others with relevant supporting details and examples Demonstrates clear awareness of the direction of the discussion	Effectively paraphrases and summarises Effectively uses key source vocabulary and grammar to precisely convey meaning Causes minimal strain for listeners to understand meaning	78+
Pass	Makes some constructive contributions Demonstrates understanding of the topic and sources Cites specific evidence from provided sources and independent research	Responds to the ideas of others (e.g. agreeing, requesting clarification), but not always appropriately or in depth Expands on the ideas of others Demonstrates awareness of the general direction of the discussion	Causes some strain for the listener, but the meaning is generally clear Paraphrases or clarifies with some difficulty, but can convey meaning Attempts to use some more advanced vocabulary, but this may affect fluency and pronunciation	65–77
Fail	Did not contribute sufficiently or meaningfully to the discussion Misunderstood topic and/or main ideas from the provided sources Did not cite any provided sources or independent research	Reads or recites prepared notes Interacts in a distracting or inappropriate way Does not demonstrate awareness of the direction of the discussion	Causes considerable strain for listeners through poor pronunciation and/or grammar Unable to sufficiently paraphrase or explain as required Very little attempt to use academic vocabulary Relies on simple language & structures	50–64

Appendix E: Partial Transcript of Small Group Discussion

[start 11:30]
01 S2: I think we should move to the last one [last question]
02 S4: [yeah I think so] how can corruption in
03 Thailand be reduced?
04 S2: personally I think the corruption in Thailand is ... um the people in Thailand
see
05 um watch ... um think that this is the normal thing that everyone do ... corrupt
06 yeah and I think the cause of corruption in Thailand in the regis... legis...
07 legislation is weak and the behaviour in the citizen do not follow the laws... yeah
08 S4: [so]
09 S3: [um-hmm] I think that ah so you you mean ah government need to like [strict]
10 S2: [yeah]
11 S3: more strict the law and this is a good point so I think like government like
12 government need to have the strong punishment to the citizens ... um-hmm
13 S2: from from the GAN Integrity the penalty for paying bribes is around um
14 10,000 Thai baht
15 S4: but um I disagree with you a little bit because um as I found in G.N. ah GAN
16 said that actually Thai law is fine... Thai is, like, more perfect, but the punishment
17 is not... is a very weak one... the government has no authority to um punish ah
18 people who have like who [uh]
19 S2: [who corrupt]
20 S4: yeah who corrupt because some people have like very strong background [yeah]
21 S2: [yeah]
22 S4: I think law is fine but they need to fix with... like
23 S5: strict more
24 S2: strict... more strict
25 S4: yeah

26 S1: um government have to um [inaudible] whole city, not their individual ... they
27 have to be consider about the society, not only their private office like Thai
28 government have to manipulate their money to develop our country like BTS
29 S5: oh facilities
30 S2: [yeah]
31 S4: [yeah]
32 S5: I think last night I found from Cornell University that we should be aware of aware
33 and avoid of corruption it means we try to understand and manage the why they
34 need to corruption and realise of why they need to corruption sometimes they
35 need to run their business or take care their family for they want their money like
36 that and the important thing that I agree that it is do the right thing even we know
37 that corruption is
38 S4: bad
39 S5: bad but sometimes they... they
40 S4: need to [do]
41 S5: [no] they gain money for us but it's bad that we know it means that we know that
42 if we corruption but we make our business run easier but it's bad so we need
43 not... we don't need to do that even we know that it's good for company... like
44 that... do you understand me?
45 S4: if we know that corruption is good [good for the company but we will not do it]
46 S5: [good for the company but yes] but it will harm
47 others so we need to stand strong that it is bad so we need to do ... to avoid it
48 S2: but I would like to say that um we need to realise why they corrupt
49 S4: yeah, um-hmm
50 S2: such as that they... for example maybe they corrupt to um help their family or uh
51 S4: and I think that the most important one is adults in Thailand should be role
52 models to children... so next generation if they saw... if they see like adults do the

53 corruption every day, so they will think it is very common things so of course
54 when they grow up maybe it will be corruption... so should be
55 S3: I have like other situation to... like... to explain to you guys from like my family
56 we are... we have business right, so sometime we have to like
57 S4: corrupt
58 S5: corruption... to pay bribe to a government official
[end 16:05]

Further Reading

Hallinger, P. (2004). Meeting the challenges of cultural leadership: The changing role of principals in Thailand. *Discourse: Studies in the Cultural Politics of Education, 25*(1), 61–73.

Keyuravong, S. (2010). Insights from Thailand. *Learning through English: Policies, challenges and prospects: Insights from East Asia*, 69–95. Retrieved from https://www.teachingenglish.org.uk/sites/teacheng/files/publication_1_-_learning_through_english.pdf

Nanni, A. (2016). *Innovation in English as a foreign language: A case study exploring a program's response to change*. Doctoral dissertation. Retrieved from https://repository.library.northeastern.edu/files/neu:cj82mx32j/fulltext.pdf

Nanni, A., & Brown, D. (2016, June). Reframing the goals of academic speaking: Targeting discussion sub-skills in curriculum design. In *The fourth FLLT conference proceedings by LITU*. Retrieved from http://litu.tu.ac.th/journal/download/pdf/Proceeding_2016_Complete.pdf

Wadden, P., & Hale, C. C. (Eds.). (2019). *Teaching English at Japanese universities: A new handbook*. New York: Routledge.

References

Bedore, P., & O'Sullivan, B. (2011). Addressing instructor ambivalence about peer review and self-assessment. *Writing Program Administration, 34*(2), 11–26.

Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning, 39*(3), 313–338.

Blue, G. M. (1994). Self-assessment of foreign language skills: Does it work? *CLE Working Papers, 3*, 20.

Bonnet, A. (2012). Towards an evidence base for CLIL: How to integrate qualitative and quantitative as well as process, product and participant perspectives in CLIL research. *International CLIL Research Journal, 4*(1), 66–78.

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15*(1), 101–111.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167.

Brown, H., & Bradford, A. (2019). Teaching subject-content through English: CLIL and EMI courses in the Japanese university. In P. Wadden & C. C. Hale (Eds.), *Teaching English at Japanese universities: A new handbook*. New York: Routledge.

Elkington, J. (1997). *Cannibals with forks: The triple bottom line of twentieth century business*. Oxford: Capstone.

Folse, K. S. (2006). *The art of teaching speaking: Research and pedagogy for the ESL/EFL classroom*. Ann Arbor: University of Michigan Press.
Gardner, D. (2000). Self-assessment for autonomous language learners. *Links & Letters (Universitat Autónoma de Barcelona), 7*, 49–60.
Hale, C. C. (2015). Self-assessment as academic community building: A study from a Japanese liberal arts university. *Language Testing in Asia, 5*(1), 1–12.
Hale, C. C., & Wadden, P. (2013). Academic writing in a liberal arts curriculum in Asia: Culture and criteria. *NU ideas, 2*(2), 19–28.
Hallinger, P. (2010). Making education reform happen: Is there an 'Asian' way? *School Leadership and Management, 30*(5), 401–418.
Harris, M. (1997). Self-assessment of language learning in formal settings. *ELT Journal, 51*(1), 12–20.
Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement* (1st ed.). New York: Routledge.
Hilgers, T. L., Hussey, E. L., & Stitt-Bergh, M. (2000). The case for prompted self-assessment in the writing classroom. In J. Smith & K. Yancey (Eds.), *Self-assessment and development in writing: A collaborative inquiry* (pp. 1–24). Cresskill: Hampton Press.
Huang, L. S. (2010). Seeing eye to eye? The academic writing needs of graduate and undergraduate students from students' and instructors' perspectives. *Language Teaching Research, 14*(4), 517–539.
Kim, J. Y. (2011). Dartmouth college commencement: Valedictory to the seniors. http://www.dartmouth.edu/~commence/news/speeches/2011/kim.html. Accessed 15 Jan 2018.
Kozulin, A., & Presseisen, B. Z. (1995). Mediated learning experience and psychological tools: Vygotsky's and Feuerstein's perspectives in a study of student learning. *Educational Psychologist, 30*(2), 67–75.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL—Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol: Multilingual Matters.
Marginson, S. (2011). The Confucian model of higher education in East Asia and Singapore. In S. Kaur, E. Sawir, & Marginson (Eds.), *Higher education in the Asia-Pacific: Strategic responses to globalization* (1st ed., pp. 53–75). New York: Springer.
Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing, 26*(1), 75–100.
McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice, 10*(2), 209–220.
Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge: Cambridge University Press.
Oldfield, K. A., & MacAlpine, J. M. K. (1995). Peer and self-assessment at tertiary level — An experiential report. *Assessment and Evaluation in Higher Education, 20*(1), 125–132.
Patri, M. (2002). The influence of peer feedback on self-and peer-assessment of oral skills. *Language Testing, 19*(2), 109–131.
Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.
Rigoglioso, M. (2007). Research: Corruption causes business inefficiency. *Stanford Graduate School of Business.* https://www.gsb.stanford.edu/insights/research-corruption-causes-business-inefficiency. Accessed 19 Nov 2018.
Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144.

Sandel, M. J. (2007). *The case against perfection: Ethics in the age of genetic engineering*. Cambridge: Harvard University Press.

Sandel, M. J. (2012). *The case against perfection: What's wrong with designer children, bionic athletes, and genetic engineering*. London: Routledge.

Santa Clara University (Producer). (2011, January 24). Five ways to think ethically [Video file]. https://www.youtube.com/watch?v=OTpwUUNepZc&feature=youtu.be. Accessed 21 Oct 2018.

Sullivan, K., & Hall, C. (1997). Introducing students to self-assessment. *Assessment & Evaluation in Higher Education, 22*(3), 289–305.

Wadden, P., Hale, C. C., Rush, E., Punyaratabandhu, D., Kleindl, M., Paterson, R., & Engler, S. (2012). English for liberal arts: Toward a new paradigm for university language teaching. In A. Stewart & N. Sonda (Eds.), *JALT 2011 conference proceedings*. Tokyo: JALT.

Chapter 4
CEFR, CLIL, LOA, and TBLT – Synergising Goals, Methods and Assessment to Optimise Active Student Learning

Claudia Kunschak

4.1 Introduction

Globalisation and internationalisation have greatly impacted educational policy and practice in Japan over the past decades, leading to the introduction of a series of concepts and approaches from different pedagogical cultures that have been tested and modified to meet the local needs. One of the most well-known and widely applied examples is the CEFR-J (Tono 2017), a modified version of the Common European Framework of Reference for Languages (CEFR). The modification takes account of the additional hours required to learn English if the mother tongue is typologically different, and does so by dividing up the lower levels to include more sub-levels (i.e. PreA1, A1.1, A1.2, A1.3). At the same time, interest in CLIL is on the rise with some pioneering programs, especially in the sciences, due to the increase in English-medium programs across both public and private universities in Japan, particularly those supported by special government grants earmarked for internationalisation. CEFR in general and CLIL lend themselves to being integrated in a task-based language teaching (TBLT) environment, as TBLT operationalises the can-do statements in the language classroom and guides learners in how to use language to manipulate content knowledge while helping them to develop their language skills through the handling of content. The final cornerstone of a proposed model that aims to synergise goals, methods, and assessment to optimise active student learning and is currently under development at a College of International Relations at a large-scale private university in western Japan that provides both Japanese and English-medium instruction, is assessment that accompanies a learner from day one, learning-oriented assessment (henceforth LOA). For a discussion of

C. Kunschak (✉)
College of International Relations, Ritsumeikan University, Kyoto, Japan
e-mail: ckr12106@ritsumei.ac.jp

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_4

learning-oriented assessment and its sibling assessment *for* learning (A*f*L), please refer to Chap. 1 of this volume.

The rationale for choosing this set of principles and practices is that it is transparent, scalable, coherent, and practical. The CEFR was adapted to the local concept to provide a curriculum framework that can be communicated to different stakeholders, from students to teachers, parents to institutional decision-makers. CLIL was adopted as the methodology of choice to increase student motivation, link language classes and major courses, and allow students to seamlessly move from integrated language classes into bridge courses, and finally into the English-medium program and/or a study abroad experience. TBLT was chosen to provide students with a practical, holistic, and intellectually stimulating setting that encourages language learning as a means to communicate ideas creatively and accomplish something specific. Finally, LOA was selected as the underlying pedagogical principle that within the process of assessment creates opportunities for students to grow and develop both autonomous and peer-supported learning skills aside from content knowledge and language proficiency.

Referring to the above-named principles, the chapter will address the following aspects from a theoretical and practical perspective: (1) how and why to create a localised framework for CLIL, (2) how to increase articulation across and within programs, (3) how to link curricular goals, methodology and evaluation, (4) how to support the teaching faculty in the implementation of the new framework, and (5) how to ensure that students can become active learners. After laying out the theoretical foundations of this curricular model, the chapter will exemplify the different components and steps by referring to the master grid (modified CEFR scales similar to CEFR-J but tailored to the college's program goals and streamed language curriculum), sample course descriptors, a range of tasks, as well as different sets of rubrics developed to support learning oriented assessment. The chapter will thus provide both a rationale for why and how to apply these principles in an internationalised university setting and step-by-step guidelines for implementation.

The following section will introduce the four pillars of the model, CEFR, CLIL, TBLT, and LOA both from a general perspective and a Japan-related applied perspective. This structure was chosen to demonstrate the parallel developments that were first taking place elsewhere and are now being retraced and expanded in Japan.

4.2 Overview of Concepts and Application in Japan

The **Common European Framework** (CEFR) (Council of Europe 2001a), which has been described in greater detail in Chaps. 1 and 3, shall be discussed here from the point of view of curriculum development and articulation with the goal of maximising active student learning. As mentioned repeatedly in the literature (e.g. North 2007), the CEFR is not a set of test specifications but a framework to develop context specific tasks, descriptors, and assessment criteria. Furthermore, its subtitle 'Learning, Teaching and Assessment' underlines the interconnected areas of

language programs, program development and program evaluation rather than focusing on bands and scores alone. The related European Language Portfolio (ELP) (Council of Europe 2001b) was designed to be used as a model for strengthening learner autonomy and originally, in the case of Europe, documenting the scope of students' achievement of the mother tongue plus two goal (European Council 2002; henceforth MT + 2; Marsh 2002). In other contexts, for example, in the program described in this chapter, it can help students monitor and document their progress in multiple languages: English and another language in the case of Japanese Medium Instruction (JMI) students; English, Japanese, and another language for English Medium Instruction (EMI) students. Just as the core document, the portfolio is flexible and adjustable to local contexts. Teachers as well benefit as they can not only choose appropriate textbooks more easily but also explore how to design their own level specific tasks by referring to the user guides for applying the framework (e.g. Trim 2001; Glaboniat et al. 2003) in addition to any site-specific workshops that need to be offered to support the curricular transition. The most recent addition to the framework, the Companion Volume (Council of Europe 2018) supports students even further by moving away from the native speaker as a referent, introducing plurilingual and pluricultural competence, and updating domains such as online communication. A final strength of the CEFR is its multidimensional expandability across program levels and branches and its portability across institutions and countries (see also Schmidt et al. 2011).

In *Japan*, the CEFR is gaining ground (O'Dwyer et al. 2017), but its spread across regions and school types varies (cf. Runnels 2015). It is mostly known through the CEFR-J, an adaptation of the CEFR to the language learning realities in Japan (Runnels 2014; Tono 2019). The focus of the Japanese version is on the lower levels of proficiency with both a pre-A1 level and a further stratification of A1 compared to the European system. This stratification not only corresponds to the majority of language learners in Japan that can be found in levels A1 and A2 (Negishi et al. 2012) but can also be seen as a precursor to the current companion volume to the CEFR which has also added a degree of diversification at the lower end of the spectrum. While the initial adaptation of the CEFR for Japanese purposes was aimed at English as the first foreign language learned in school, its application has expanded to other languages (Negishi and Tono 2014), not least due to the testing and certification orientedness prevalent in the Japanese educational system. Research into applications of the CEFR and CEFR-J in Japan is available in both English (e.g. Runnels 2015; Hatasa and Watanabe 2017) and Japanese (Watanabe et al. 2011, 2012, 2016) and its applicability to current efforts of internationalisation has bolstered support at high school and university level. For CEFR studies linked to CLIL programs in Japan, see O'Dwyer et al. (2013) and O'Dwyer and deBoer (2015), who focus on a combination of TBLT and different formats of learning-oriented assessment. At the same time, the CEFR has influenced the teaching of Japanese as a Foreign/Second Language in Japan and abroad via the AJE-CEFR project (Association of Japanese Language Teachers in Europe) sponsored by the Japan Foundation (AJE and JF 2005) and the development and publication of the JF Standards for Japanese Language Education (Japan Foundation 2010). A textbook

series *Marugoto* followed and the JLPT proficiency test was adapted to align more closely with the CEFR principles (Shigemori Bučar et al. 2014). As Tono (2019) mentions in his recent article, CEFR and CEFR-J have the potential to be useful to a range of languages beyond those that were originally targeted.

Content and Language Integrated Learning (CLIL), and the various combinations of its components, content, language, integration and learning, have been discussed in great detail in Chap. 1 of this volume. What is relevant in this context is the main focus of implementation and research in Europe on primary and secondary education guided by the MT (mother tongue) + 2 directive (European Council 2002; see also Marsh 2002), whereas outside Europe, particularly in Asia, CLIL is being introduced at the tertiary level to boost economic competitiveness of its graduates (cf. Tsuchiya and Perez Murillo 2015; Yang 2017) often in conjunction with an increase in English-medium instruction (EMI), if we consider EMI the end point of CLIL (cf. Dearden 2017). Another relevant issue related to CLIL is teacher qualification (Mehisto and Lucietto 2011), which becomes increasingly central as CLIL moves up through the educational levels. Teacher qualification here refers to both their qualification in the language and content domain as well as their specific training in CLIL methodology (Shohamy 2012). Finally, the potential for moving across the CLIL continuum with a shifting focus from increasing language proficiency with content-based materials to prioritising content in EMI-mode within one program will be discussed. McLellan (2018) provides another example of how to implement such a gradation in Japan.

Although Content and Language Integrated Learning (CLIL) can still be considered to be in its infancy in *Japan* as far as institutionalization is concerned compared to Europe, or at the 'grassroots level' (see Chap. 1, this volume), it has seen an upsurge in interest over the past several years. Initially linked to its sister discipline of Languages/English for Specific Purposes (LSP/ESP), CLIL has established itself at the tertiary level with textbooks, teacher training and a research agenda (e.g. Cope 2014; Ikeda 2013; Ito 2016; Kashiwagi and Tomecsek 2015; Sasajima 2013). Kusumoto (2018) and O'Dwyer and deBoer (2015) provide evidence and suggestions for the development of critical thinking, learner autonomy, and collaboration, all twenty-first century skills (Partnership for 21st Century Learning 2007). At this point, less activity in this area can be observed at the primary or secondary level (for primary, see, e.g. Yamano 2013; for secondary, Clark 2013) due to the focus on testing discrete language items and receptive skills for the purposes of selectivity in admission to upper-level schools and universities. However, the overall push towards internationalisation may lead to an increase in CLIL at the lower educational levels in due course. This would be the expected washback effect emanating from the ministerial goal of moving to external four-skills evaluation of English for university admission (McCrostie 2017), i.e. the use of four-skills based standardised tests by commercial providers instead of university-developed instruments focusing on receptive skills and discrete point items only. Two complementary views on the opportunities and challenges of CLIL/EMI at the tertiary level are provided by McLellan (2018) and Ng (2017). The 2013 Special Edition of the Asian EFL Journal

on *CLIL in Asian Contexts: Emerging Trends* offers a total of 10 studies focusing on Japan covering a variety of disciplines.

Learning-oriented assessment, and assessment *for* learning (Black et al. 2004), both discussed at great length in the introductory chapter, may include both assessments designed from the outset to have formative function and types of assessment initially designed with summative function in mind. In either case, the overarching goal is to support learning by letting both the teacher and the learner know about the stage of learning the student has reached with respect to the ultimate learning goals of a class while serving to inform subsequent teaching and learning practice. For LOA, Turner and Purpura (2016) propose a framework of seven dimensions that influence the process of learning under the LOA approach with 'learning' being the central one encompassing cognition, feedback and assistance, and self-regulation (see also Chaps. 1 and 5, this volume). In order to fulfil the pedagogical goal of LOA, several dimensions such as feedback *vs* feed forward (Mehisto and Ting 2017), descriptive *vs* evaluative feedback (Tunstall and Gipps 1996), diagnostic *vs* dynamic assessment (Alderson et al. 2015; Lantolf and Poehner 2010), self *vs* peer *vs* teacher-generated evaluation (Andrade 2010; Topping 2010), task *vs* strategy-oriented feedback (Alderson et al. 2014) need to be considered (see Chap. 1, this volume). In the context of CLIL, particular emphasis has to be directed to the integration not only of language and content over the course of a semester but also of different dimensions of assessment whose goal is the promotion of learning over time. Effective and efficient for teachers, timely and transparent for learners, constructive, focused, and interactive can be considered the goal posts in this approach.

Japan can be considered in the experimental stages of LOA as more university teachers are beginning to set their own agenda within the parameters of program goals. They are also choosing their methodologies and materials, and are experimenting with assessment that promotes learning. In other words, LOA is not part of assessment policy. Besides the previously mentioned O'Dwyer et al. (2013) and O'Dwyer and deBoer (2015), Weaver (2012) examines formative assessment in a TBLT context, Saito (2008) investigates peer-assessment, and Wicking (2017) concludes that teachers do have a learning-oriented approach to assessment despite the exam-driven system of Japan's education sector. However, whereas international schools and feeder schools for the top-ranked universities also enjoy a certain leeway in adapting their programs and progressions within the general framework stipulated by the Ministry of Education, teachers at regular schools are under immense pressure to achieve high scores on high-school and university entrance exams for their students (Karlsson 2016; Sasaki 2018). Assessment whose goal is to promote learning is still an option though according to Shimojima and Arimoto (2017) if we look at the capacity of school-based professional learning communities, who in their daily practice can and do integrate the seemingly opposing poles of teaching to the test and assessing to promote learning (see also Davison and Leung 2009, for teacher-based assessment and Inbar 2008, for changes in assessment culture). Given the socio-political context, Japan might thus benefit by considering the cases of Hong Kong (Carless 2010) and Singapore (Leong and Tan 2014),

two high-achieving nations as far as international standardised tests are concerned that have been moving closer to learning-oriented assessment practices.

Task-based language teaching (TBLT) due to its action-orientation, naturally aligns with the CEFR's can-do statements and the principles of CLIL of using language to convey content or using content to develop language. Another fitting parallel, similar to the CLIL continuum (Ball 2009) or Quadrants (Leung and Morton 2016) discussed in Chap. 1 of this volume, resides in the flexibility and adaptability of this approach to various contexts. Choosing a 'weak' or basic form of TBLT is suitable for more traditional learning contexts whereas 'strong' or complex TBLT extends the authenticity of individual tasks to create sustained authenticity by linking them together (cf. Benevides and Valvona 2008; O'Dwyer et al. 2013). In its basic form, TBLT involves a 'real-life' output, some information gap, negotiation of meaning and the linguistic means to achieve the end (Ellis 2003). In the context of CLIL in higher education, a task may be an academic one, a work-related one, a discipline-specific one, an integrated one, to name just a few ways of approaching task design. TBLT is both performance-based and criterion-referenced (González-Lloret and Nielson 2014); it can serve to integrate not only language and content but also academic skills, soft-skills, strategy training, and more. At the same time, by offering such a plethora of possibilities, it requires a strong methodological foundation and high degree of assessment literacy for teachers to implement. One way of addressing this challenge would be by implementing a task-based approach in language teacher education to let teachers experience the methodology themselves (Jackson 2012). Support can also be found in two recent special issues of the TESOL Quarterly (2017) and Language Teaching Research (2016), which focus on TBLT and whose contributions examine the topic from all angles including learner perspectives, teacher education, task design and performance, as well as research methodology. Two contributions, Fukuta (2016) on the effects of task repetition and Yasuda (2017) on combining genre pedagogy and TBLT, demonstrate how this approach has been applied in Japan.

Overall, the case for Task-based Language Teaching in *Japan* follows a similar pattern as the three other pillars previously described. A growing interest in task-based methodology can be observed in the literature and at professional meetings such as the biannual TBLT in Asia conference organised by the TBLT Special Interest Group of the Japanese Association of Language Teachers (JALT). However, the strong focus on traditional teaching and testing structures in primary and secondary school precludes this approach from gaining ground throughout the educational system. Sato (2010), Sybing (2011), and Harris (2018) provide some insight into the debate around the feasibility and desirability of such an approach in the Japanese educational system. Notwithstanding these reservations, some promising studies demonstrate that tasks can be used flexibly at all educational levels in Japan. (See Shintani (2016) on the comparison of traditional PPP (Present, Practice, Produce) and TBLT among young learners, Sasayama and Izumi (2012) on the effects of task complexity on performance at high school level, and Horiba and Fukaya (2012) on the effects of task instructions on text processing and learning at a college level nursing program.) On the other hand, a lack of TBLT-based

textbooks and teacher training contributes to the limited spread of this methodology in Japan with the exception of higher education. Among noteworthy exceptions are the recent publication of *On Task* (Harris and Leeming 2018a, b, c) as well as localised material design in the field of tourism, for example. In support of TBLT, increased student motivation has been mentioned as one of the main benefits of using this approach in Japan (e.g. Anthony 2012). This motivation can be increased by using learner-created content as the basis for the task to be achieved (Lambert et al. 2017). Furthermore, as tasks often are accomplished in groups, they may also lead to the development of collective efficacy (Leeming 2017), a concept ready to be exploited in a social context such as Japan that encourages collaboration.

4.3 Description of Site with a View to Applicable Concepts

The site for experimenting with a combination of the above-mentioned approaches to language teaching, learning and assessment, CEFR, CLIL, LOA, and TBLT, is a College of International Relations at a private large-scale university that belongs both to the Global 30 and Top Global group of institutions (MEXT 2009, 2014). The college offers a range of programs of study, including a Japanese-medium degree with a focus on English in years one and two, an English-medium degree that offers Academic skills training and English support mainly in year one, and a recently unveiled joint-study degree with a mandatory component of study abroad contingent upon reaching a required level of English proficiency. In order to make student achievement comparable and transparent among the different degree options so that students can switch between them if needed or so desired, the CEFR was chosen to provide the framework for assessing language proficiency across the board. Some form of combining language and content has long been the methodology of choice due to the possibility of moving from a soft CLIL or CBLT in year one to more emphasis on content in year two and providing language support for content classes in the EMI program. In order to develop learner autonomy and life-long learning beyond the college classroom, LOA is being encouraged in its various manifestations across all levels. Finally, to prepare students for study abroad, moving into the EMI program, or entering the job market, TBLT is being used to provide students with concrete deliverables that they can accomplish only by combining the various categories of skills, including the so-called twenty-first century skills, that they are asked to develop in their classes. The following paragraphs will flesh out these four pillars in more detail both to illustrate their independent contribution to the program and to demonstrate the synergies that are created by applying them jointly.

4.4 Step-by-Step Development and Integration

Before delving into the process of development, it should be noted that one of the underlying motivations for the curricular overhaul was the attempt to bridge the request for Assessment of Learning (A*o*L) from the institutional stakeholders and the culture of assessment *for* learning (A*f*L) practiced by the English language faculty (see Chap. 1, this volume, for a discussion of the underlying concepts). As is the case in most institutions, test scores matter but will only tell half the story if not aligned with curricular goals (contextual validity), assessment purpose (assessment use) or internal validity (Weir 2005; Bachman and Palmer 2010; Mislevy et al. 2003). That is, test scores that are based on discrete points and receptive skills will not be able to reflect gains in integrated skills or argumentation. In the programme under discussion, contextual validity would require both a focus on the content of international relations and the academic tasks to be accomplished in various classes and seminars. Assessment use would include both proof of mastery of program goals and the more crucial purpose of promoting learning in the process of reaching those goals. Finally, internal validity would require the assessment result to be explainable by the input, i.e. the teaching and (collaborative) learning taking place in the classroom. Based on these requirements, traditional standardised testing would not reflect the achievement of program goals.

An alternative way of demonstrating mastery is based on level-specific can-do descriptors (CEFR) that determine the extent of achievement of a task (TBLT) related to a domain of study (CLIL) and potentially created in a dynamic, interactive process (LOA). The model is, furthermore, based on the elements of curriculum design proposed by Brown (1995), which include needs analysis, objectives, testing (assessment), materials, teaching, and evaluation. The model aims at integrating objectives (CEFR), assessment (LOA), materials (CLIL), and teaching (TBLT), although there is of course some overlap among the four pillars as they apply to more than one component in Brown's model, which makes them not only mutually compatible but complementary. This further implies that assessment is not an externally imposed procedure but a core component of teaching and learning that informs and empowers students as well as teachers to modify their practice according to perceived needs (see also Chap. 1, this volume). In the following paragraphs, the role of the various components as well as the steps in developing program specific standards, documents and instruments will be outlined. As the documents related to the model (Fig. 4.1) are of a proprietary nature, the purpose of the following paragraphs is to provide guidelines for independently developing a model to be adapted according to institutional needs rather than present a ready-made set of templates.

CLIL, as can be seen from Chap. 1 of this volume, is an umbrella term and a bundle of principles that can be tailored to fit individual program needs. Before launching into a curriculum reform, two main questions need to be addressed by the faculty in charge: *how and why to create a localised framework for CLIL***.** It may seem easier to just take a readymade structure such as the CEFR and fit all local course offerings into the framework. However, the CEFR was created based on

Fig. 4.1 Step-by-step development of curricular model

language learning conditions in Europe. It was also specifically conceptualised as a framework that could and should be adapted to local needs. Hence, the development of CEFR-J that incorporates a more fine-grained distinction at lower levels and serves as a guideline for vocabulary to be included at different levels in the school system (Negishi and Tono 2014). For university settings, a more academically-oriented, content-sensitive version is needed to cater to the varying degrees of language and content, different purposes of different programs, as well as different student populations at different types of institutions. More importantly, university administrators and teachers need to strike a balance between satisfying student and parent demands and providing an internationally recognised framework of reference. Finally, while core university administrative units, such as language centres may be happy to simply adopt a validated framework, individual colleges, faculties, and programs and their instructors may resist a structure that does not reflect their curricular realities. It is, thus, essential to clearly spell out the rationale and goals for any localisation to all stakeholders before launching into the reform process. In this particular case, demands for program results expressed in scores of internationally recognised criterion-referenced tests, also used for norm-referenced purposes such as streaming and scholarship allocation and thus doubly valuable to stakeholders, had to be reconciled with a program that favoured a task-based approach and included varying degrees of content with a resulting lack of test-specific preparation.

After collecting feedback from all parties involved regarding the need for and direction of this specific localisation of a framework for CLIL, the current *program structure, program goals, and demands for proof of program effectiveness need to be considered*. Regarding program structure, different modalities of CLIL, often moving from a language-heavy first year program to a more content-focused second-year and beyond, need to be mapped out as described in the following paragraph. At the same time, language levels, such as spread of proficiency among yearly intake or mixed-level vs. streamed classes, need to be considered. Differentiating further, skill-specific needs such as a focus on listening/speaking for aviation English but reading/writing for the legal professions, as well as EMI modalities and their place in the spectrum need to be examined (cf. McLellan 2018). Program goals may include international competitiveness on the job-market, ability

to participate in exchange programs, or mandatory semesters abroad in the case of joint-degree programs, or further education in their field of study, as well as preparation to fully benefit from a globalised society in political, economic, and social terms. At the same time, current practices of transparency and accountability across different university stakeholders regarding assessment use need to be reviewed and any planned new system will need to be vetted not only among the faculty concerned but also university administrators. That is to say, a variety of assessment uses, from admission to placement to achievement, with their respective divergent assessment modalities need to be reviewed for their fit for purpose, reconsidered from an educational, budgetary and human resource perspective, and hybridised to reflect more closely and consistently what the program is offering and expecting students to achieve.

Based on the structure and goals of the program, demands for alignment with an internationally recognised system, and the need for an open system that could be adapted to local circumstances, the CEFR suggested itself as the underlying structure for language proficiency development and measurement across the different program components. These components comprise CBLT or soft CLIL in year one of the Japanese-medium program (Quadrant 3), a sliding scale towards a harder form of CLIL or scaffolded content courses in year two of the Japanese language program (Quadrant 2), as well as Academic English or Academic skills in year one (Quadrant 1) of the English-medium and joint-degree programs (Quadrant 4). This classification follows Leung and Morton's (2016, p. 220) matrix (see Chap. 1, this volume), which can help visualise how to build or clarify a program according to the different language and content needs of its student population as they move towards graduation. By plotting the different streamed levels and course offerings across the main programs of study, students can move vertically across through different levels and horizontally through different programs. In order to adapt the CEFR in its current version, that is, the global descriptors for levels A2 through to C1, which correspond to the range of levels found among the student population, covering the skills of listening, reading, interactive speaking, productive speaking, and writing, the original CEFR can-do statements were compared with the course descriptors of the different programs and modified mainly in the tasks that students were expected to be able to perform in the different skills at the various levels. Before fleshing out the skill-specific requirements, general program goals for each proficiency level were stipulated based on student trajectories within and beyond the program of studies. This master grid can then serve as a framework for developing course descriptors and attainment objectives, task descriptors and rubrics, as well as test specifications, all essential elements to support assessment whose goal is the promotion of learning as a task shared between teachers and students by providing perspective, feedback, and opportunities for self-directed and peer-supported learning.

In order to draft this master grid, overall program goals for students at different proficiency levels need to be established (e.g. to what extent they are expected to be able to study abroad, work abroad or work in an international context in their home country). Next, for each skill, global descriptors need to be developed (e.g. what

kind of genres they are able to understand or produce, from lectures or journal articles to workplace instructions or email messages, what kind of interactions they are able to engage in, from moderating discussions or critiquing an argument to managing basic requests for information or expressing an opinion). These descriptors need to reflect both program requirements and CEFR level benchmarks, that is, what users at a specific level are able to accomplish regarding domain, function, and complexity.

Besides the localised master grid that can be used to build a multi-dimensional framework to increase articulation within and across programs, the design of the content ladder and its branches will define the coherence of a CLIL program. Without such a blueprint, students may end up missing out or duplicating their program of study. Depending on the type of program, credit hours dedicated to content, language or CLIL subjects, as well as student proficiency level, program goals, and attainment levels in content need to be clarified. This will invariably lead to negotiations beyond the language teaching faculty members to include not only those content faculty members who teach through the medium of English but ideally also those who deliver their classes in the home language. While some overlap between content-based language classes and content delivered in the home language may be helpful, a clear progression needs to be delineated between first- and second-year content-based language classes and scaffolded content courses on one hand and academic skills classes and content classes in an EMI program on the other. Students need to be carefully supported through this progression in order to make sure that EMI classes not only serve inbound international students but can be fully utilised by domestic students to prepare themselves for study abroad (McLellan 2018). Based on content progression, course descriptors can be adjusted to define the degree and depth of content focus, such as moving from an introductory integrated skills course in the first semester to a theme-based course in the second semester to a discipline-specific content course with language support and research component in year two. This general trajectory can be modified according to proficiency level to include a variable percentage dedicated to content vs. language-focused activities, e.g. the amount of content covered vs. the number of linguistic support activities such as paragraph development, paraphrasing or integrating sources. In order to increase/decrease the cognitive and linguistic load, the extent and nature of out-of-class assignments such as simple surveys or more complex literature reviews can be adjusted as well.

This may serve as a segue into the next level of consideration, *how to link curricular goals, methodology, and evaluation* when implementing a progressive CLIL program. Task-based (language) teaching can span the content-language continuum and be easily modified for a variety of disciplinary/genre requirements (Yasuda 2017) and adjusted to the language proficiency level. TBLT fits with the CEFR grid as can-do statements refer to a task that can be accomplished with the required linguistic tools. When adding in content, to give an example from the area of international relations, policies may be compared, critically evaluated or used as the basis for designing an implementation plan. Depending on the desired student profile upon graduation, or the required level of content knowledge and language

competence for passing a specific course, target tasks for each class such as a model UN debate or an oral presentation and write-up of survey results as well as task components to be developed throughout the semester should ideally be cross-checked across different program elements to maximise student learning. That is to say, goals, tasks, and assessment modalities should be gradated both across CBLT-Academic Skills-CLIL-EMI courses (i.e. depending on the quadrant in Leung and Morton 2016) and across streamed proficiency levels within each course type. For example, if students in the EMI program are expected to be employable in international organisations or global companies in their home countries, they will have to develop a different profile from those in the JMI program who are aiming at an entry-level job with an internationally oriented Japanese company. If we define curricular goals by what students are able to accomplish upon completing a course, and choose specific tasks that illustrate the desired competence, methodology, and evaluation must necessarily follow. TBLT and its sister approaches, problem-based learning and project-based learning, lend themselves to interactive classrooms, the development of learner autonomy, and collaborative work. Students work together to accomplish a specific task such as reporting on an innovative environmental approach, solve a real-world problem such as waste management or reduction, or create a meaningful project such as raising environmental awareness through social interaction with community members and learn to take on responsibility for their work. Thus, besides content and language gains, students also acquire the necessary twenty-first century skills (Kusumoto 2018) supplemented by computer-based applications as needed. Regarding assessment, both CEFR and TBLT require clear task descriptions for students to follow as well as detailed rubrics for teachers to evaluate level of achievement and for students to identify strengths and weaknesses to ensure that assessment covers both, learning accomplished and learning in progress. These rubrics can be adjusted to program needs by including one or more content-relevant categories or weighting content and language related categories differently according to type of class or level of proficiency of students (see, e.g. Chaps. 2 and 3, this volume). Obviously, just like task components will be developed during the course of the semester, different types of quizzes including micro tasks such matching of concepts and definitions, paraphrasing set expressions or choosing appropriate word forms will prepare students for the final task (cf. Christodoulou 2017).

When creating rubrics (e.g. Stevens and Levi 2005), similar to the process of creating the master-grid, overall goals for a task or performance and their assessment need to be established (e.g. to assess the draft of an academic paper with particular focus on citations and references and the option to resubmit, or to evaluate the task accomplishment of a customer service encounter with a particular emphasis on sociolinguistic appropriateness). Next comes the question of whether to create a holistic or analytical rubric. Keeping in mind LOA, the fourth pillar of the model, an analytical rubric provides better opportunities for feedback that targets different subdomains, which is why this type of rubric was chosen. In the following step, criteria (such as content and organization) and levels of performance (e.g. exceeds, fulfils, approaches, fails) will be determined and then, descriptors illustrating those

levels will be formulated (e.g. uses scholarly sources or displays audience focus). As with the master-grid, descriptors will have to fall within the range of proficiency outlined by the global and detailed descriptors of the CEFR.

As can be seen from the above paragraphs, CLIL, CEFR, and TBLT are very easily integrated into a whole that provides a structure but also allows for room to adjust and modify (see O'Dwyer et al. (2013) for a combination of CEFR, TBLT, and LOA). When implementing any combination of these approaches, *teacher background and familiarity with the concepts as well as teacher dedication* (full-time, contract, part-time) *and teacher preference or self-efficacy need to be considered*. Regarding CLIL, some faculty members may have a content background but require some support on how to structure tasks; other teachers come from a teacher education program and may not feel comfortable with a high degree of content in their classes. More importantly, most teachers struggle with a significant workload and may be reluctant or unable to attend workshops or apply a new grading scheme. In order to work with these givens, including teachers beyond the core faculty members in the development process can ensure that the outcome is feasible and practical and will more easily be adopted by a larger number of teachers as the total number of teachers offering courses at any one faculty can easily be 10 times the number of core faculty in charge of program development. In addition, similar to the principles of the CEFR, which is presented as a "flexible tool" that "does not set out to tell practitioners what to do, or how to do it" (Council of Europe n.d.), the master-grid and related documents were created to provide an orientation for teachers to position themselves and their students within the overall system. Adjustments based on a particular group of students will always be necessary and remain in the realm of a teacher's professional practice. At the same time, teachers need a clear view of the overall program structure and progression. Only then can they be expected to actively contribute to a culture of assessment whose goal is to promote learning that is based on providing feed up, feedback and feed forward (Hattie and Timperley 2007). With a view to increasing consistency across classes and facilitating mentoring for teachers new to the program, establishing contact among teachers in charge of a sequence of classes or similar parallel programs can help jump-start collaboration and support teachers from the beginning. This has to be launched at the time of planning, after class assignments have been communicated, i.e. well before the syllabus is due approximately 2–3 months before the start of the new academic year. Suggested materials for specific classes, be they textbooks or other resources, can also raise consistency across sections and coherence across levels and types of classes. Finally, guidelines on expected workload for students need to be communicated to avoid teacher and student overload by trying to do both, a traditional language focused program on top of an integrated CLIL/TBLT syllabus.

From the perspective of *active learning for students*, all three pillars, CEFR, CLIL, and TBLT, contribute to the goal. CLIL in its combination of content and language encourages discovery of content matter while providing the tools to communicate about the latter while TBLT engages students in meaningful manipulation of language and content with a specific outcome, often combining in- and out-of-class work as well as pair and group activities. The CEFR with its empowering

focus on can-do as opposed to "got xyz wrong" gives students a goal to strive for as well as a sense of accomplishment, and if used for self-assessment purposes, an instrument to monitor their own learning. Beyond the use of the CEFR, other elements of learner-oriented assessment (LOA) as assessment whose goal is to promote learning (see also Chap. 1, this volume) can support a CLIL/TBLT-based program. Integrating a term-based assessment cycle with diagnostic, mid, and final assessment shows students where they are, how much they have improved, and what still needs to be worked on, which puts them in the driver's seat and allows teachers to consider both progress and product and adapt their teaching plan accordingly. Combining peer, tutor, computer-based, teacher, and self-evaluation allows for multiple perspectives and roles as well as reflection and revision of the main task or project. Scaffolding tasks or building task-networks such as outlines, drafts, and bibliographies, oral, written, and multi-media products, individual, pair, and group components, allows students to build their repertoire, receive timely and varied feedback and apply their skills in various ways. By integrating content and language in assessment, students can be subtly guided towards a more holistic approach to learning or deep learning (Coyle 2018) with a view to applying their skills rather than studying vocabulary for a test or memorising concepts by heart. If so desired, this network of steps, components, and perspectives can also be combined into a portfolio or e-portfolio for an added focus on reflection, development and achievement, connection to the respective discipline and show-case for what students can do (Lam 2017).

It can thus be argued that CEFR, CLIL, TBLT, and LOA should not be considered as separate features of the learning environment but interlocking systems that work together to support student learning, increase teacher efficacy, and provide a coherent and transparent map for all stakeholders (Fig. 4.2). At this point, it should also be noted that the four-dimensional framework outlined above as well as the

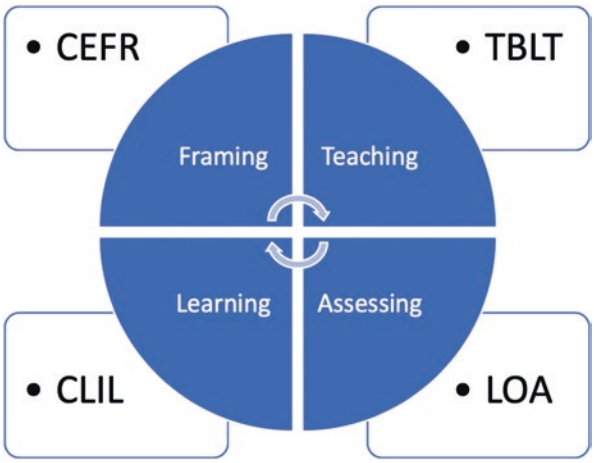


Fig. 4.2 Interlocking four-pillar model

following references to both challenges and possible solutions need to be viewed as a dynamic process embedded in a structure tailored to the specific needs of a local context. That is why, the reader will not find a ready-made model to replicate at their institution but hopefully glean some pointers for developing their own original framework and approach. This can best be accomplished by starting with a needs analysis (e.g. study abroad options, post-graduation trajectories), facilitating discussions among language and content faculty about desired learning outcomes, and setting up a formative process of program development, monitoring, and improvement that is participatory, qualitative and inquiry-based (cf. Norris 2016). Some stakeholders will undoubtedly be asking for effectiveness and accountability, and their claims will need to be considered. One way that this can be addressed is by suggesting a portfolio of assessment measures. These measures may include standardised test scores but also evidence of achievement of learning outcomes such as papers or videos or other authentic samples of performance tied to a specific rubric of learning objectives (e.g. a term paper on a policy issue or a group presentation on a mini-research project). Evidence of progress such as multiple drafts, peer review sheets, and reflective comments or diagnostic, mid-term, and final in-class timed writings could also be included. Crucially, and complementary to external evaluations that may indicate some weak points for improvement but not always a clear remedy appropriate for the context, evaluations of innovative programs need to be cyclical like action research, moving from planning to implementation, assessment to reflection and on to the next round. In the following paragraphs, some of the main challenges in the implementation of the framework described above will thus be discussed and suggestions for overcoming these challenges will be outlined.

4.5 Challenges in the Implementation of the Design

While all four elements of the program seem to fall into place to create a coherent whole, each of them can and has been seen from a critical angle. The CEFR, although adapted and adopted in many educational settings in Japan, has long been doubted or rejected for considering language without its cultural dimension and not being applicable to ESP or CLIL due to its focus on language only. Both reservations can be addressed by considering the CEFR as the framework that it purports to be, not an exhaustive ready-made system but a building set that can be linked up with other existing structures. Regarding the cultural dimension, the newly published Companion Volume (Council of Europe 2018) includes descriptors for plurilingual/pluricultural competence, which can be integrated into the learning goals of any class. Along those lines, mediation is another central concept in the new approach to framing language learning that can be utilised to add a translingual/transcultural aspect to the framework (MLA 2007). Initial case studies supporting these concepts have been conducted and can serve as a blueprint for localisation (e.g. Berceruelo 2018). Similarly, content features or discipline-specific genre dimensions and language needs can be defined for each level, skill, and task and

used to anchor teaching, learning, and assessment by dual focus experts in the content and language domains. Such an adaptation involves considerable curriculum planning efforts but may be the only way to ensure the development of appropriate competencies among the learners.

This naturally leads the discussion towards related concerns about the implementation of CLIL. Two main long-standing issues with this approach are teacher background and available materials/material design. As has been discussed in the literature (Pavon and Ellison 2013), depending on the focus of CLIL classes, a content or a language teacher or, if workload and funding permit, a combination of both may be in charge of instruction. One way to supplement teacher training in this area is to include it into the professional development requirement for new hires at Japanese universities for example (Ito 2016). Another pathway that can be taken is to include a dual focus into the hiring criteria, that is, either a content expert with a language teaching certificate or a language teaching expert whose area of research interest overlaps with the content domain of the college or faculty, a process used at this particular college (cf. McLellan 2018). In any case, the teacher in charge has to be qualified and feel confident to tackle the combination of content and language. With respect to materials, depending on the language proficiency level of students, appropriate GSCE, O-level, or A-level textbooks used in English-speaking countries may be chosen and supplemented with localised content. Alternatively, a college may decide to create a searchable database with relevant articles or other authentic documents that all instructors have access to. Sometimes, if student numbers make it feasible, a specific textbook based on a thorough needs analysis may be tailor-made to meet this demand. All in all, it can be said that much still depends on teacher grassroots engagement which also implies teacher professionalisation, which in turn requires a suitable working environment with pedagogical support and adequate release time to build up the repository.

Moving on to the next pillar, TBLT, some similar concerns arise regarding the availability of textbooks. While some tasks may be integrated into mainstream textbooks, purely task-based textbooks are far and few between. While a pioneering work, *Widgets* (Benevides and Valvona 2008), was published by Pearson in Hong Kong, a recent example compiled by TBLT teachers/researchers based in Japan is the previously mentioned *On Task* (Harris and Leeming 2018a, b, c). One of the main underlying reasons for this dearth of teaching materials lies in the continuation of learner progression along notional/functional lines or a combination of thematic area and related vocabulary plus necessary structures according to level of proficiency, as well as the theoretical/experimental focus in TBLT (Benevides 2016; as cited in Harris 2018). This may not only lead to frustration among learners who need to rehash similar texts and exercises over time but often prevents a meaningful and coherent output-based development of language skills. Once more, teachers are called upon to design tasks that constitute a meaningful link between the language of the textbook and the real-world tasks learners can be expected to be able to perform. Another repeated point of criticism regarding TBLT concerns the priority given to task completion over the mastery of particular language features, such as accepting a paraphrase if the vocabulary item does not come to mind or indulging

faulty grammatical structures as long as no misunderstanding is caused. This misperception, also addressed in Ellis (2009) can be corrected by applying a similar perspective as with the CLIL continuum. It is up to the institution or alternatively the teacher in charge to define the criteria for successful completion of a task. Aviation English may serve as a case in point where clear and audible pronunciation is a prerequisite.

Regarding LOA, few pedagogues will argue with the main purpose of the approach, that is, to make assessment meaningful for the learner, a learning experience in and of itself, and one that has a beneficial effect with a view to further learning. Some concerns that have been voiced in this respect relate to teacher literacy in assessment (Fulcher 2012; Vogt and Tsagari 2014), teacher workload, and the perceived need for a standardised assessment structure both within an institution and in relation to external comparison. On one hand, not all teachers of language have language teaching credentials, some may be linguists, literary scholars or cultural studies experts. On the other, even among those that have majored in language teaching, the majority will have taken a class on assessment but not be necessarily familiar with alternative approaches to the extent of being able to design a sequence of task-based assessments that provide reliable data for high-stakes decision-making. This can be partly addressed by including a workshop in one of the orientation sessions and providing guidelines and samples to teachers but may not be fully realised without providing ongoing support. Another issue concerns teacher workload, as LOA tends to be qualitative in nature (although, as is mentioned in Chap. 1, this volume, summative evaluation also has a place in the toolkit of LOA). That implies more time investment on the part of the teacher, e.g. in the case of multiple drafts, which may be unmanageable with large class sizes. One way to address this challenge would be the planning of various sources of feedback, such as an online platform, peer review, and self-evaluation. Spot-checking, asking students to highlight and/or explain changes, as well as basing a final task on previously submitted ones may aid in reducing the workload. Finally, while most institutions allow for a margin of effort and achievement in the final grading criteria, due to scholarship issues or access to upper-level classes, norm-referenced assessment may still be the ultimate mode of identifying and evaluating students' levels of competency. Streaming can be implemented to alleviate this potentially demotivating set-up but only a conscious effort of the teacher to integrate LOA into their classes will bring maximum benefit for all learners.

4.6 Moving Forward

Having put together the main building blocks for a student-centred experience in a CLIL environment, some further breaking-down of the process, updating of the goal posts and criteria and support for stakeholders may be needed. Regarding the choice of the CEFR as the basis for curriculum planning and evaluation, the new Companion Volume (Council of Europe 2018) and its shift towards, receptive, productive,

interactive, and mediational skills, as well as the additional online communication and plurilingual/pluricultural dimensions need to be incorporated into the master-grid. These additions need to be reflected in the rubrics available per each task/skill/level, the currently available set of rubrics representing a first step towards localising benchmarks. Mediation skills, to give just one example, are already being developed in group projects and group presentations, are culturally accessible to learners, but require the development of a more fine-grained localised scale for evaluation keeping intercultural pragmatics in mind. As a next step, a simplified version to be used by learners in their self- and peer-review activities is currently under construction. Compared to the teacher set, which covers all criteria in great detail and from a pedagogical standpoint (e.g. denoting the level of guidance a learner requires), learner rubrics provide a depiction of abilities in broad strokes using simpler language and less detail. This addendum inspired by the ELP aims at promoting active learning, learner autonomy, and transferable skills.

In fine-tuning the overarching CLIL approach, continuous cooperation with content area faculty to avoid overlap and ensure provision of the basics is essential even without team-teaching. While not all overlap is harmful or even avoidable, the basic principles of complementarity and progression should be taken into consideration. A more structured network of support for contract and part-time teachers without placing an undue burden of workload or restricting academic freedom would be desirable. Such a framework entails some challenges in implementation, however, due to the pattern of teaching assignment that spreads teachers out across colleges, campuses, or even universities. Although the overall framework is provided through the master-grid and the set of rubrics and CLIL levels and degrees are present in course descriptors and handbooks, the actual implementation of the program guidelines, particularly in regard to a TBLT approach, is largely left to teachers. This is due to the composition of the teaching staff as well as institutional culture. Unless or until teachers have taught at all levels and in all areas of the program, lateral complementarity and vertical progression may need to be made more transparent. This can be done by providing a pathway document mapping out the different programs and requirements as well as the time and structure for collaboration and exchange among teachers assigned to the same quadrant of the program. Meanwhile, a combination of all staff, small group, and individual meetings serve as the support structure for teachers.

As regards LOA, teachers may either not have the necessary assessment literacy or the required time for designing such a sequence independently for their classes. If an institution is committed to providing a student-centred learning environment, support for teachers to engage in LOA is essential. Assigning 1 h less of teaching time on a rotating basis for teachers to develop a classroom research project may be funds well spent. A system of monthly lunchtime round-tables that address above named to-do assignments in order to facilitate the desired synergy among the four pillars supporting student learning may also help to make the goal more tangible.

As mentioned in the introduction, the proposed model is still being fine-tuned as part of the current curriculum reform. This chapter has thus aimed to provide the conceptual background of the model, its rationale, structure, and synergistic

potential, while laying out the concrete steps taken by a program to formulate and implement the various components. As a framework, it is meant to be applied to varying degrees, i.e. adapted by different stakeholders based on their type of class, student cohort and professional judgment. It is at the same time a common resource and a flexible tool, a means of articulation, communication, and demonstration, a point of reference for everyone involved in the provision of education that is integrated, purposeful, clearly defined and supportive of overall program goals.

4.7 CLIL and Assessment—Coda

As with any kind of teaching approach, assessment needs to support the main mission of the program, which here is to develop both language skills and content knowledge among learners. Obviously, many related concepts, notably critical thinking and collaboration to name just two of the so-called twenty-first century skills (Partnership for 21st Century Learning 2007), could be integrated into the system, and have been in fact included into descriptors and objectives. Those can also be subsumed under the 'I' of integration, beyond language and content proper. TBLT, on the other hand, can be regarded as the bridge that intertwines the two opposing poles of the CLIL continuum or links up the four quadrants of the matrix and the CEFR could be seen as the trellis that props up progression across the levels. However, as the acronym so aptly states, this is about C-L-I-Learning, thus it is high time that our focus on assessment shift to LOA or assessment whose goal is promotion of learning. LOA needs to be conceptualised as an integral part of CLIL when content is learned through the medium of a language or vice versa, with LOA itself becoming part of the equation. It is the ideal platform for combining feed-back, feed-up, and feed-forward (Hattie and Timperley 2007), accompanying learners and teachers by helping them plan, evaluate and revise their respective work. In addressing both content and language as well as their interrelatedness, LOA has the potential to transform assessment into a multi-dimensional, multi-directional, dynamic, and negotiable process. In fact, CLIL may well be considered the model context for understanding and demonstrating LOA in action that could inform LOA theory-building and application in other contexts. In short, it is a flexible, interactive, transferable pedagogical approach that can be considered a valuable addition to every teacher's toolkit.

Further Reading

Bygate, M. (Ed.). (2015). *Domains and directions in the development of TBLT: A decade of plenaries from the international conference*. Amsterdam: John Benjamins Publishing Company.

Byram, M., & Parmenter, L. (2012). *The common European framework of reference: The globalisation of language education policy*. Bristol: Multilingual Matters.

Carless, D., Bridges, S., Chan, C., & Glofcheski, R. (Eds.). (2017). *Scaling up assessment for learning in higher education*. Singapore: Springer Nature.
Valcke, J., & Wilkinson, R. (Eds.). (2017). *Integrating content and language in higher education: Perspectives on professional practice*. Frankfurt: Peter Lang.

References

AJE & JF. (2005). *Learning and teaching Japanese language in Europe and the common European framework of reference for languages*. Tokyo: Japan Foundation. http://www.jpf.go.jp/j/publish/japanese/euro/. Accessed 25 Mar 2019.
Alderson, J. C., Haapakangas, E.-L., Huhta, A., Nieminen, L., & Ullakonoja, R. (2014). *The diagnosis of reading in a second or foreign language*. New York: Routledge.
Alderson, J. C., Brunfaut, T., & Harding, L. (2015). Towards a theory of diagnosis in second and foreign language assessment: Insights from professional practice across diverse fields. *Applied Linguistics, 36*(2), 236–260. https://doi.org/10.1093/applin/amt046.
Andrade, H. L. (2010). Students as the definite source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 344–351). New York: Routledge.
Anthony, G. C. (2012). Task-based language teaching in a low-proficiency Japanese university context. *Hachinohe University Journal, 45*, 41–50. http://jairo.nii.ac.jp/0305/00000118/en. Accessed 25 Mar 2019.
Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: OUP.
Ball, P. (2009). Does CLIL work? In D. Hill & P. Alan (Eds.), *The best of both worlds?: International perspectives on CLIL* (pp. 32–43). Norwich: Norwich Institute for Language Education.
Benevides, M. (2016, June). *Widgets: A TBLT textbook postmortem*. Plenary paper presented at the third TBLT in Asia conference, Kyoto: Ryukoku University.
Benevides, M., & Valvona, C. (2008). *Widgets: A task-based course in practical English*. Hong Kong: Pearson Education Asia Ltd.
Berceruelo, M. T. (2018, July). *Piloting CEFR companion volume descriptors.* Paper presented at the launching conference for the CEFR companion volume with new descriptors, Strasbourg, pp. 16–17.
Black, P. J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Assessment for learning: Putting it into practice*. Berkshire: McGraw-Hill Education.
Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle.
Carless, D. (2010). Classroom assessment in the Hong Kong policy context. In B. McGaw, E. Baker, & P. Peterson (Eds.), *International Encyclopedia of education* (3rd ed., pp. 438–442). Oxford: Elsevier.
Christodoulou, D. (2017). *The future of assessment for learning*. Oxford: Oxford University Press.
Clark, G. (2013). Snapshot of a lower secondary CLIL program in Japan. *Asian EFL Journal, 15*(4), 383–394.
Cope, T. (2014). *Designing supplementary CLIL materials for a MEXT-authorised English textbook in Japan.* https://www.researchgate.net/publication/262643160. Accessed 25 Mar 2019.
Council of Europe. (2001a). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Modern Languages Division/Council of Europe.
Council of Europe. (2001b). *European Language Portfolio (ELP)*. www.coe.int/portfolio. Accessed 25 Mar 2019.
Council of Europe. (2018). *Common European framework of reference for languages: Learning, teaching, assessment*. Companion volume with new descriptors. Provisional edition. Language

Policy Programme. Education Policy Division. Education Department. Council of Europe. www.coe.int/lang-cefr. Accessed 19 Mar 2019.
Council of Europe. (n.d.). *Use of the CEFR*. https://www.coe.int/en/web/common-european-framework-reference-languages/uses-and-objectives. Accessed 19 Mar 2019.
Coyle, D. (2018). *Designing 'successful' principles and practices for deep learning in CLIL contexts*. https://www.ehb.swiss/bili-tagungconference-bili-documentation. Accessed 15 Jan 2020.
Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.
Dearden, J. (2017, February 2). *EMI (and CLIL) – A growing global trend.* https://oupeltglobalblog.com/2017/02/02/emi-and-clil-a-growing-global-trend/. Accessed 25 Mar 2019.
Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.
Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics, 19*(3), 221–246. https://doi.org/10.1111/j.1473-4192.2009.00231.x.
European Council. (2002). *Resolution on improving and diversifying language learning and teaching within the education systems of the European Union*. Barcelona: EUP.
Fukuta, J. (2016). Effects of task repetition on learners' attention orientation in L2 oral production. *Language Teaching Research, 20*(3), 321–340. https://doi.org/10.1177/1362168815570142.
Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041.
Glaboniat, M., Müller, M., Rusch, P., Schmitz, H., & Wertenschlag, L. (2003). *Profile Deutsch: Lernzielbestimmungen, Kannbeschreibungen, kommunikative Mittel, Niveau A1, A2, B1, B2 [CD-ROM mit Begleitbuch]*. Berlin: Langenscheidt.
González-Lloret, M., & Nielson, K. B. (2014). Evaluating TBLT: The case of a task-based Spanish program. *Language Teaching Research, 19*(5), 525–549. https://doi.org/10.1177/1362168814541745.
Harris, J. (2018). Responding to the critics: Implementation of TBLT in Japan. *Indonesian Journal of Applied Linguistics, 8*(1), 139–148. https://doi.org/10.17509/ijal.v8i1.11473.
Harris, J., & Leeming, P. I. (2018a). *OnTask 3, English language teaching textbook upper intermediate level*. Tokyo: Abax.
Harris, J., & Leeming, P. I. (2018b). *OnTask 2, English language teaching textbook lower intermediate level*. Tokyo: Abax.
Harris, J., & Leeming, P. I. (2018c). *OnTask 1, English language teaching textbook beginner level*. Tokyo: Abax.
Hatasa, Y., & Watanabe, T. (2017). Japanese as a second language assessment in Japan: Current issues and future directions. *Language Assessment Quarterly, 14*(3), 192–212. https://doi.org/10.1080/15434303.2017.1351565.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487.
Horiba, Y., & Fukaya, K. (2012). Effects of task instructions on text processing and learning in a Japanese college nursing setting. In C. A. Coombe & A. Shehadeh (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 89–107). Amsterdam: John Benjamins.
Ikeda, M. (2013). Does CLIL work for Japanese secondary school students? *International CLIL Research Journal, 2*(1), 31–43.
Inbar, O. (2008). Language assessment culture. In N. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 2418–2432). Boston: Springer.
Ito, M. (2016, July 3). *CLIL teacher education: A case at a Japanese university.* Paper delivered at the CLIL special symposium, Tokyo.
Jackson, D. O. (2012). Task-based language teacher education in an undergraduate program in Japan. In C. A. Coombe & A. Shehadeh (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 267–285). Amsterdam: John Benjamins.

Japan Foundation. (2010). *JF standard for Japanese language education* (2nd ed.). Tokyo: Japan Foundation. https://jfstandard.jp/pdf/jfs2010_all_en.pdf. Accessed 25 Mar 2019.

Karlsson, P. (2016, June 8). Is the Eiken doing Japan's English learners more harm than good? *The Japan Times.* https://www.japantimes.co.jp/community/2016/06/08/issues/eiken-japans-english-learners-harm-good/#.W1902bo4DIU. Accessed 25 Mar 2019.

Kashiwagi, K., & Tomecsek, J. (2015). How CLIL classes exert a positive influence on teaching style in student centered language learning through overseas teacher training in Sweden and Finland. *Procedia—Social and Behavioral Sciences, 173*, 79–84.

Kusumoto, Y. (2018). Enhancing critical thinking through active learning. *Language Learning in Higher Education, 8*(1), 45–63.

Lam, R. (2017). Taking stock of portfolio scholarship: From research to practice. *Assessing Writing, 31*, 84–97.

Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research, 21*(6), 665–680.

Lantolf, J. P., & Poehner, M. E. (2010). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11–33. https://doi.org/10.1177/1362168810383328.

Leeming, P. (2017). Investigating collective-efficacy in the foreign language classroom. *The Language Learning Journal*. https://doi.org/10.1080/09571736.2017.1416424.

Leong, W. S., & Tan, K. (2014). What (more) can, and should, assessment do for learning? Observations from 'successful learning context' in Singapore. *The Curriculum Journal, 25*(4), 593–619.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol: Multilingual Matters.

Marsh, D. (2002). *CLIL/EMILE–The European dimension: Actions, trends and foresight potential public services contract DG EAC: European Commission*, 1–204.

McCrostie, J. (2017, July 5). Spoken English tests among entrance exam reforms Japan's students will face in 2020. *The Japan Times.* https://www.japantimes.co.jp/community/2017/07/05/issues/spoken-english-tests-among-entrance-exam-reforms-japans-students-will-face-2020/#.W19xbbo4DIU. Accessed 25 Mar 2019.

McLellan, P. (2018). Scaffolding excellence: Content-language integration and the development of 'global leaders'. *International Journal of Research Studies in Education, 7*(3), 33–48.

Mehisto, P., & Lucietto, S. (2011). *CLIL essentials*. https://www.icfabriani.edu.it/allegati/443/WHAT%20IS%20%20CLIL%20BY%20PROF.SSA%20SANDRA%20LUCIETTO.pdf. Accessed 25 Mar 2019.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge: Cambridge University Press.

MEXT. (2009). The global 30 project. Ministry of Education, *Japan.* http://www.mext.go.jp/en/policy/education/highered/title02/detail02/sdetail02/1373894.htm. Accessed 25 Mar 2019.

MEXT. (2014). Top global universities. Ministry of Education, Japan. https://tgu.mext.go.jp/en/about/index.html. Accessed 25 Mar 2019.

Mislevy, R. J., Almond, R. G., & Lucas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report 03–16). Princeton: Educational Testing Service.

MLA. (2007). Foreign languages and higher education: New structures for a changed world. https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Teaching-Enrollments-and-Programs/Foreign-Languages-and-Higher-Education-New-Structures-for-a-Changed-World. Accessed 25 Mar 2019.

Negishi, M., & Tono, Y. (2014, April 10–11). *An update on the CEFR-J project and its impact on English language teaching in Japan.* Paper delivered at 2014 ALTE Paris, http://events.cambridgeenglish.org/alte-2014/docs/presentations/alte2014-masashi-negishi.pdf. Accessed 5 Feb 2019.

Negishi, M., Takada, T., & Tono, Y. (2012). A progress report on the development of the CEFR-J. *Studies in Language Testing, 36*, 137–165.

Ng, P. C. L. (2017). English-medium instruction in Japanese universities: Policy implementation and constraints. *Current Issues in Language Planning, 18*(1), 57–67. https://doi.org/10.1080/14664208.2016.1204053.

Norris, J. M. (2016). Language program evaluation. *Modern Language Journal, 100*(1), 169–189. https://doi.org/10.1111/modl.12307.

North, B. (2007). *The CEFR levels: Key points and key problems*. Paper presented at the symposium of the Association of Language Testers in Europe (ALTE): The common European framework of reference for languages: From plurilingualism to intercultural competences, April 18–20 at Sevres. http://www.ealta.eu.org/documents/resources/B%20North%20Presentation%20Sevres%202007.pdf. Accessed 4 Nov 2019.

O'Dwyer, F., & deBoer, M. (2015). Approaches to assessment in CLIL classrooms: Two case studies. *Language Learning in Higher Education, 5*(2), 397–421. https://doi.org/10.1515/cercles-2015-0019.

O'Dwyer, F., Imig, A., & Nagai, N. (2013). Connectedness through a strong form of TBLT, classroom implementation of the CEFR, cyclical learning and learning-oriented assessment. *Language Learning in Higher Education, 3*(2), 231–253. https://doi.org/10.1515/cercles-2013-0012.

O'Dwyer, F., Hunke, M., Imig, A., Nagai, N., Naganuma, N., & Schmidt, M. G. (Eds.). (2017). *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (English profile series 6). Cambridge: Cambridge University Press.

Partnership for 21st Century Learning. (2007). *Framework for 21st century learning.* http://www.p21.org/storage/documents/docs/P21_framework_0816.pdf. Accessed 25 Mar 2019.

Pavon, V., & Ellison, M. (2013). Examining teacher roles and competences in Content and Language Integrated Learning (CLIL). *Linguarum Arena, 4*, 65–78. http://ler.letras.up.pt/uploads/ficheiros/12007.pdf. Accessed 25 Mar 2019.

Runnels, J. (2014). The CEFR-J: The story so far (2012–2014). *FLP SIG Newsletter, 12*, 9–19.

Runnels, J. (2015). Usage of the CEFR and CEFR-J in Japanese universities: Preliminary survey results. *FLP SIG Newsletter, 14*, 8–18.

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing, 25*(4), 553–581. https://doi.org/10.1177/0265532208094276.

Sasajima, S. (2013). How CLIL can impact on EFL teachers' mindsets about teaching and learning: An exploratory study on teacher cognition. *International CLIL Research Journal, 2*, 1. http://www.icrj.eu/21/article5.html. Accessed 25 Mar 2019.

Sasaki, M. (2018). Application of diffusion of innovation theory to educational accountability: The case of EFL education in Japan. *Language Testing in Asia, 8*(1), 1–16. https://doi.org/10.1186/s40468-017-0052-1.

Sasayama, S., & Izumi, S. (2012). Effects of task complexity and pre-task planning on Japanese EFL learners' oral production. In C. A. Coombe & A. Shehadeh (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 23–42). Amsterdam: John Benjamins.

Sato, R. (2010). Reconsidering the effectiveness and suitability of PPP and TBLT in the Japanese EFL classroom. *JALT Journal, 32*(2), 189–200.

Schmidt, M., Naganuma, N., O'Dwyer, F., Imig, A., & Sakai, K. (Eds.). (2011). *Can-do statements in language education in Japan and beyond*. Tokyo: Asahi Press.

Shigemori Bučar, C., Ryu, H., Moritoki Škof, N., & Hmeljak Sangawa, K. (2014). The CEFR and teaching Japanese as a foreign language. *Linguistica, 54*, 455–469. https://doi.org/10.4312/linguistica.54.1.455-469.

Shimojima, Y., & Arimoto, M. (2017). Assessment for learning practices in Japan: Three steps forward, two steps back. *Assessment Matters, 11*. https://doi.org/10.18296/am.0023.

Shintani, N. (2016). *Input-based tasks in foreign language instruction for young learners*. Amsterdam: John Benjamins.

Shohamy, E. (2012). A critical perspective on the use of English as a medium of instruction at universities. In A. Doiz, D. Lasagabaster, & J. M. Sierra (Eds.), *English-medium instruction at universities: Global challenges* (pp. 196–2012). Bristol: Multilingual Matters.

Stevens, D. D., & Levi, A. J. (2005). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. Sterling: Stylus Publications.

Sybing, R. (2011). A response to criticism of TBLT in Japan's language classrooms. *JALT Journal, 33*(1), 67–69.

Tono, Y. (2017). CEFR-J Version 1.0 (English). Retrieved from https://tufspods.files.wordpress.com/2017/03/cefr-j-english-version.pdf

Tono, Y. (2019). Coming full circle – From CEFR to CEFR-J and back. *CEFR Journal – Research and Practice*, 5–17.

Topping, K. J. (2010). Peers as a source of formative feedback. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 61–74). New York: Routledge.

Trim, J. L. M. (Ed.). (2001). *Common European framework of reference for languages: Learning, teaching and assessment. User guide*. Strasbourg: Council of Europe. https://rm.coe.int/1680697848. Accessed 25 Mar 2019.

Tsuchiya, K., & Murillo, M. D. P. (2015). Comparing the language policies and the students' perceptions of CLIL in tertiary education in Spain and Japan. *Latin American Journal of Content & Language Integrated Learning, 8*(1), 25–35.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–404. https://doi.org/10.1080/0141192960220402.

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–274). Berlin: De Gruyter. https://doi.org/10.1515/9781614513827-018.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly, 11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046.

Watanabe, Y., Ikeda, M., & Izumi, S. (Eds.). (2011). *CLIL: New challenges in foreign language education* (Vol. 1). Tokyo: Sophia University Press.

Watanabe, Y., Ikeda, M., & Izumi, S. (Eds.). (2012). *CLIL: New challenges in foreign language education* (Vol. 2). Tokyo: Sophia University Press.

Watanabe, Y., Ikeda, M., & Izumi, S. (Eds.). (2016). *CLIL: New challenges in foreign language education* (Vol. 3). Tokyo: Sophia University Press.

Weaver, C. (2012). Incorporating a formative assessment cycle into task-based language teaching in a university setting in Japan. In C. A. Coombe & A. Shehadeh (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 287–312). Amsterdam: John Benjamins.

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Wicking, P. (2017). The assessment beliefs and practices of English teachers in Japanese universities. *JLTA, 20*, 76–89.

Yamano, Y. (2013). Utilizing the CLIL approach in a Japanese primary school: A comparative study of CLIL and EFL lessons. *Asian EFL Journal, 15*(4), 160–183.

Yang, W. (2017). Tuning university graduates for high mobility and employability under the content and language integrated learning approach. *International Journal of Bilingual Education and Bilingualism, 20*(6), 607–624. https://doi.org/10.1080/13670050.2015.1061474.

Yasuda, S. (2017). Toward a framework for linking linguistic knowledge and writing expertise: Interplay between SFL-based genre pedagogy and task-based language teaching. *TESOL Quarterly, 51*(3), 576–606.

Chapter 5
Assessment for Learning in Bilingual Education/CLIL: A Learning-Oriented Approach to Assessing English Language Skills and Curriculum Content in Portuguese Primary Schools

Ana Xavier

5.1 Introduction and Background

Assessment, or the process of collecting information leading to beneficial consequences (Bachman and Palmer 2010, cited in Council of Europe 2010), has often been regarded as a delicate issue in bilingual education and CLIL contexts. These learning contexts are relatively new to the Portuguese education system, where summative assessment has traditionally been highly valued and where there is little knowledge of formative assessment practices, especially in CLIL contexts. It has been a few years since the framework for CLIL assessment in early primary (ISCED[1] 1) was published online[2] to respond to the needs of the teachers in Portugal implementing the Bilingual Schools Project, the predecessor of the Bilingual Schools Programme (Ministério da Educação n.d.-a; n.d.-b).[3]

The establishment of an institutional cooperation between the Directorate-General for Education of the Portuguese Ministry of Education and the British Council Portugal implemented both the pilot and the programme in a number of state schools in mainland Portugal. These initiatives are linked together as they share common aims, namely to provide bilingual education from an early start, by helping learners gradually achieve higher proficiency levels in English in line with

[1] International Standard Classification of Education

[2] Xavier, A. (2016) *Assessment for Learning in EBE/CLIL: a learning-oriented approach to assessing English language skills and curriculum content at early primary level* [online]

[3] Additional information on the former pilot project and the current programme is available online at the website of the Directorate-General for Education of the Portuguese Ministry of Education

A. Xavier (✉)
Direção-Geral da Educação/Ministério da Educação, Lisbon, Portugal

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_5

curriculum content learning, raise the children's multilingual and intercultural awareness, and promote an inclusive education for all. Conversely, the project and the programme slightly vary in terms of timeframe and scope. The Bilingual Schools Project, implemented from 2011 to 2015, was a pilot involving a small number (6) of school clusters[4] and approximately 800 6–10 year-old learners attending early primary. Following a successful evaluation study of the experimentation of bilingual education in lower primary, the Bilingual Schools Programme was launched in 2016 comprising more school clusters and more educational levels in addition to early primary. These included preschool (ISCED 0) targeting 3–6 year-olds, upper primary (ISCED 1), and lower secondary (ISCED 2) targeting 12–15 year-old learners. There were *circa* 3% (21) school clusters implementing the programme at national level in September 2019. The framework for CLIL assessment has gained momentum both in the curriculum and the assessment arena (see Sect. 1.3, Chap. 1, this volume).

Within both the pilot and the programme, the curriculum has always been regarded as a reference point for learning, teaching, and assessment, involving a principled selection of curriculum content topics and learning aims of the subjects to be taught through the medium of English using a CLIL approach. One key principle was to teach different content in the national language (Portuguese) and the target language (English) to avoid content repetition which could lead to student boredom in the classroom and ultimately disruption to their learning. For example, if students had already learned the body systems in Portuguese, they would likely be less eager to pay attention to the same topic if it were repeated in English. Furthermore, the language education has been integrated with the content lessons to encourage literacy development that *stretches beyond* the everyday foreign language to that required to convey content meaning and ensure "ongoing language growth (being alert to plateauing)" (Bertaux et al. 2010, p. 8).

Throughout the pilot project and the programme, early primary learners were taught part of the primary curriculum through the medium of English by content teachers (or generalist teachers) and English as a foreign language by English specialist teachers. The subjects involved have been *Estudo do Meio* (a combination of History, Geography, and Science), *Expressões* (or Self-Expression which includes, arts and crafts, music, drama) or *Educação Artística* (Art Education), and English as a Foreign Language. Since the programme currently involves more educational levels in addition to early primary, as a result of scaling up, new curriculum reference documents and new subjects have been included in the programme.

In preschool, there are major curriculum content areas entitled *Área de Formação Pessoal e Social* (Personal and Social Development), *Área de Expressão e Comunicação (*Self-expression and Communication), and *Área do Conhecimento do Mundo* (Awareness of the World). These are included in the preschool

[4] Most state schools in Portugal are organised in clusters comprised of one main school and several other schools. The main school is typically a secondary school and it is where the steering, administrative and managing bodies are based. The remaining schools are usually kindergartens and primary schools.

curriculum guidelines (Ministério da Educação 2016) and are supposed to be developed in a cross-curriculum, integrated, and holistic approach which takes into account young learners' preferences. Within the Bilingual Schools Programme, 5 weekly hours of the preschool curriculum (20%) is developed in English.

This integrated approach to content learning is still fostered in early primary where content areas begin to be organised into subject areas and these subject areas gradually become more specific in upper primary and in lower secondary. For example, Awareness of the World, in preschool, will develop into *Estudo do Meio*, in early primary, and into History and Geography of Portugal and Science in upper primary. Further on, in lower secondary, the latter will disaggregate into more specific subjects, namely History, Geography, Science, Physics and Chemistry. Similarly, Self-expression and Communication will turn into *Expressões* or *Educação Artística*, in early primary, and then in Art, Technological Education, Music or Physical Education in upper primary and lower secondary. In the Portuguese curriculum, it is important to note that content areas are a learning continuum gradually moving from general awareness to more and more specific knowledge. This impacts the Bilingual Schools Programme because it promotes the sequentiality of learning in one of two languages. For example, if *Estudo do Meio* is partially taught in English, in lower primary, it is expected that the more specific subjects drawing on this content area, such as Science in upper primary or Natural Science and Physics and Chemistry in lower secondary are also taught in this language. Moreover, it is expected that more challenging content matter within these subjects is selected to be taught in the target language given the fact that it can more likely activate cognitive/thinking skills as discussed in Sect. 5.4.1.

Accordingly, within the programme, in lower primary education 7–9 weekly hours (31–36%) are used for developing English as a foreign language and *Estudo do Meio*, *Expressões* or *Expressáo Artística*, and Physical Education through the medium of English. This is different in upper primary and lower secondary, where the schools implementing the programme decide which subjects are to be taught through the medium of English. The time allocated to the subjects taught in English varies from 9–10 weekly hours (30–37%) in upper primary to 11–12 weekly hours (33–40%) in lower secondary.

In this vein, and considering the expected alignment between curriculum and assessment, tensions have arisen throughout the implementation of bilingual education in Portugal. It did so at the time of the pilot project and it still does today when the Bilingual Schools Programme is being implemented. These tensions have been mainly related to a fluctuation of the value attributed to either summative or formative assessment across education trends, a need for teacher assessment literacy, and how to assess curriculum content through the medium of a foreign language. These tensions are discussed in more detail below.

In the Portuguese education context, summative assessment has traditionally been highly valued in practice to the detriment of formative assessment (Fernandes 2014). Implications were that the administrative purposes were overrated and the pedagogical ones needed to be fostered so as to serve both purposes of assessment, assessment *of* and assessment *for* learning (for the discussion of terms, see Chap. 1,

this volume). That no guidelines were in place to support teachers in classroom-based assessment let alone assessment in CLIL classrooms added to those tensions, including at the primary level of education. Even though many pilot project teachers valued formative assessment, this was not actually embedded in their teaching practice (Xavier 2016). Awareness of this mismatch between these teachers' beliefs about assessment and what they essentially did in their classroom-based assessment has put pressure on the need to tackle this issue. A reflection upon such tensions and implications laid the foundation for devising a framework with practical suggestions for assessing young learners' development of language skills and curriculum content knowledge.

Still in the wider context, within the process of curriculum redesign in Portugal, there has been a shift towards formative[5] assessment as the main practice of assessment in compulsory education. The current legislation establishes the purpose of assessment, it endorses assessment as part of curriculum development, it states that it should serve and inform learning and teaching, and include the use a variety of tools.[6] The importance of formative assessment has also been reflected in the new curriculum reference documents. One example is the *Aprendizagens Essenciais* for English as a Foreign Language in year 3 (lower primary; Ministério da Educação 2018).[7] One of the three competence areas within the new foreign language curricula, strategic competence (the remaining are communicative and intercultural competences), includes such learning objectives as self- and peer- assessment activities using portfolios, journals or learning logs, and progress reports or checklists. Therefore, Portuguese teachers need guidance in assessment *for* learning so as to become more confident in its key concepts and make effective use of related practices. Such guidance could be introduced at the level of policy experimentation at national level by producing formative assessment guidelines and tools to support these teachers so as to have a positive impact in the classroom nationwide.

Within the former project and the current programme at the primary level, it was made clear to the group of content and language teachers co-teaching *Estudo do Meio* and *Expressões* who were involved in the pilot, that what was learnt and taught through the medium of English was expected to be assessed in the target language. However, how to assess content knowledge and language skills with a view to promoting learning still required objective answers, particularly in a context where assessment has never traditionally been a consensual or straightforward issue in terms of conceptual understanding and terminology. Devising this framework was

[5] The Portuguese curriculum uses the term 'formative assessment', which is largely congruent with how assessment for learning is defined in the present volume.

[6] Legislative Order no. 17-A/2015, of the 22nd September (revoked) versus Legislative Order no. 1-F/2016, of the 5th April, and Ministerial Implementing Order no. 223-A/2018, of the 3rd August, on assessment in primary and lower secondary education (both in force)

[7] Ministério da Educação (2018) *Aprendizagens Essenciais/Articulação com o Perfil dos Alunos – Inglês 3.° ano – 1.° ciclo do ensino básico* (new curriculum reference document for English as a foreign language based on core learning skills, knowledge and attitudes targeting year 3 of lower primary and 8 year-old learners)

hence an attempt to help the group of teachers involved in this bilingual pilot, notably, how to cope with the assessment of the content of *Estudo de Meio* and *Expressões* through the medium of a foreign language in addition to the assessment of the English language. However, the framework is yet to be implemented on a larger scale.

The recent curriculum redesign also poses new challenges for assessment as it means adjusting the framework so as to include new learning objectives and success criteria drawing on new subjects so as to build assessment from. For example, as part of a national strategy for citizenship education, a new subject has been added to the curriculum structure entitled *Cidadania e Desenvolvimento* (Citizenship Education) which could be an interesting opportunity for developing learning objectives linked to culture/intercultural awareness and the concept of 'otherness' as the key for discovering self through pluricultural citizenship and multiculturalism (Coyle 2005).

Across Portuguese governments and the educational policies, assessment has oscillated between the grading and certification purposes or diagnostic and formative purposes. Regardless of the political trends, assessment *for* learning has scarcely been part of teacher assessment literacy, despite being broadly acknowledged by most teachers as being important (Xavier 2016). Until recently, there has been a greater focus on summative than on formative assessment. Currently, it is the opposite. Nonetheless, Portuguese teachers have struggled with the practical use of classroom-based assessment and now it is expected that they use it. This is probably the most significant of the tensions previously mentioned as it suggests a need to build teacher capacity on assessment literacy.

Even though the context for the design of this framework in this study was a pilot project in Portugal, the framework is directed at any teacher working in CLIL settings at early primary level since it promotes assessment practices which are suitable for the CLIL classroom, particularly as they can help learners and teachers know what success looks like and how to get there. This chapter will be an opportunity to revisit and discuss the framework for CLIL assessment.

5.2 A Case for Learning-Oriented Assessment

This section will discuss learning-oriented assessment as to its appropriateness to the primary CLIL context in Portugal. Accordingly, two general directions relevant for the present chapter emerge from the discussion in Chap. 7: one has to do with adapting the frameworks (such as CLIL assessment as part of European projects) for assessing for the primary level of education; the other is related to the relevance of learning-oriented assessment as *the* assessment approach that could fit this particular context.

'Assessment in the Primary School Curriculum' (2007) by the National Council for Curriculum and Assessment (NCCA[8]) from the Republic of Ireland and the 'Curriculum for Excellence' (2010), by the Scottish Government can provide assessment frameworks envisioning how assessment *for* learning approaches can help build towards assessment of achievement. In doing so, these frameworks offer guidelines to support primary teachers and school management to ensure that assessment meets the needs of all young learners. Both these documents cover primary curriculum subjects, target similar age-range learners and are designed to support teachers by conveying one key message: that assessment in primary has to entail a formative approach.

Similarly, two European projects dedicated to CLIL assessment at early primary level provide a model for assessing language and content: the Assessment and Evaluation in CLIL Project (AECLIL) (Barbero 2012; Maggi 2012) and the CLIL Learner Assessment Project (CLILA) (Massler et al. 2014). The AECLIL defines reference points, steps, and foci of assessment, namely content, language, and cognition. The CLILA project contributes with a model for integrating content and language assessment. The model aims to link assessment tasks to frameworks of reference for language, learning objectives of content subjects, and the dimensions for scientific literacy. It can be represented in three dimensions: the curriculum content topics of a specific subject, the communicative competence in the foreign language, and the levels of scientific literacy (Massler et al. 2014, p. 142).

These projects suggest that the first step towards CLIL assessment in early primary is setting a reference point that determines what learners need to know. This is where learning objectives for content and language are extracted from and aligned with success criteria and assessment foci are designed. Furthermore, these projects suggest that assessment *for* learning is an intrinsic part of CLIL. To better illustrate this point, taking the curriculum of *Estudo do Meio* as reference for a moment, we could think of CLIL as being the content topic of the circulatory system and formative assessment as being the heart. This analogy may be useful to envision assessment as supportive of learning and teaching. It keeps the latter on track by promoting the use of a repertoire of methods, techniques, and tools which enable the process of seeking and interpreting evidence of learning leading to further learning and more adjusted teaching.

As to the second direction emerging from Sect. 5.1, learning-oriented assessment can be suited for the Portuguese context, as it can best serve the double goal of assessment (Boud 2000, cited in Carless 2007, p. 1): (a) to decide what comes next and (b) to judge what has gone before, notably because it merges formative and summative assessment. The latter is especially relevant, as it keeps balanced the pedagogical purposes of assessment with its administrative and grading purposes. To better understand why, it is worth outlining what learning-oriented assessment is.

[8] The NCCA is an advisory body of the Ministry for Education and Science in charge of curriculum and assessment for primary and post-primary schools in the Republic of Ireland.

Purpura (2004, p. 236), in the scope of language assessment, defined learning-oriented assessment as "the collection and interpretation of evidence about performance so that judgements can be made about further language development." The definition was later echoed by Keppell and Carless (2006, p. 181), who emphasised the learning centrality of assessment and to its applicability for both formative and summative assessment. Learning-oriented assessment can, therefore, be seen as enabling a collection of evidence through a combination of formative and summative assessment of learning that will feed back into assessment planning and adjusting, which will promote further learning. It is worth emphasising that learning is always at the centre of this process.

Learning-oriented assessment introduces a clear model for conceptualising classroom-based assessment, which includes three main principles, also discussed as components, elements, processes, or strands elsewhere (Keppell and Carless 2006, Carless 2007, 2009). These are:

- learner involvement;
- the closing of feedback loops (entailing feedback which reduces the gap between the learner's understanding and performance and the aimed learning objective and success criterion, which direct the following action);
- and tasks seen as learning tasks.

All these are part of a 'unified whole' aligned with learning objectives, which makes assessment and teaching have a clear purpose to help focus learner time and effort. Below is a synopsis of the three components.

First, learners' involvement entails learner training into self-assessment and peer-assessment, which can actively involve learners and helps them from an early age to develop lifelong learning. This links to the formative assessment concepts of self-regulation and self-monitoring (Sadler 1989, p. 121) or the act of regulating learning processes which can enable learners to assess the quality of their work and systematically monitor what they are doing (Black and Wiliam 1998).

If assessment is made transparent to learners and they are taught to reflect on and assess their work and performance, and those of their peers, against success criteria, they will improve their evaluative skills. This so happens because they develop a better understanding of objectives, criteria, and what quality work and performance mean. In this primary CLIL context, learner involvement is particularly relevant as it can activate the cognitive processes of thinking and reasoning which help learners process, understand, and convey information more and more autonomously; and this is crucial to both language and content development (see Chap. 6, this volume). Training learners into becoming active stakeholders in assessment or active assessors will also eventually develop learner autonomy and provide teachers with information that enables them to keep track on learner progress and adjust their teaching based on evidence of this (Wiliam 2008).

Secondly, timely and feed-forwarding feedback is highlighted as regards the specific features and conditions that make it effective. These draw on the model of feedback proposed by Hattie and Timperley (2007), discussed in Chap. 1 of this volume. Accordingly, these features of effective feedback rely on (a) its timeliness,

or the ability to use feedback that can feed into both current (where and how is the learner going?) and future learning (where to is the learner going next?); (b) teacher involvement with challenging objective and support; and (c) learner and peer involvement in feedback with actual effort and active strategies. To help the teacher provide effective feedback to learners, it is important to understand feedback's functioning on four levels, namely:

- on task, to check how well learners are accomplishing it
- on process, to see if learners understand the task and the processes associated with accomplishing it
- on self-regulation, to confirm if learners are able to monitor their work
- on self, to praise or provide positive reinforcement to the learners, based on their personal characteristics but also not to discourage them.

Considering these levels, it is essential that feedback acts on more than the task and/or self-levels, helping to bridge the gap between the learning objective and the outcome or between the actual and desired performance, to direct, and to regulate learner actions. Feedback about process and self-regulation highlighted in Chap. 1 of this volume as being particularly effective, seems to more closely relate to the concepts of timeliness and feed-forwardness of feedback. Accordingly, these levels are able to specifically point out to learners how they can move forward towards specific learning objectives. Therefore, both learners and teachers have key roles in assessment.

The third component is the development of assessment tasks also as learning tasks aligned with curriculum objectives and success criteria and relying on assessment foci or constructs. The use of a varied repertoire of such tasks occupying learners with continuous work in the course of a content unit that fosters learning is regarded as being more beneficial for learning (Bachman and Palmer 2010, p. 26, cited in Council of Europe 2010, p. 13) rather than separate assessment events not based on and linked with learning, as it reflects their learning progress against benchmarks. This so happens because assessment tasks taking the form of authentic learning tasks can be carried out more regularly than traditional tests. Therefore, instead of being based on rote learning, they can have the potential to reflect the learners' best performance and encourage challenging learning.

Likewise, learning-oriented assessment can be also relevant for this primary CLIL context because it challenges the traditional view that formative assessment and summative assessment are not combinable. They can be, as long as the *for* learning purpose of assessment is reinforced more than the measurement ones. This enables a clear understanding of learning being central to feeding learning and teaching forward. This is accomplished by aligning the quantitative (measurement) dimension of summative assessment (testing and exams) and the qualitative (individual) dimension (through teaching expertise in formative assessment techniques that promote learning). The challenge remains in the ability of assessment to promote effective learning.

In addition to the conceptual basis of learning-oriented assessment outlined in the present overview of its three components, this approach also offers a system comprised of a sequence of systematic steps taken in the classroom so as to help

teachers plan, use, record, and feedforward the assessment information. The system can be explained through the learning-oriented assessment cycle (Cambridge English Language Assesment 2013) which suggests taking the following steps in a given lesson so as to include assessment: from setting and sharing of learning objectives; having the learners carry out activities or tasks aligned with those; observing, monitoring, interpreting evidence of learners' progress and taking quick notes; to providing feedback to the class which will eventually feed into adjusted objectives that will restart the cycle once again (see also Davison 2008; Chap. 1, this volume). Both the monitoring and the record keeping can become more formal and structured registers at the end of the lesson which can eventually be used as sound records of progress and achievement underlying formative and summative assessment purposes. Even though this cycle was developed for language assessment, it enables adaptations to several other educational contexts as done by Bentley (2014) at the British Council Regional Policy Dialogues 2013/2014, where she presented an adaptation of the cycle including competence-based assessment foci. The same was done for this CLIL primary context, as will be presented in the following section.

Still, one of the main challenges in any educational context is how to approach institutional level promotion and policy level experimentation so as to develop teacher assessment literacy. Carless (2007) admits that the institutional promotion of learning-oriented assessment can be hard and challenging due to the different conceptions of assessment both on the part of education systems and on the part of the different teacher conceptions on assessment. Attempts to build teacher capacity and disseminate classroom-based assessment practices include module-level implementation of this approach, such as the learning-oriented assessment project (LOAP), at the level of initial teacher training programmes (Keppell and Carless 2006; Carless 2007), which provides suggestions for teacher training in learning-oriented assessment. One issue identified by Carless was how the promotion of student reflection on how these assessment practices could be used with the learners in the classroom. In this CLIL primary context, this could mean acting on the current educational context both on a strategic level and on a remedial level. The first would not only include the need to come to understand to what extent teacher training programmes in Portuguese higher education institutions include formative assessment modules and develop and apply them, in case they are not in place; and/or the second would mean to organise continuous professional development so as to include in-service teachers already involved in CLIL provisions.

This review has attempted to account for the choice for learning-oriented assessment in this specific context. In doing so, this section kept a close eye on the aim of learning-oriented assessment (to strengthen the learning features of assessment) which can be achieved through either formative or summative assessments. Moreover, it extracted from learning-oriented assessment a structured system for planning, using, and recording assessment that can be useful to turn theory into practice, speculating about possible institutional promotion of this approach to assessment. The subsequent section will discuss how learning-oriented assessment can contribute to the design of a framework for primary CLIL assessment with a view to supporting teacher assessment literacy and fostering learning-oriented assessment practices.

5.3 Designing a Framework

The design of the framework was meant to guide the Bilingual Schools Project teachers into putting CLIL classroom-based assessment into practice. Accordingly, it was developed so as to gradually evolve from a more theoretical and conceptual approach to assessment to actual practice, in terms of planning, using, and recording. Thus, the relevant literature review was summarised and reflected upon and subsequently assessment questions set the tone to provide the expected answers to the teachers. A colour scheme was used to easily match the assessment questions with their corresponding answers. In view of this, the framework is divided into three parts: (1) it presents a rationale for supportive and pedagogical assessment; (2) it suggests a continuum of assessment methods, techniques, and tools; and (3) offers a model for planning, using, and recording an assessment sequence.

The first part lays the theoretical background for assessment that CLIL primary teachers need to know. In order to do so, it emerges from the learner as the centre of the assessment process and relies on the five assessment questions, namely *Why* (assess)? *What* (to assess)? *Who* (assesses)? *How* (to assess)? and *When* (do we assess)? as these can be a clear and objective way of simplifying the main concepts to the teachers before classroom-based assessment planning. Figure 5.1 below was inspired by the conceptual basis of learning-oriented assessment (Carless 2007), whereby the learning element is more enhanced and presented as a unified whole. The learner is in the middle of the figure as a reminder that it is the learning that drives the assessment.

Following Fig. 5.1 clockwise will show how the answers to the questions are aligned with this conceptual basis as they mainly focus on: (1) the pedagogical purposes and supportive principles that should be enhanced in the Portuguese educational context given its tradition of favouring the administrative and grading purpose of assessment (cf. introduction and background); (2) the setting of reference points for assessment coming from the primary curricula, including CEFR (Common European Framework of Reference; Council of Europe 2001) levels for English language proficiency (see Chaps. 2, 4, and 6, this volume), as well as from the CLIL literature regarding cognition and language, in addition to the foci or constructs of assessment emerging from those reference points, namely, content, language, learning to learn and behaviour/attitude, which can support young learners' overall development in a foreign language; (3) the use of learning-oriented assessment in the classroom; (4) the roles of learners and teachers in assessment; and (5) the (systematic) frequency these stakeholders should consider when using assessment in the classroom.

The second part of the framework presents a set of eight guidelines drawing on the rationale discussed previously:

- set objectives & criteria and share them with the learners;
- observe to document and reflect on progress;
- involve learners and get feedback from them;
- provide feedback to support progress;
- scaffold learning tasks/activities;

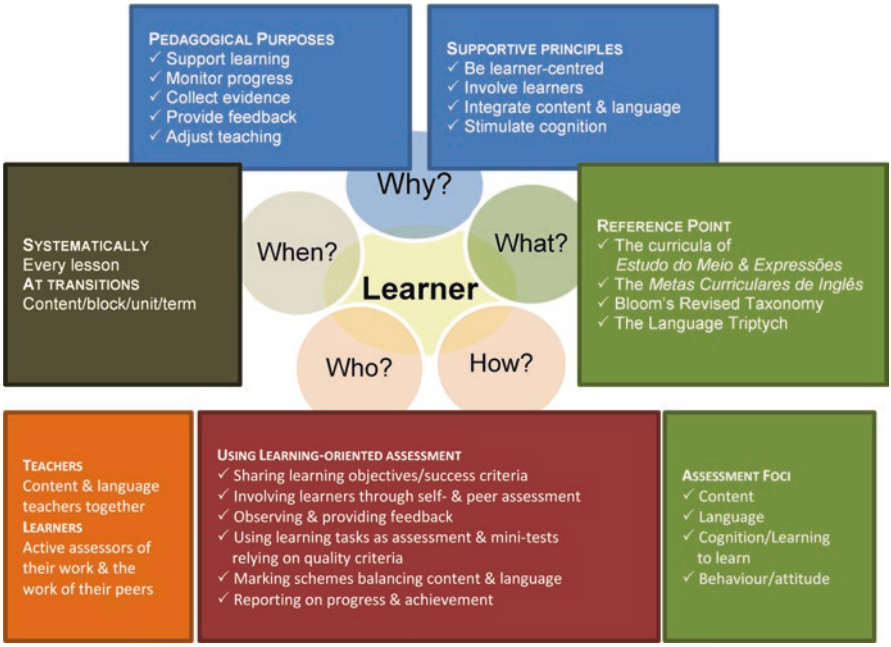


Fig. 5.1 A rationale for supportive and pedagogical assessment

- use mini-assessments to assess progress & achievement;
- use separate marking schemes and different weightings; and
- report on progress and achievement.

These guidelines are linked to a continuum of assessment methods that can help teachers plan for the assessment of a part of a curriculum content topic, integrating it with learning and teaching. Practical techniques and tools adapted from online primary and CLIL resources or from the Bilingual Schools Project teachers' own material complement the methods suggested. The key idea underlying the design of this part is that teachers can research, select and adapt the techniques and tools that seem to fit the learning outcomes and the needs of their learners after having become more confident regarding the theoretical background that supports the choice of assessment methods. All the guidelines emerge from the answers to the *how* question, whereby the *manner* of assessment is the focus which will, in turn, link to the methods drawing on the key concepts of formative assessment and learning-oriented assessment. Guidelines 1, 3, 4, and 6 more directly connect to the components of learning-oriented assessment in this CLIL primary setting so they will subsequently be discussed.

Guideline 1 suggests that learning objectives and success criteria are devised for the assessment foci or constructs that "support children in becoming more

cognitively and communicatively empowered" (Xavier 2016). These need to be shared in the classroom so as to make the assessment process transparent and prime the children for deeper learning experiences as they progress towards the established benchmarks (Carless 2007). This guideline will influence the following ones as those are always dependent on objectives and criteria (see Chap. 6, this volume).

Guideline 3 recommends teachers to actively involve learners in self- and peer -assessment since this will make them feel more engaged in objectives and criteria, better understand the quality of their work and performance as those of their peers'. Self- and peer -assessment can foster learner involvement through the sharing of learning objectives/success criteria and when teachers observe learners try out tasks and get feedback from them on process and completion. And, indeed, all this can be achieved if assessment tasks are designed as learning tasks.

Guideline 4 advocates for teachers to monitor learner progress by observing how they perform and by giving feedback on their performance against the previously established learning objectives. In a learning-oriented mode, it is crucial that feedback is timely and helpful in terms of serving its purpose to help learners better learn and perform now and in the future. The more specific and immediate this feedback is conveyed to the learner, the more it will support and feed learning forward.

Finally, **guideline 6** endorses the design and use of learning tasks as assessment and relying on quality criteria that ensure they are valid and reliable assessments that have a beneficial impact on learning and are practical to administer. Section 5.4.2 discusses examples of this by suggesting the use of child-friendly mini-assessments to assess progress and achievement.

Moreover, this guideline proposes that assessments are aligned with objectives and criteria to generate learning and are carefully scaffolded so as to support learning in a foreign language to primary school children. To appeal to these young learners, these tasks are expected to reflect authentic 'real-world tasks', be 'cooperative' and cater for learners' choices and preferences (Keppell and Carless 2006). In fact, the motivational trait of tasks is key to encourage learners to produce their best performance "and mirror the kind of learning we wish to promote, so as to stimulate complex learning" (Carless 2009). This guideline draws on the learning-oriented assessment component of assessment tasks as learning tasks which properly fits this setting because it enables the use of frequent and varied mini-assessments resembling learning activities and tasks rather than the administration of long and one-off written tests taking place only once or twice a term only involving writing.

A repertoire of assessment methods, techniques and tools that illustrate the previous guidelines can be found in the second part of the framework based on a specific primary curriculum content topic. The point was to show teachers that they can use this as reference to devise their own assessment material with different learning objectives and success criteria. The guidelines were explained as to what they were, why they were beneficial for assessment and how and when they could be integrated in the lesson. Figure 5.2 summarises the methods (upper arrow) and accompanying techniques and tools (lower arrow) which attempted to clearly show where the learner is going, where the learner is now and how to help them get there (Leahy

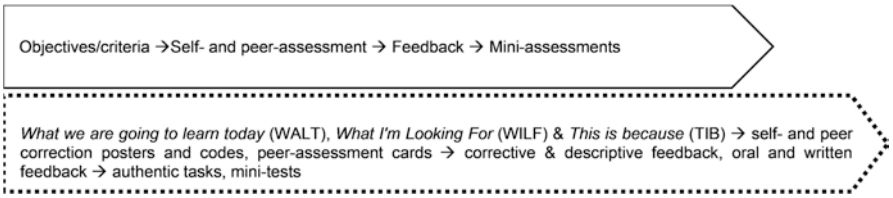


Fig. 5.2 A continuum of assessment methods/techniques/tools; Abridged and adapted from Xavier (2016)

et al. 2005, cited in Wiliam 2014). Some practical examples will be provided in Sect. 5.4.

This second part of the framework is aligned with learning objectives and success criteria originating from the primary curricula and the CLIL features of assessment as these were regarded as the starting point for assessment which produced relevant constructs or foci for primary CLIL, namely content, language, behaviour/attitude, and learning to learn/cognitive skills. This reflects one of the components of the learning-oriented assessment framework addressing the need for alignment between objectives, content and assessment (Biggs 1999; in Keppell and Carless 2006). Such alignment revealed as crucial to the design of the tasks in this second part and, as this was being devised, there were recurrent references to objectives and criteria throughout the assessment guidelines proposed. Furthermore, the development of objectives and criteria as being the first guideline to be explored was intentional as it would set the tone of the remaining guidelines.

The third and final part of the framework creates a system inspired by the learning-oriented assessment cycle (Cambridge English Language Assessment[9] 2013) and modified by Bentley (2014) so as to fit CLIL. It attempts to demonstrate how the previous guidelines and related methods, techniques, and tools can be used as a unified whole or a cycle during the time of a CLIL lesson at early primary level. As shown in Fig. 5.3, this adapted system or cycle adds a step for involving learners through self- and peer -assessment as this is a key component of learning-oriented assessment. On the left, there are the assessment foci or constructs suggested for this CLIL primary context as being relevant for the child's learning development through a foreign language in addition to language levels which can represent benchmarks for the four early primary years. These operate as reference points for assessment to depict the starting and ending point of the cycle. As to the actual cycle, it includes the steps teachers can take during the time of a learning sequence in a lesson so as to integrate the assessment continuum which can support effective learning.

A focus on the learning aspect of learning-oriented assessment is the common denominator to the design of this primary CLIL assessment framework. This was shown in the first part of the framework whereby the answer to all the five assessment

[9] Currently Cambridge Assessment English.

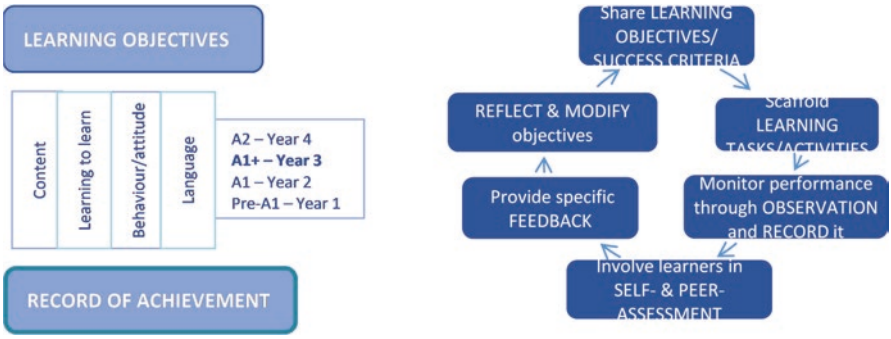


Fig. 5.3 Planning, using and recording an assessment sequence. (Xavier 2016)

questions drew on the learner as the centre of assessment and was in line with the supportive nature and pedagogical principles of assessment in an early primary CLIL setting. The guidelines in this second part tried to help promote productive learning by including learner involvement and feedback. Finally, in the third part, there has been an attempt to show how the previous methods, techniques, and tools can be used as sequential steps that integrate a cohesive learning ethos to assessment. The following section of this chapter will provide some practical examples withdrawn from the framework which attempt to demonstrate how learning-oriented assessment can be done in the classroom.

5.4 Assessment in Practice

The practical examples selected from the primary CLIL assessment framework aim to reflect the components of the learning-oriented assessment model, namely the design and implementation of assessment tasks as learning tasks; the involvement of learners in assessment as peer- or self-assessors; and feedback as feedforward.

The first component is associated with the guidelines *set objectives & criteria and share them with the learners* and *use mini-assessments to assess progress & achievement*. These will be addressed first.

5.4.1 Set Objectives & Criteria and Share them with the Learners

Under the assumption that there should be alignment of objectives, content, and assessment (Biggs 1999; in Keppell and Carless 2006), objectives are the beginning of all, as without them quality assessment promoting learning will not happen. As such, this alignment is necessary for task design and implementation.

Considering the assessment reference points and foci discussed in Sect. 5.2, it is necessary that, at the assessment planning level, the overall learning intentions or aims stated in the curriculum are differentiated from more specific learning objectives and criteria. The first should be used as reference but then be reformulated into simple and short statements in a language that is clear and understandable by the children.

The reference point for devising the assessments in this part comes from the intersection of the curriculum of *Estudo do Meio* and *Expressões,* the *Metas Curriculares* for English (Ministério da Educação e Ciência ed 2015), which contains ability level descriptors ranging from A1 to B1+ in line with the CEFR, verbs fostering learner's cognition moving from lower order thinking skills (LOTS) to higher order thinking skills (HOTS) embedded in Bloom's Revised Taxonomy (Krathwohl 2002)[10] and the specific CLIL language *of* and *for* learning and *through* learning (Coyle et al. 2010). Hence, by combining learning objectives and criteria drawing on several curriculum reference documents, the integration of content and language discussed in Chap. 1 of this volume is being assured right from the start when planning assessment.

Accordingly, objectives are devised in infinitives to convey what children need to learn and what needs to be assessed. Criteria can be expressed in the form of can-do statements focusing on the action the children will have to perform to show they reached that objective. That action is the evidence of learning. Further down is a suggestion inspired by Clarke (2003) for using objectives and criteria as reference points for supporting the learning that is going to take place, namely for sharing these objectives and criteria with the learners in the classroom. This has been adapted to the pilot so it included the specific *Estudo do Meio* content of the circulatory system as reference and a language and a learning to learn focus.

Learning objectives and success criteria can be introduced in the CLIL primary classroom by first scaffolding the children's understanding of the concept. This can be achieved by showing what is meant by the prompts *What we are going to learn today* (WALT) (a standard for success), *What I'm Looking For* (WILF) ('can-do' statements conveying the action that shows that children achieve the objective), and *This is because* (TIB), or showing a reason why children are learning something specific. Then, it is important to make these statements visible in the classroom, which can be more meaningfully accomplished if children take part in writing them on paper strips, illustrating them with purposeful drawings conveying their own perceptions of WILT, WILF, and TIB and sticking them onto the board or on a wall poster. Children can also be involved in the making of the paper strips and wall poster in an *Expressões* lesson focusing on arts and crafts. The WALT, WILF, and TIB can be used (a) in a formative assessment perspective, at the beginning of the lesson routine, during the lesson, at the end of an activity/task, or at the end of a

[10] LOTS and HOTS categorise learning objectives according to cognitive processes, ranging from the least cognitively demanding, such as remembering, understanding or applying (LOTS) to the most challenging ones, such as analysing, evaluating or creating. The combination of LOTS and HOTS can create opportunities for deeper learning and dismiss rote learning.

lesson when checking understanding; and (b) in a summative assessment perspective that shows what children's achievements are after a content unit or a school term.

A learning activity observed in a content lesson of *Estudo do Meio* and *Expressões* in a Bilingual Schools Project school consisting of making a part of a booklet on the circulatory system with a paper and woolen heart and aiming at identifying parts of the heart and understanding how the circulatory system works could be used as assessment using the previously mentioned tool. Below is an example of this tool combining a double focus on the content of *Estudo do Meio* and *Expressões* to be in line with the expected cross-curriculum and integrative approach to learning in primary CLIL. The descriptors include HOTS, namely *analyse* and *create*:

Learning objectives/WALT
We are learning to make a model of the heart.
We are learning to describe how the heart works
Success criteria/WILF
I can follow instructions
I can cut out a paper model of the heart
I can put/pull a thread of wool through a hole
I can describe how the circulatory system starts and ends
This is because/TIB
I am making a booklet on the circulatory system
The paper heart can help me remember and talk about the circulatory system with my classmates.
Adapted from <https://www.pinterest.com/pin/310537336783517867/> and Xavier (2016) [Accessed 01/06/2018]

Objectives and criteria should guide classroom-based assessment and be a benchmark when getting feedback from our learners. In this vein, learners will be able to compare their learning progress against a reference point. This also enables tracing how learners' knowledge changes over time, is based on the learning objectives, and realised as a set of can-do statements related to content, language, and metacognition.

Nevertheless, the relevance of objectives and criteria does not confine to setting up benchmarks for success and getting learners familiar with them at the beginning of the lesson. In fact, they set up or even raise the level or standard of the design of assessment tasks as learning tasks. Further suggestions are discussed below.

5.4.2 Use Mini-assessments to Assess Progress & Achievement

Alignment of assessment tasks as learning tasks with objectives is key to making assessment valid (assessing what it is supposed to assess), considering the different foci or constructs of assessment. These mini-assessments as part of learning tasks can resemble a child-friendly authentic experience that learners can better relate to

and which will enable them to express their age-related abilities. This can be more engaging and appealing to children because they can be more hands-on and involve touch, move, play, and pretend. This can be helpful to understand how to organise the classroom activities such that they contribute to learning; hence, considering the learner age, short activities requiring kinaesthetics and touching are selected.

Moreover, mini-assessments can be administered frequently in the classroom instead of written tests administered once or twice a school term. Therefore, mini-assessments can not only be beneficial for learners but also teachers. For the learners, they can cater for age-related short attention spans, need for frequent movement and use of several senses such as touch, sight, and ultimately learn by play, which will more positively impact on learning. For the teachers, they can be more practical and less time consuming as they take less time to devise, administer, and grade. They will also more likely ensure reliability as they can only focus on one assessment construct at a time and a number and variety of items and techniques which can give learners several opportunities to show their learning progress and how engaged they are which is a strong predictor of motivation and success.

For example, drawing on the content learning activity mentioned in Sect. 5.4.1, if the focus of assessment now turns to language, more specifically the productive skills, the following objectives and criteria can be suggested for a writing task in the form of a simple fact file which can be included in the booklet on how the circulatory system works can be devised. Since the aim is to write a fact file on the circulatory system targeting the writing skill, the objective/criteria pair can be as follows: "to write a fact file on the circulatory system."/ "I can spell and write the names of the organs."; "I can write short sentences about the organs and their functions." The writing process can be previously scaffolded by first presenting a similar model to the class, clearly prompting and elucidating how to do it and afterwards setting up a group work activity whereby they do a similar fact file of their own on an organ of the circulatory system or the types of circulation, using the same model.

Furthermore, in an assessment sequence vein, a speaking task could follow the previous task, whereby the learners would be expected to report on how they made the booklet. This could include the following objective/criteria pairs: "to present oral work previously prepared"/"I can greet the audience."; "I can speak clearly and loudly."; "I can say what the presentation is about."; "I can face the audience." and/or "to recount"; "to retell"/"I can tell the class what I did." HOTS and LOTS verbs are combined in the formulation of descriptors as they involve *apply* and *understand*. This can be scaffolded first so that children can be assisted in the process of getting prepared for an oral presentation by using the poster as an anchor chart to recap roles, namely of the speaker and the audience, with short sentences such as: "Be prepared, greet the audience and introduce yourself", "Say what the presentation is about …", and/or "Listen, show interest, ask questions, raise your hand to speak …" As learners get set for this task, they can take part in questioning about the circulatory system, using correct intonation in statements, questions, and exclamations, when greeting, introducing themselves, and thanking others. Then, children can devise their scripts and rehearse the oral presentation in groups. For the

actual presentation, children will present their recount or fact file in plenary. This can be an exciting opportunity for children to show what they know about the content topic.

This assessment sequence may propitiate the adjustment of teaching that follows assessment; a lesson with a linguistic focus devoted to learners' presentation of the previously learned content followed the assessment of the content activity (making a woollen and paper heart and marking its parts); to make a stronger link between the two, insights from assessment of the former activity—this could be gaps in terminology or in the understanding of what happens in different parts of the heart—should inform the lesson with the linguistic focus.

In terms of language assessment, tasks and activities can be used to focus on one skill at a time in order to assess it in more depth. Even though language skills are integrated in nature, assessing language skills separately will make assessment more valid and reliable because it will enable learners to have a range of opportunities to show what they know and/or how well they can perform a given task. This also informs how learners' knowledge is elicited in the mini-assessments; the assessments focus on one aspect at a time and the techniques are varied so learners are given multiple opportunities to display their knowledge.

The examples focused on language skills can be suitable for the English language lesson, whereas the content one would better fit the content lesson of *Estudo de Meio* and *Expressões*, nevertheless the language teacher and the content teacher should collaboratively discuss their assessment procedures and what constructs to value more and less in content or language lesson. In fact, the integration of content and language in this context requires close collaboration between colleagues. Both content and language teachers need to liaise when they manage the curriculum considering the planning, delivery, reflection and assessment stages. This way it is possible to integrate content and language so that one can serve and *stretch* the other.

Any of the previous tasks offers a range of possibilities for assessment which link to the second component of the learning-oriented model, namely learner involvement in assessment as peer- or self-assessors. This connects with one of the key concepts of formative assessment related to assigning roles to all stakeholders in assessment, in this case to learners.

5.4.3 Involve Learners and Get Feedback from Them

It is directly linked to self- and peer -assessment or systematically and consistently engaging and training learners into assessing their own work or the work of their peers, against learning objectives and success criteria so as to adjust learning and teaching. This can include varied child-friendly tools that keep learners engaged and challenged. Self- and peer correction of the learners and their peers' exemplars can be an idea to involve learners in assessment by correcting their own exemplars or

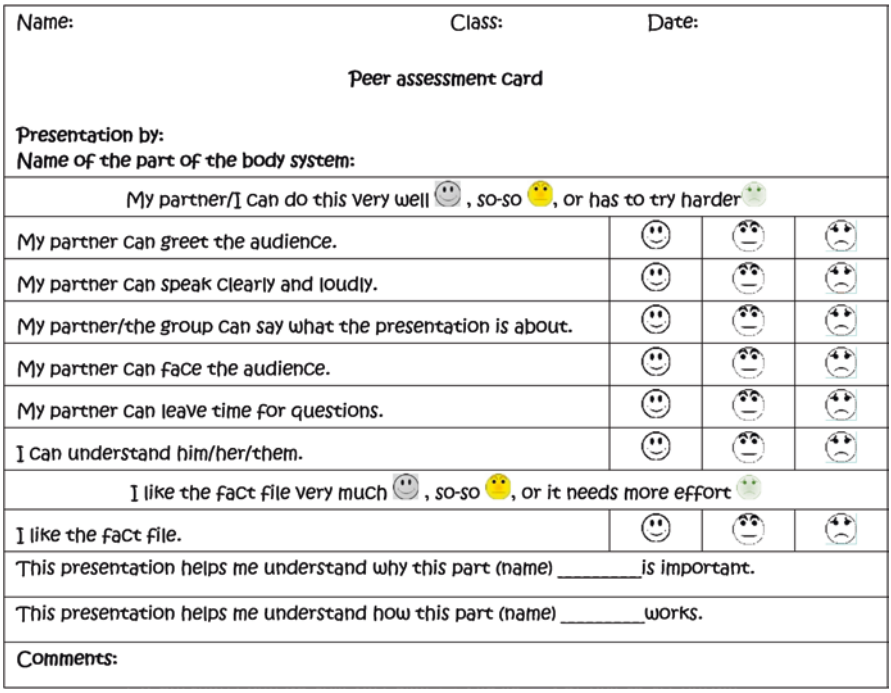
those of their peers using self- and peer correction posters and correction codes as these can assist learners create the routine of self-checking their written work and checking the work of their peers. For example, as regards the writing task previously mentioned on a written fact file on the circulatory system, learners can build a COPS poster in lessons combining language and *Expressões*, which will work as an anchor chart to help them focus on aspects of accuracy, such as remember how to correct their work using capitalisation, order and organisation, punctuation, and spelling (COPS). This can be done by first familiarising learners with the importance of editing their work before handing it highlighting that this way their work can be even better; and then be trained by using the learners' written exemplars such as sentences, small paragraphs, a writing sample produced by the teacher including the learners' frequent mistakes. The exemplars can be read aloud or written on the board, then the learners can be helped notice, underline mistakes by asking questions such as: "Are there capital letters missing?", "Where are they missing?" (in proper nouns – names of people, places; beginning of sentences), "Is there a beginning, middle end to the paragraph?", "Don't miss out punctuation at the end of sentences, in the middle of sentences, in quotes." (accompanied by the corresponding graphic representation of the punctuation); in spelling: "Look for mistakes! Look up for the correct word in your glossary." In *Expressões* lessons, learners can build their COPS posters and correction codes as using their own pieces of material will develop their sense of ownership and boost their self-confidence.

Using COPS posters and correction codes as visual support in language lessons will enable learners to correct their writing pieces or those of their classmates since these will help them underline and signal the type of mistake. Learners can be engaged in a peer-correction activity where, divided in groups of four, each group member can be assigned a specific role: one corrects the capitalisation, another the organisation, the punctuation or the spelling.

Following an *Estudo do Meio* and *Expressões* lesson, a language lesson can focus on developing presentation skills and include *peer assessment cards* to foster and maintain learner involvement in oral presentations, such as presenting a fact file of the circulatory system. This can involve direct support to the learner by allowing them to first rehearse their presentation in their group while their peers assess the performance in the rehearsal by completing the card below (Fig. 5.4).

Secondly, after the rehearsal, the actual presentation of a fact file on one part of the circulatory system by a group of learners before the whole class can involve the use of a similar card, which will help the learners who are part of the audience be engaged in the oral presentation activity.

Learning can be conceptualised differently in different assessment activities: determining whether learners can or cannot yet identify different parts of the heart, which presupposes a stage-like development, allows for making different inferences from those emerging in assessment when learners assess whether their peers can present on the same topic and help their peers to complete the presentation.



Name:	Class:	Date:		
Peer assessment card				
Presentation by: Name of the part of the body system:				
My partner/I can do this very well, so-so, or has to try harder				
My partner can greet the audience.				
My partner can speak clearly and loudly.				
My partner/the group can say what the presentation is about.				
My partner can face the audience.				
My partner can leave time for questions.				
I can understand him/her/them.				
I like the fact file very much, so-so, or it needs more effort				
I like the fact file.				
This presentation helps me understand why this part (name) _____ is important.				
This presentation helps me understand how this part (name) _____ works.				
Comments:				

Fig. 5.4 Peer assessment card for individual presentation; adapted from Ioannou-Georgiou and Pavlou (2003, p. 97) and Xavier (2016)

Any of the previously presented self- and peer -assessment tools can contribute for learner involvement against objectives and criteria, which can ultimately favour learner perception of quality (Falchikov 2005, cited in Keppell and Carless 2006).

Last but not (by far) the least, the guideline *provide feedback to support progress* links to Sadler's idea of 'feedback as feedforward', 'the closing of feedback loops' and 'feedback as closing the gap' discussed in the formative assessment literature (Sadler 1989).

5.4.4 Provide Feedback to Support Progress

The idea of supporting progress is so entwined as a key feature of feedback that makes it unthinkable to feature the information provided to learners as such if it does not cause a timely future action that promotes deeper and further learning in

relation to the objectives and criteria defined. Feedback can be given in a natural way nonetheless it should also have a corrective focus whether it is on language (for example, fluency or accuracy) or content, behaviour/attitude or a learning to learn/cognition. Accordingly, it is key to use techniques and tools that take full advantage of feedback in practice as a catalyser for learners to act upon. In this context, oral and written feedback encouraging self-regulation and self-monitoring seem more suitable and dynamic ways of closing the gap between their current and desired performance and see how this can feedforward into future work, that is, can be used by the learners (Keppell and Carless 2006). In view of that, it is relevant that the selection of feedback techniques and tools is supported by research. For example, in Hattie and Timperley (2007) study results on the correlation of effect sizes to feedback effects provide evidence that some feedback techniques are more powerful than others in terms of promoting learner achievement. Immediate feedback is regarded as more powerful than delayed feedback. Corrective feedback has, too, been found effective.

Oral feedback can be expressed verbally or non-verbally. In any form, oral feedback can take a dialogic form, including questioning to foster cognition, during learner/teacher interaction on activity/task process or after this is finished. Non-verbal techniques can range from little red flags to signal mistakes in a non-scary way if the focus is listening skills; crossing fingers for word order mistakes resulting from code switching, such as when learners mix up 'carbon dioxide' with 'dioxide carbon' because this resembles the Portuguese term. In addition, verbal techniques can be used to correct language errors that can harm fluency when conveying meaning. Some examples are provided by deBoer (n.d.) in his blog entitled *CLIL Media* which provides practical suggestions for CLIL teachers, such as:

- *recasting or reformulating* so as to replace the mistake with the correct form;
- *clarification requests*, where the teacher can elicit the learner to clarify what has been said but there is no actual correction of the mistake on the part of the teacher (only the learner);
- *elicitation*, where an incomplete sentence is provided by the teacher so that the learner can then complete the sentence after an emphatic pause from the teacher; and
- *repetition*, whereby the mistake separately repeated and emphatically highlighted.

Whatever the technique, the learners should be expected to be aware of the mistake and say the correct version again in order to internalise it.

In addition to being corrective, written feedback has the advantage of being descriptive as they can focus on progress or achievement and clearly convey forms of improvement even if it follows the dialogic style of oral feedback. For example, when giving a mark on an achievement such as *poor* or *good*, this mark can be accompanied by a more descriptive comment suggesting a motivational action for improvement, such as when handing out to the learners a writing activity, already

marked, on an invitation for another class to visit a body system exhibition. This can entail a corrective feedback comment focusing on accuracy, such as

Good work.
You only made 3 spelling mistakes. Can you find them?
You forgot to capitalise the first word in line 3. Can you correct it?

Adapted from <https://www.youtube.com/watch?v=5cWjOegvkCU> and <http://datause.cse.ucla.edu/DOCS/pt_tea_1996.pdf> [Accessed 15 August 2015] in Xavier (2016)

This kind of comment can develop learning to learn objectives, as it can help learners become more aware of the quality of their work. By making such specific feedback comments the teacher will help the learners gradually become more aware of how to critically review their work and do better next time.

After learners are more familiar with corrective comments, the teacher's feedback can represent an opportunity to elicit a language extension from the learners. For example, when comments are added to the correction of the same invitation piece of writing which show curiosity by what they wrote: "I really liked the title you gave to the exhibition."; "Don't miss our exciting body systems exhibition."; "Can you add one more sentence about what is exciting about the exhibition?" or "How are you planning to give out this invitation? Who are you going to invite? Please tell me more!" (Xavier 2016).

The interaction in this example involves feedback to learners also that coming from other learners; it is due to the age of the learners, that the examples of feedback given have a strong focus on correction, however, this same feedback should support progress (feed forward) and be given systematically (building on the emerging understanding of learners abilities and on the development of these abilities).

All in all, feedback should be given systematically in every lesson as learning and teaching develop and occur both on task process and product (Massler and Stotz 2013, pp. 78–80) but combined with a focus on the learner as self from a developmental perspective because these learners are children learning in a foreign language. For this reason, whether it is conveyed orally or in writing, immediately or delayed, it is important that it is expressed in a friendly, positive, corrective, and descriptive way to point directions for improvement in relation to objectives without harming children's vulnerability. Furthermore, for reasons of clarity and effectiveness, the use of the target language in any form of feedback should be carefully planned and conveyed in a simple way that learners can easily understand. This will more likely establish more solid connections between planned objectives and tangible outcomes.

Summary of the feedback types discussed

ORAL FEEDBACK

should be

- ✓ immediate/delayed
- ✓ natural, friendly, positive way but corrective focus
- ✓ focusing on language/content
- ✓ using questioning to foster cognition

includes

non-verbal techniques

- ✓ signalling word order mistakes from code-switching, e.g. "carbon dioxide" with "dioxide carbon"
- ✓ counting fingers or crossing fingers


- ✓ raising little red flags to signal mistakes

verbal techniques

Teacher: What does the circulatory system do?

Learner: The circulatory system moves blood the heart and our body.

Recasting/reformulating

Teacher (T): Good, so the circulatory system moves blood **through** the heart and **around** our body?
Can you try again?
Learner (L): Yes, the circulatory system moves blood **through** the heart and **around** our body.

Clarification requests

T: That's almost correct. Would you like to you try again?
L: Yes, the circulatory system **moves** blood **through** the heart **and around** our body.

Elicitation

T: The circulatory system moves blood …
L: The circulatory system moves blood through the heart and around our body.

Repetition

T: *Moves blood the heart and our body …*
L: The circulatory system **moves blood through the heart and around our body.**

Adapted from <http://clilcoursesonline.com/how-to-provide-effective-feedback> [Accessed 20/07/2015] in Xavier (2016)

WRITTEN FEEDBACK

- ✔ should be corrective/descriptive
- ✔ can focus on progress or achievement
- ✔ should provide a clear message on how to improve

Corrective feedback comments

Good work.

You only made 3 spelling mistakes. Can you find them?

You forgot to capitalise the first word in line 3. Can you correct it?

Adapted from <https://www.youtube.com/watch?v=5cWjOegvkCU> and <http://datause.cse.ucla.edu/DOCS/pt_tea_1996.pdf> [Accessed 15/08/2015] in Xavier (2016)

This section has attempted to illustrate how the components of the learning-oriented assessment model can be applied to learning and teaching in an early primary CLIL setting. Accordingly, some practical examples embedding learner involvement, feedback, and assessment tasks as learning tasks were withdrawn from the framework and presented here in order to help CLIL primary teachers plan a content unit using assessment and embed assessment in their daily practice. There was a discussion on how assessment activities and feedback should be organised such that learners are engaged and motivated; allowing learners multiple opportunities to demonstrate their learning, giving them agency in finding and acting upon gaps, and eliciting learner strengths alongside their weaknesses are key to enabling this. This will assist teachers gather sound evidence on an ongoing basis of how children progress so as to inform and adjust learning and teaching and report on achievement to parents.

5.5 Critical Analysis of the Framework

Since this framework has not been tested, and the original CLIL learning context no longer just involves early primary but other education levels as well, the fact that it has been revisited in the scope of this chapter may inspire an upgrade of the framework in the future. This could, for example, entail adding intercultural awareness as a new assessment focus that relates to one of Coyle's four Cs (Coyle 2006)—the C standing for *culture* that was left out of the original framework. As such, new assessment tasks as assessment tasks could be devised for intercultural awareness which could also provide opportunities for addressing the assessment of *Cidadania e Desenvolvimento* (Citizenship Education), the new subject introduced in the Portuguese curriculum. Furthermore, at the level of planning a content unit for assessment, it could involve the definition of new learning objectives and success criteria for more advanced English language levels and new content subjects, in upper primary or lower secondary education, such as Physics and Chemistry. The design of new adjusted learning tasks as assessment tasks could, in addition to the Portuguese curriculum, use the recently published CEFR companion volume (Council of Europe 2018) as it contains opportunities to devise fine-tuned descriptors. Additionally, at the level of planning, using, and recording an assessment sequence in a lesson, a possible upgrade of the framework could also include learning-oriented assessment cycles for not only language lessons but also content ones. This reflection upon how the framework can be adjusted in the future, may also enable a renewed opportunity for research at the level of CLIL assessment that could support more teachers from several education levels.

In this vein, a potential upgrade of the framework could respond to the educational momentum in Portugal whereby formative assessment began to be more valued and depth not breadth in learning is favoured across education levels (Sect. 5.1). If the framework were institutionally promoted, module-level implementation could be organised as continuous professional development targeting in-service primary

and secondary school teachers. They could trial a learning-oriented assessment toolkit and adjust it to their specific context. This way teachers could be supported to diagnose, improve, evaluate, and scale up good practice that would in the end promote deeper learning as learning-oriented assessment intertwines with cognition and the development of thinking processes.

5.6 Conclusion and the Way Forward

This chapter has revisited an assessment framework proposed for an early primary education CLIL setting and discussed how it was designed so as to reflect a learning-oriented approach to assessment and suit an educational content where such assessment practices still need to be in place. The framework suggested a variety of methods, techniques, and tools drawing on specific curriculum content topics and other relevant reference points for CLIL assessment at early primary level.

Additionally, revisiting this framework enabled an in-depth reflection about ways to update, improve and trial it in the future, notably in the light of the recent changes targeting curriculum and assessment in the Portuguese education system. Drawing on this reflection, possibilities were also tackled as to how to adjust the framework to the current scope of the Bilingual Schools Programme which involves several education levels and subject areas.

Accordingly, and as suggested in the previous section, trialling a CLIL assessment toolkit based on the original framework could therefore represent a valid opportunity to help teachers keep the progress/achievement combination in mind as one does not happen without the other and always keep the learning function of assessment working.

Acknowledgements I would like to thank Vanessa Boutefeu, former Master's supervisor and friend, for having informed me about this opportunity and to its editors, Mark deBoer and Dmitri Leontjev, for having enabled me to contribute with this chapter and for all their feedback and support in this process. Without them this work would not have been possible.

I take this opportunity to again express my gratitude to the Bilingual Schools Project teachers and head teachers, the British Council Portugal and the Directorate-General for Education of the Portuguese Ministry of Education for their contributions to the development of this framework for assessment.

Further Reading

Bentley, K. (2010). *The TKT Teaching Knowledge Test course: CLIL module: Content and language integrated learning*. Cambridge: Cambridge University Press.

Fuentes, M. A. (2013). What score is adequate: Approximation to the assessment rationale used in a Science through English CLIL written test. *Bellaterra Journal of Teaching & Learning Language & Literature, 6*(4), 54–73. http://www.raco.cat/index.php/Bellaterra/article/viewFile/297851/386834. Accessed 11 Dec 2018.

Gallacher, L. (2004). *Learner Training with Young Learners*. https://www.teachingenglish.org.uk/article/learner-training-young-learners. Accessed 25 May 2018.
Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
Mehisto, P., Marsh, D., & Frigols-Martin, M. J. (2008). *Uncovering CLIL: Content and language integrated learning in bilingual and multilingual education*. Oxford: Macmillan.
Pinter, A. (2009). Learning to learn. In *Teaching young language learners* (pp. 99–114). Oxford: Oxford University Press.

Legislation

Ministry of Education and Science. (2015). Legislative Order no. 17-A/2015, of the 22nd September: assessment in primary and lower secondary education. Diário da República, 2.ª série, n.° 185, 27380 (2) – 27380 (10) (PT).
Ministry of Education. (2018). Ministerial Implementing Order no. 223-A/2018, of the 3rd August: Assessment in primary and lower secondary education. Diário da República, 1.ª série, n.° 149, 3790 (2) – 3790 (23) (PT).
Secretary of State for Education's office. (2016). Legislative Order no. 1-F/2016, of the 5th April: assessment in primary and lower secondary education. Diário da República, 2.ª série, n.° 66, 11440 (3) – 11440 (10) (PT).

References

Barbero, T. (2012). Assessment tools and practices in CLIL. In F. Quartapelle (Ed.), *Assessment and evaluation in CLIL* (pp. 38–56). Como-Pavia: Ibis. http://aeclil.altervista.org. Accessed 1 Feb 2019.
Bentley, K. (2014). *Panel discussion: CLIL issues held at the British Council regional policy dialogue CLIL 2013–2014 policy and practice: Competence-based education for employability, mobility and growth*. http://www.britishcouncil.org/sites/default/files/bc_regionalpolicydialogues_baja.pdf. Accessed 5 Nov 2018.
Bertaux, P., Coonan, C. M., Frigols-Martín, M. J., & Mehisto, P. (2010). *The CLIL teachers' competence grid*. http://tplusm.net/CLIL_Competences_Grid_31.12.09.pdf. Accessed 5 Nov 2018.
Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.
Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66. https://doi.org/10.1080/14703290601081332.
Carless, D. (2009). *Learning-oriented assessment: Principles, practice and a project*. https://www.researchgate.net/publication/242679418_Learning-oriented_Assessment_Principles_Practice_and_a_Project. Accessed 8 Dec 2018.
Clarke, S. (2003). *Enriching feedback in the primary classroom*. London: Hodder and Stoughton.
Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Strasbourg: Modern Languages Division, Council of Europe.
Council of Europe. (2010). *Guide for development and implementation of curricula for plurilingual and intercultural education*. Satellite Study no. 2. Assessment in Plurilingual and Intercultural Education. https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016805a1e55. Accessed 6 Nov 2018.
Council of Europe. (2018). *Common European framework of references for languages: Learning, teaching, assessment*. Companion Volume with New Descriptors. Provisional Edition.

Language Policy Programme. Education Policy Division. Education Department. Council of Europe. www.coe.int/lang-cefr. Accessed 19 Mar 2019.
Coyle, D. (2005). *CLIL: Planning tools for teachers*. Nottingham: University of Nottingham. https://www.unifg.it/sites/default/files/allegatiparagrafo/20-01-2014/coyle_clil_planningtool_kit.pdf. Accessed 6 Nov 2018.
Coyle, D. (2006). Content and language integrated learning. Motivating learners and teachers. *The Scottish Language Review, 13*, 1–8.
Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.
deBoer, P. (n.d.). *How to provide effective feedback part 1 and how to provide effective feedback part 2* [Blog posts]. https://www.clilmedia.com/?s=feedback. Accessed 9 Dec 2018.
Direção-Geral da Educação. (2018). *Direção-Geral da Educação website*. http://www.dge.mec.pt. Accessed 10 Oct 2018.
Fernandes, D. (2014). Avaliação das aprendizagens e políticas educativas: o difícil percurso da inclusão e da melhoria. In M.L.Rodrigues (Org.), *Quarenta anos de políticas de educação em Portugal: a construção do sistema democrático de ensino (Volume I)* (pp. 231–268). Coimbra: Almedina. Retrieved from https://repositorio.ul.pt/handle/10451/16010. Accessed 11 August 2020.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487. Accessed 8 Dec 2018.
Ioannou-Georgiou, S., & Pavlou, P. (2003). *Assessing young learners*. Oxford: Oxford University Press.
Keppell, M., & Carless, D. (2006). Learning-oriented assessment: A technology-based case study. *Assessment in Education: Principles, Policy & Practice, 13*(2), 179–191. https://doi.org/10.1080/09695940600703944. Accessed 7 Dec 2018.
Krathwohl, D. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice, 41*(4), 212–218. Retrieved from http://www.jstor.org/stable/1477405. Accessed 5 Nov 2018.
Maggi, F. (2012). Evaluation in CLIL. In F. Quartapelle (Ed.), *Assessment and evaluation in CLIL* (pp. 57–74). Como-Pavia: Ibis. http://aeclil.altervista.org. Accessed 6 November 2018.
Massler, U., & Stotz, D. (2013). Chapter 6: Assessment in bilingual classrooms. In D. Elsner & J.-U. Keßler (Eds.), *Bilingual education in primary school: Aspects of immersion, CLIL, and bilingual modules* (pp. 73–92). Tübingen: Narr Studienbücher.
Massler, U., Stotz, D., & Queisser, C. (2014). Assessment instruments for primary CLIL: The conceptualisation and evaluation of test tasks. *The Language Learning Journal, 42*(2), 137–150. https://doi.org/10.1080/09571736.2014.891371.
Ministério da Educação. (2016). *Orientações Curriculares para a Educação Pré-escolar*. Lisboa: Ministério da Educação. http://www.dge.mec.pt/orientacoes-curriculares-para-educacao-pre-escolar. Accessed 12 Oct 2018.
Ministério da Educação. (2018). *Aprendizagens Essenciais/Articulação com o Perfil dos Alunos – Inglês 3.° ano – 1.° ciclo do ensino básico*. Lisboa: Ministério da Educação. http://www.dge.mec.pt/aprendizagens-essenciais-0. Accessed 10 Oct 2018.
Ministério da Educação. (n.d.-a). *Implementação do Projeto Ensino Bilingue Precoce no 1.° ciclo do Ensino Básico*. http://www.dge.mec.pt/implementacao-do-projeto-ensino-bilingue-precoce-no-1o-ciclo-do-ensino-basico. Accessed 8 Oct 2018.
Ministério da Educação. (n.d.-b). *Programa Escolas Bilingues/Bilingual Schools Programme*. http://www.dge.mec.pt/programa-escolas-bilinguesbilingual-schools-programme. Accessed 8 Oct 2018.
Ministério da Educação e Ciência ed. (2015). *Metas Curriculares de Inglês do Ensino Básico – 1.°, 2.° e 3.° Ciclos*. Lisboa: Ministério da Educação e Ciência. https://www.dge.mec.pt/sites/default/files/Basico/Metas/ING/eb_metas_curriculares_ingles.pdf. Accessed 10 Oct 2018.
National Council for Curriculum and Assessment. (2007). *Assessment in the Primary School Curriculum: Guidelines for Schools*. http://www.ncca.biz/guidelines/intro.htm. Accessed 10 Oct 2018.

Purpura, J. (2004). Learning-oriented assessments of grammatical ability. In *Assessing Grammar* (pp. 212–250). Cambridge: Cambridge University Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119–144. https://doi.org/10.1007/BF00117714.

The Scottish Government. (2010). Building the curriculum 5: A framework for assessment. In *Curriculum for Excellence* (pp. 1–55). Edinburgh: The Scottish Government. https://www2.gov.scot/resource/doc/341834/0113711.pdf. Accessed 29 Mar 2019.

Wiliam, D. (2008). *Taking assessment for learning to scale*. Paper presented at the OECD/CERI 40th Anniversary International Conference on Learning in the 21st century: Research, Innovation and Policy. http://www.oecd.org/site/educeri21st/40756772.pdf. Accessed 10 Dec 2018.

Wiliam, D. (2014). *Formative assessment and contingency in the regulation of learning processes*. Paper presented in a Symposium entitled Toward a Theory of Classroom Assessment as the Regulation of Learning at the annual meeting of the American Educational Research Association, Philadelphia. http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers.html. Accessed 10 Dec 2018.

Xavier, A. (2016). *Assessment for Learning in EBE/CLIL: A learning-oriented approach to assessing English language skills and curriculum content at early primary level*. Master's thesis, FCSH Nova Lisbon, Portugal. https://run.unl.pt/handle/10362/17973. Accessed 11 Dec 2018.

Chapter 6
Assessment of Young Learners' English Proficiency in Bilingual Content Instruction (CLIL) in Finland: Practices, Challenges, and Points for Development

Taina Wewer

6.1 Introduction

Finland has a considerably long tradition in *Content and Language Integrated Learning* (CLIL), and it is even considered to be one of the trailblazers and birth countries of contemporary European CLIL (e.g. Marsh 2013, pp. 132–133). Teaching and learning subjects and subject areas through two languages has been prevalent especially in larger cities since early 1990s when legislation allowed the practice. Following from the bilingual CLIL approach typical for Finland, this chapter will interchangeably use alongside with CLIL the more descriptive, unambiguous term *bilingual content instruction*. Implementation of and research on bilingual content instruction in Finland can be roughly categorised in three decades with different emphases (see also Wewer 2014a, pp. 44–45) but notable absence of assessment research or focus on teachers' assessment literacy (see e.g. Chap. 1, this volume) until the previous decade.

The first decade, 1990s, was the decade of enthusiastic launching of CLIL mainly based on the Canadian immersion model and supported by the policies of the European Union and Finnish National Agency for Education (FNAE). The first Finnish CLIL study (Järvinen 1999) heralded the subsequent decade, the 2000s, as one establishing the scientific CLIL research in multiple fields, e.g. development in thinking skills (Jäppinen 2003), literacy skills in Finnish language (Merisuo-Storm 2007), classroom discourse (e.g. Nikula 2005), stakeholder experiences (Rasinen 2006), as well as achievement and affective issues such as motivation (Seikkula-Leino 2007). CLIL, under the term *instruction in foreign language*, was included in the National Core Curriculum for Basic Education (NCC 2004). The NCC asserted the dual role of the target language (TL): it is not only the target of learning and

T. Wewer (✉)
Finnish National Agency for Education: European School Luxembourg I, Kirchberg, Luxembourg
e-mail: taina.wewer@gmail.com

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_6

teaching but also the instrument of it (NCC 2004, p. 270). The FNAE passed an Ordinance (25/011/2005) on CLIL teachers' markedly high language proficiency requirements, e.g. minimum 80 credits[1] of university studies in the TL.

The 2010s was the decade of growing professionalism. The body of CLIL research expanded by including new topics, such as communicative competence (Rahman 2012), student experiences and viewpoints (Pihko 2010), teacher beliefs, skills, and characteristics (e.g. Järvinen 2012; Bovellan 2014; see also Chap. 4, this volume), and assessment practices (Wewer 2014a, 2015). The NCC 2004 was revised in a multi-staged, widely collaborative process under the supervision of the FNAE. A new chapter was drafted in the NCC 2014 (2016, English translation) on CLIL, now called *bilingual education*, with the intention of emphasising the involvement of two languages. Additionally, the NCC 2014 differentiated between two extents in bilingual content instruction: *large-scale bilingual education* with the minimum of 25% of teaching and learning through the foreign TL and *small-scale bilingual education* with less than 25% TL exposure, e.g. language showers. The context of this chapter is large-scale bilingual education.

6.2 CLIL in Finnish Basic Education

Finland represents a decentralised education system in which teachers have the authority to make their own decisions and create individual practices. Another characteristic of Finnish compulsory basic education (grades 1–9, ages 7–15) is acting independently and locally within the common national frame. The main principles of education are guided by legislation (the Basic Education Act 628/1998 and Basic Education Decree 852/1998) and the given normative National Core Curriculum in force. The NCC, rather than being highly prescriptive, detailed, and binding, acts as a framework ensuring some national congruence simultaneously allowing, even encouraging, incorporation of local characteristics and emphases. Local municipal and/or school-specific curricula are composed in alignment with the NCC guidelines. Hence, the more detailed local decisions regarding CLIL provision such as the extent, affected subjects and contents, language objectives etc. must be communicated in the municipal and/or school-level curriculum practically drafted by local CLIL teachers with the mandate of municipal authorities, school principals, or the curriculum working group.

Such individualised organisation of bilingual CLIL instruction has resulted in a multitude of ways CLIL is implemented in Finland. Additionally, teacher recruitment requirements and preferences seem to fluctuate at municipal level (see e.g. Kangasvieri et al. 2012; Peltoniemi et al. 2018) regardless of the FNAE Ordinance (25/011/2005) concerning CLIL teacher language proficiency; therefore, varies also

[1] credit – One credit is equivalent of 27 h of study and study-related work in Finnish universities, while 80 credits corresponds approximately to 1,5 years of full-time studying.

the linguistic background of teachers teaching CLIL, the materials and pedagogy used. At primary level, class teachers holding a Master of Art (MA) degree in educational sciences typically teach CLIL, whereas foreign languages are mostly but not exclusively taught by subject teachers. Finnish primary teachers increasingly opt for acquiring double qualifications as both class and subject teacher which enables them to teach both. A point worth mentioning is that even without a formal language education, class teachers with an MA are qualified to teach English and Swedish at primary level due to the few language courses included in the MA curriculum. Another factor contributing to the multitude of CLIL implementations is Finnish teachers' high pedagogical independence, also reflected in assessment practices.

The most common language combination in bilingual content instruction is Finnish and English following the European trend of favouring English in language education (see e.g. Eurydice 2017). CLIL instruction is given separately from traditional foreign language instruction, most often English as a Foreign Language (EFL), which used to begin in the third grade but will be preponed to begin in the spring of the first grade in every municipality from 2020 onward. First grade also is the normal onset of CLIL study with emphasis laid on the acquisition of numeracy and literacy in Finnish or the language determined as dominant in the local curriculum.

6.2.1 Content and Target Language Objectives in Primary CLIL

Learners in CLIL must master the same content defined in the NCC as learners in monolingual instruction in order to become eligible to apply for upper secondary studies, which also explains why content drives CLIL in Finland instead of language. Thus, content learning objectives in CLIL are identical to monolingual instruction unless selected content areas have been excluded from bilingual study by the education provider in the local CLIL curriculum. As to the target language objectives, they must always be considered, determined, and proportioned to the extension of the given TL provision, large-scale or small-scale, in the local curriculum since the two CLIL-recognising NCCs from 2004 and 2014 do not introduce any other than general language learning goals.

The 2004 document mentions the general goal of acquiring "a firmer language proficiency" than can be attained in traditional foreign language lessons (NCC 2004, p. 270). The NCC 2014 states that "an effort is made to achieve solid and versatile language skills" in both languages of instruction, and it mentions long-term goals such as "laying a foundation for lifelong learning of languages and appreciation of linguistic and cultural diversity" (2016, p. 93). The previous NCC 2004 named the conventional four language skill areas of reading, writing, listening, and speaking, and cultural skills as ones in which the objectives for desired

language learning levels should minimally be defined. The NCC 2014 suggests using the proficiency scales and descriptors of the Common European Framework of Reference for Languages (CEFR 2001) as an aid in defining the TL objectives even though they are not discipline-specific but general in nature such as the following example of the second lowest level A2 on spoken interaction (CEFR 2001, p. 30).[2]

> Can answer questions and respond to simple statements. Can indicate when he/she is following but is rarely able to understand enough to keep conversation going of his/her accord.

Noteworthy is that the current NCC 2014 recognises subject-specificity, in other words the academic nature of language needed in content study in different disciplines, and the correct use of language. The document ushers defining language conventions for each discipline (NCC 2016, p. 94):

> The conventions of language use and text types that are typical of each subject should be jointly considered. That way, linguistic objectives are defined for different subjects.

Incorporation of age-appropriate academic language in content study is one of the key characteristics in which CLIL differs from traditional foreign language instruction which concentrates more on social, communicative language, at least in basic education (see, e.g. Wewer 2015 for a comparison). Additionally, the NCC 2014 highlights several other aspects in CLIL instruction: interaction, communicativeness, functionality, student activity in language usage, and sufficient support in different subjects taught bilingually.

6.2.2 Assessment Principles in Finnish Basic Education and CLIL

Identical assessment principles apply to CLIL as to any instruction in Finnish basic education. The task of assessment is to guide, encourage, and enhance learning. The basis for such a view is laid both in the Basic Education Act (628/1998) and Basic Education Decree (852/1998). The Basic Education Act accentuates the multimodality of assessment and enhancing learners' ability to capitalise on self-assessment (628/1998, 22§). The Basic Education Act cements the tripartite nature of assessment by stating that, during the course of study, both the learner and the guardian must receive information on learner's progression of studies, work and behaviour frequently enough (852/1998, 10§). Such an assessment approach represents *assessment for learning* (A*f*L), or *teacher-based assessment* (TBA) as defined by Davison and Leung (2009), and more generally, assessment promoting learning. The function of assessment, thus, is primarily formative; officially summative assessment is

[2] See e.g. The Structured Overview of all CEFR scales for more descriptor examples: https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168045b15e (20.10.2019)

reserved to the very end of basic education, although the year-end reports of grades 1–8 could also be seen as summative with formative usage (see Davison and Leung 2009, p. 397 for discussion). The role of assessment *for* learning has been described in more detail in the NCC 2014.

The NCC 2014 introduced a new term in the Finnish curriculum context schools are expected to develop: *assessment culture* (See Chap. 1, this volume, for general discussion). The NCC describes its pivotal features verbosely, thereby stressing its importance. For example, assessment in the desired assessment culture should be transparent and based on ample, diverse evidence from different sources. It is encouraging and recognises effort. The process of learning (learning to learn) is occasionally considered more important than the actual end result. Learners' role is participatory; they are not objects, but equal partners in the process of assessment, which calls for interaction and dialogical methods. The majority of assessments and feedback should be grounded on reciprocal communication between teachers and learners. Guardians and peers also have an important role.

The aim of assessment is to help learners understand their individual learning processes which, translated to actual learning, should be continuously made visible thereby making the learning curve concretely available for all stakeholders. In other words, learners should be able to recognise their own potential and build on it. High ethical standards, transparency, and fairness are key elements in the assessment culture required. Fairness and transparency are sought by adopting *criterion-referenced assessment* in which learning is mirrored against objectives and criteria that are communicated in both national and local curricula. Assessment *of* learning (see Chap. 1, this volume), study practices and behaviour does not entail comparing or contrasting learners to one another.

Assessment in CLIL must follow these very same ideals and regulations of the Finnish assessment culture presented above. Both assessment foci, content and the TL, must be addressed in CLIL. Content mastery is assessed according to the national criteria available in the NCC. Assessment of the TL proficiency should be specified locally, as language objectives should vary in concert with the extent of language exposure. Neither NCCs do not provide precise guidance on how to organise the TL assessment, but the necessity of carrying it out is unambiguous, as the excerpt from the NCC (2004, p. 273) shows:

> Assessment must give the teacher, pupil, and parents or guardians adequate information about the pupil's language proficiency in relation to the given objectives. Growth in comprehension of a foreign or immersion language is to be monitored, especially when instruction in the foreign or immersion language begins in other subjects.

The current NCC 2014 repeats the same principle but includes also content mastery and highlights monitoring the development of both languages of instruction in relation to predetermined objectives. Both NCCs require versatile feedback on the TL. Nonetheless, in the final report at the end of basic education, the ninth grade, the TL in CLIL must exceptionally be assessed using the same criteria as in conventional foreign language instruction (e.g. EFL) in order to ensure an equal starting point for all when applying for places at secondary level study.

Unlike its predecessor, the NCC 2014 mentions the European Language Portfolio by name as an example of actual assessment methods applicable for CLIL. It also notes that assessment should take place in collaboration between teachers and by capitalising on student self-assessment as a valuable source of information, as the Basic Education Act (628/1998) obliges. Peer assessment is also mentioned as a method to produce assessment information and viewpoints. However, the vast pedagogical autonomy of Finnish teachers in administering assessment includes choosing assessment methods of their preference and/or deliberation. The decentralised education system and long-standing tradition in TBA allow teachers to develop their own assessment practices as long as they observe the regulations and principles in the NCC. This results in a situation where there is no completely unified assessment methodology in Finnish basic education, including bilingual content instruction CLIL.

6.3 Prior Literature and Research on Young Learners' TL Assessment in CLIL

Literature on young learners' language assessment, a subfield of language assessment pertaining to children aged approximately 6–13 (the primary years in Finland), has pinpointed principles applicable as such to language assessment in CLIL circumstances. These principles closely overlap with the characteristics of TBA (Davison and Leung 2009), alternative assessments (e.g. Dochy 2001; Brown & Hudson 1998), and resemble those of the Finnish assessment culture described in the NCC 2014. The list below, inspired by Gottlieb (2012) and Ioannou-Georgiou and Pavlou (2003) and compiled from Brown and Hudson (1998), Hasselgreen (2005), and McKay (2006), provide clues on such principles (see Sect. 6.6; see also Chap. 1, this volume, for further discussion).

- Content assessed must be familiar and the genres used simple enough for the learners.
- Assessors should be familiar adults and the environment psychologically safe.
- Scaffolding and immediate feedback are recommendable.
- Assessment should be varying, multimodal and closely resemble children's typical tasks and activities in the learning environment.
- Personal-response (e.g. portfolios), constructed-response (e.g. fill-in tasks) and selected-response (e.g. multiple choice) assessments, introduced in this order, are appropriate for young learners.
- Assessment tasks should be interesting and contain elements of gaming and enjoyment.
- All stakeholder perspectives (teacher, guardians, and learner) should be included in the communication of assessment information.
- Assessment concentrates more on strengths than weaknesses – it is encouraging in nature.

As stated above, research on language assessment in CLIL has been utterly scarce until the 2010s, and it still cannot be regarded as numerous. Particularly scarce has been the assessment research concerning young CLIL learners, as the following few instances show. Serragiotto (2007) looked into co-teaching of class and language teachers in CLIL and their assessment practices at primary level in Italy and concluded that teachers' background education has an effect on what they tend to emphasise in assessment: content teachers stress content; language teachers are keen on language. In Finland, two young language learner-related studies on any kind of assessment in CLIL have been conducted so far: the present study by the author (Wewer 2014a) and another investigating the use of portfolios in grades 1–3, ages 7–10 (Wewer 2015). Practical CLIL projects have also contributed to the effort of bridging the assessment research gap. In Germany and Austria, the project investigation of Massler, Stotz, and Queisser (2014) looked into assessment methods in primary science and suggested an instrument for structured assessment with descriptors combining the CEFR, content objectives, and the curriculum. Wewer (2014a, b) experimented with the affordances of computer simulations as a language assessment method in primary CLIL. A more recent Greek study by Zafiri and Zouganeli (2017) with some assessment examples concluded that assessment in CLIL indeed is a challenging endeavour for primary teachers.

In Europe, assessment research has mainly focussed on more advanced, older students at secondary or tertiary level. For example, the study of Hönig (2010) discovered that the intended assessment focus of academic achievement may be deviated or distorted, particularly in cases the test taker's level of language proficiency is great or poor. Hence, the test taker's language proficiency may have an effect on the rating of content mastery. Czura and Papaja (2010) have made similar observations in Poland: the separation of content and language in assessment presents difficulties to teachers (see also Poisel 2007; Serragiotto 2007). The double dual focus (two languages and two foci, content and language) both in instruction and assessment makes assessment in bilingual content instruction particularly challenging. To complicate assessment in CLIL further, in addition to the two main foci, assessment in CLIL may also pertain to procedural aspects such as attitude toward learning, communication or practical skills, and learning to learn (Bentley 2010). Following from this, scholars seem to recommend integrated (combined) gathering of data but discrete (completely separate) analysis and assessment of it with separate criteria for language and content (e.g. Barbero & Järvinen 2009; Leal 2016; Massler 2011; Papaja 2014; Serragiotto 2007; Short 1993).

6.4 The Study

The starting point for this study was twofold: (a) the discrepancy between the national norm predicated in the NCC and the actual school-level CLIL practices observed by the author as well as (b) a will to develop language assessment in CLIL. As there were no prior studies or surveys in the field of assessment of young

CLIL learners' target language proficiency in Finland, the purpose of this descriptive and developmental study was to form an overview of assessment in CLIL in the context of grades 1–6 in Finnish primary education and particularly investigate the assessment practices of the TL proficiency. Assessment was defined in the study context taking a wide approach (Wewer 2014a, p. 73):

> Assessment is either the systematic and well-grounded process of information gathering or the product which describes the extent and/or quality of second language acquisition, its degree of correspondence with the objectives of language acquisition and its relationship with the CLIL environment for the purposes of making decisions or judgements about individuals for various purposes.

The wider process and product approach was chosen not to exclude any assessment-related phenomena as the field was an unexplored territory.

6.4.1 Methods of the Study

The three-phased, mixed-methods study was conducted in years 2012–2013 and published in the transition period between two National Core Curricula (2004 and 2014), while the NCC 2004 was still in force and the draft version of the new NCC 2014 nearly finalised. This chapter will focus on the first of the study phases, the *assessment survey*, which will be unfolded in respect of teachers (n = 42) only, although also pupils and their parents were included in the study. The interested reader may wish to scrutinise the perceptions and experiences regarding the adequacy and sufficiency of language assessment and feedback of CLIL pupils (n = 109) in grades 3–5 from two different subject schools and their parents (n = 99) in Wewer (2014a). The set of research questions concerning teachers was the following:

1. **How is assessment of English language proficiency in CLIL organised according to class teachers?**
   - 1.1 What kind of assessment practices do CLIL class teachers employ for assessment of English language proficiency in CLIL?
   - 1.2 To what extent do CLIL class teachers assess language and provide pupils and their parents' feedback?
   - 1.3 What kinds of challenges are related to assessment?
   - 1.4 How should assessment in CLIL be developed?

The assessment survey consisted of a semi-structured questionnaire and theme interview for volunteers (n = 12). The web-based teacher questionnaires were sent out to CLIL providing primary schools all over Finland listed at the CLIL Network website available at that time. Furthermore, teachers in both subject schools, designated by the convenience of location and availability, completed paper versions. Both schools, situated in South-West Finland, had offered large-scale bilingual content instruction (scope >25%) for a longer period of time, and the number of CLIL

pupils was approximately the same. As for the teacher backgrounds, qualifications, composition of pupil material, extent of curriculum, resources, etc., the two schools differed from each other. The questionnaires were identical in both formats, and it was underlined that the questions concerned assessment of English in disciplinary subjects to ensure that CLIL assessment practices are not confused with those of EFL.

The questionnaire was designed to target teachers' practices in assessing the TL in CLIL subjects and to look into their underlying views and beliefs about assessment which possibly were guiding the practices. Teachers were also asked to identify pitfalls in assessment in CLIL and outline plausible future assessment policies and practices. In addition to these main themes, teachers' background as CLIL teachers (teacher and language qualifications and experience) was determined. The questions were open-ended, or included an open field for elaboration. A few multiple-choice questions were included. The questionnaire did not differentiate between summative and formative assessment for two reasons: (1) the norm in Finnish primary education is that all assessment is teacher-based and formative except for the final report at the end of basic education, and (2) any kind of assessment, feedback and report practices were considered interesting and valuable.

In order to receive more detailed information on the research topic, participants were invited to take part in a thematic, audio-recorded interview. A theme interview is guided by themes referenced in every interview, but they are likely to appear in different order, and their depth of handling as well as the style and formulation of questions may vary (Hirsjärvi & Hurme 2010, p. 45–46) to keep the situation natural and informal. The conversations started with a general reflection upon assessment in CLIL and the necessity of it. The subsequent topics revolved around the grand themes of official CLIL practices of the school (local CLIL curriculum and the quantity, frequency, and quality of assessment in CLIL), assessment practices and methods of the given teacher and their colleagues, issues impacting the assessment methods used, how to make pupils' learning curve visible, challenges and problems in assessment, and future visions.

Altogether 12 teachers out of 42 volunteered to be interviewed on their own time, which suggests that the topic was regarded as important. Eventually, 10 teachers were interviewed producing over 8 h of recordings transcribed at general orthographic level without false starts, hesitations, and alike irrelevant features. The collected data, the questionnaires and transcribed interviews, were analysed both quantitatively by calculating frequencies and percentages and qualitatively by carrying out thematic content analysis in which the textual data, through intense reading and occasional clarification requests, was reduced and categorised into recurrent themes and patterns identified in the data body.

6.5 Main Results of the CLIL Teacher Assessment Survey

All 42 teachers held the minimum basic education teacher qualification in Finland, the MA degree, but only 10 of them fulfilled the language requirements of the FNAE Ordinance (25/011/2005). A fifth of them (9/42) had undertaken no language studies of any kind which corroborates the notion made by Nikula and Järvinen (2013): there is a new teacher category in Finland – one teaching a foreign language without being a trained language teacher. Furthermore, CLIL class teachers' TL background varied vastly from basic studies in English language worth 25 credits, double qualifications as both class and English subject teacher to teachers who had earned their qualifications abroad, had subject teacher qualification only, or had passed a university programme preparing for CLIL instruction.[3] Their experience in CLIL varied as well: 26% of the teachers had taught 15 or more years in CLIL, four of them only 0–2 years, and the rest a varying amount of years in between. As for the age groups instructed, almost half of the teachers were teaching grades 1–2 (ages 7–8); the rest were divided rather evenly between grades 3–4 and 5–6.

Most teachers recognised the importance of TL assessment. Almost half of the teachers (20/42), regardless of grade level, considered it either highly or very important, 16 of them rather important, while the rest (6/42) did not place any significant value to the assessment of the TL in CLIL subjects. Teachers' perceptions of importance were in contradiction with the findings indicating that assessment of the TL in CLIL subjects was not an established practice. The survey results pointed out that language assessment appeared to be "infrequent, incidental, implicit and based on impressions rather than evidence or the curriculum" (Wewer 2014a, p. iii). The main results supporting this claim are grouped under three main headings: 6.5.1 Assessment and Feedback Practices, 6.5.2 Challenges in Assessment, and 6.5.3 Development Points and Future Visions. The findings are introduced with selected quotes from the questionnaire comments or interviews translated into English from the Finnish original not included here due to lack of space. The results are followed by a discussion and a set of recommendations.

6.5.1 Assessment and Feedback Practices

One fifth (9/42) of teacher informants representing all grade groups without specific scatter disclosed not assessing pupils' English proficiency by any means. The *practice of not assessing the TL* appears to be teacher-dependant and related to the lack of proper linguistic training, as the majority of such teachers had no or minor background studies in English. Another reason may lie in the fact that first of all, not all

[3] The JULIET Programme has been offered by the University of Jyväskylä as a minor since the 1990s. For more information, see https://www.jyu.fi/edupsy/fi/laitokset/okl/opiskelu/sivuaineet/juliet/en/intro/what (20.10.2019).

municipals or schools have composed a local curriculum for CLIL, i.e. defined language objectives for reference, and second, they have not developed common assessment and feedback practices for CLIL. The latter point is clearly delivered in the first quote below by a teacher with extensive CLIL experience in a school with one of the longest CLIL provisions in Finland. The second quote from a different teacher reveals how an ambiguous CLIL curriculum leads to teacher-driven, eclectic solutions. The third quote, yet from another teacher, explains the lack of assessment by the energy put into the inception of CLIL.

> Actually, during this spring, I have become aware of the necessity of it [language assessment in CLIL]. Firstly, regarding the child perceiving his/her own language proficiency but also regarding feedback for parents stating where we are going in the language development. This [awareness] has awakened in the discussions with parents because, at the moment, we don't have it [assessment in CLIL] in our school and the decision-maker [principal] has not taken a positive attitude towards it. Regardless, children, already when little, and their parents, have the right to receive feedback on language acquisition.

> In our curriculum, we just have sort of frames, what CLIL is, but we haven't separately and explicitly defined what should be taught in English in different grades. It is a decision for teachers to make. They consider class-specifically which project or theme includes English. So that varies year in, year out.

> It [assessment in CLIL] is an issue we haven't had the energy to tackle so far. We have just developed CLIL; we have not had the time and energy; we have not been able to go into it.

The lack of proper CLIL curricula was corroborated by several teachers from different schools as well as noted by the researcher in her investigations, as curricula are documents openly accessible for anyone interested, mostly also online. However, some teachers represented schools in which the local CLIL curriculum was compiled in an exemplary manner according to the NCC guidelines, and they also executed a detailed assessment plan common to the municipality or school. In the following subsections, the key results concerning self-reported assessment methods (6.5.1.1), means of giving feedback (6.5.1.2), and frequency of feedback (6.5.1.3) will be introduced.

6.5.1.1 Assessment Methods

The overall result of the most frequent assessment methods reported by teachers, i.e. means of gathering information for assessment purposes, is shown in Fig. 6.1. According to the study, traditional assessment methods such as teacher observation (86%), testing in different formats (69%), dialogic interaction (59.5%) and self-assessment (43%) were the most frequent.

Teacher observation as the most frequent method of gathering assessment data echoes the Finnish assessment culture favouring softer approaches instead of excessive or high-stakes testing almost completely non-existent in Finland. It seems that teachers pay attention to learners' use of language and make use of observation opportunities "on-the-run" (McKay 2006, p. 141) in the classroom thus forming

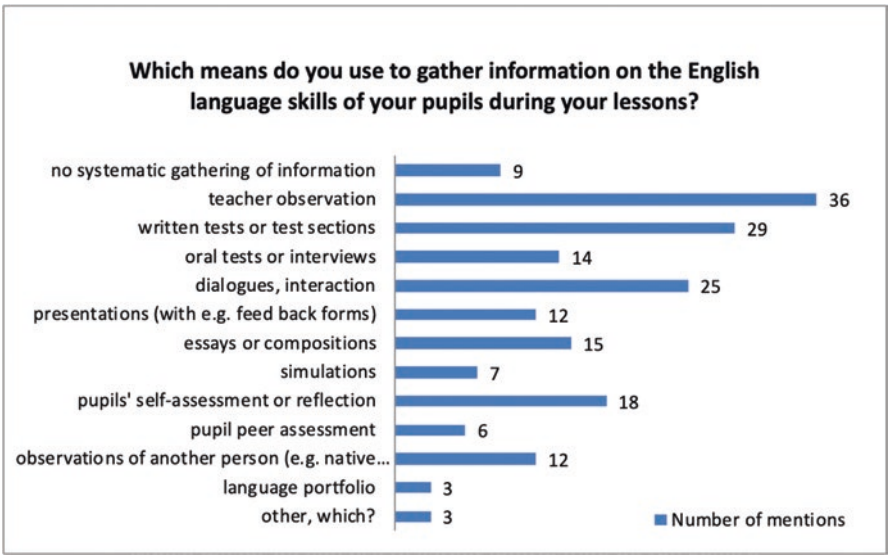


Fig. 6.1 Means of gathering information of TL proficiency in CLIL contexts. (Adapted from Wewer 2014a)

"mental records" (Wewer 2014a, p. 133) of pupils' language use and proficiency. However, it remained fairly unclear how teachers further used such mental records. The gathering of data is the necessary starting point, but making inferences or judgements, or giving feedback based on the data gathered concludes the process of assessment (see the definition on p. 195; cf. the definition of assessment in Sect. 1.1 of Chap. 1, this volume). One must not forget that according to Finnish practices, there should be three parties involved in assessment *for* learning: the teacher, learner, and guardians. Reporting to or discussing observations on learning with the stakeholders does not make assessment final or summative since the intention is to improve learning; rather it can be perceived as having both of these functions. Furthermore, assessment is partial if only one party, the teacher, is involved.

The danger is that a partial assessment process, not including the full assessment cycle as discussed by Davison (2008; see Chap. 1, this volume), does not result in enhanced learning, or the learner does not become more aware of his/her own progress as the gathered information stays with the teacher only. There were signs of such a conduct also in relation to observations made by other persons, called *externalised observations* in the research report. The practice of gathering *external observations*, for example those of native speakers, or language enrichers with various backgrounds, were present in the data, although not amongst the most frequently used. Native speakers and alike were viewed as 'proper language models or experts' by teachers, and their role was to support the class teacher in the enhancement of spoken language and create authentic language use situations by eliciting the TL in small groups. The sessions or samples of children's language use were periodically captured, as an interviewee describes in the following quote:

> The native teacher records everybody [talking]. […] It is a system that has been working in this school at least the while I have been here. I thought, then, that this would be a nice way to assess, to gather a sample from each pupil during six years. […] I don't know if the teachers get to listen to what has been recorded. But I do that out of interest and I have files for grades 1, 2, 3 and 4, and just a while ago I listened to them.

The teacher used to give these recordings to pupils when they moved over to lower secondary after the 6th grade, but they were not used for assessment *for* learning or feedback during the primary years.

> For the first time this year, children asked what's going to happen with these recordings. For the first time: 'What are they for?' Now they would like to listen themselves. […] It would be nice that I'd let everyone listen to their own production here in the middle [of primary years] because they are interested in them now.

The excerpts reveal that externalised observations, i.e. the practice of recording language samples, were not carried over to actual feedback by neither the teacher nor the native speaker.

It has been noted that in Finland, instruction in primary CLIL still appears to be fairly traditional and teacher-centred (Bovellan 2014). It is therefore not surprising that the teaching habits are reflected in teachers' self-reported assessment practices. The second most used assessment method was *testing* in diverse forms. Examples of such were weekly spelling tests, word tests in different content areas, and bilingual content tests partly English, i.e. traditional paper tests with sections in the TL. The following testing principles were distinguished in the data:

- Teaching through English is the prerequisite for giving tests or test sections in English. What has not been taught through English cannot be tested through English.
- Code switching and choice of language is allowed.
- Use of English is rewarded.
- Teachers are willing to make favourable interpretations of test answers and disregard minor language errors.
- Teachers attempt to unify scoring and grading within grade levels and school-internally.
- Content is considered to be more important than language.
- English is adopted for less demanding sections in the test.
- Language support in pre- and post-test situations is provided.

Teachers also seemed to put emphasis on interactive language use, and they attempted to create diverse TL use situations such as projects, group work, situational language use (e.g. English only during lunch), performance evenings, pedagogic drama, talks, presentations, and interviews. *Dialogic interaction*, as it was designated in the study, was the third most mentioned means of gathering information on children's language skills, presumably for teacher observation purposes. Input of subject-driven, academic-type of language seemed to be absent. Production of spoken language was naturally emphasised in earlier years. The fact that nearly a half of the participating teachers instructed children in the beginning of their school

path probably affected the finding about assessment methods used – or not used, as the following quote illustrates. The teacher justifies the practice of ignoring language assessment with affective factors, children's young age and their individual ways of learning, as well as her own views of CLIL.

> They were first and second graders, little pupils, so I didn't test it [language] in any way. In the primary school where I have mainly worked, to my view, it's about building passive language proficiency. And many are slow in starting to speak and produce. I'd like to think that we create a positive attitude towards the language and learn to use it without measuring too much. And then, another thing is that pupils are different language learners, so encouraging is very important. It might be awkward to notice that "I haven't learnt at all" and compare a lot.

In unravelling the data, it became evident that teachers' personal perceptions of CLIL, its premises, and methodology had an influence on how they approached to assessment in CLIL. Three approaches or foci were differentiated in this study; the quote above represents eclectic view (see Sect. 6.5.2.1 on the foci in this chapter).

The fourth most frequent assessment method preferred by teachers was *pupils' self-assessment*, which was slightly surprising, especially since availing self-assessment as an aid in enhancing learning is prompted in the Basic Education Act (1998) and again stressed in both NCCs. A small number of teachers expressed their view of self-assessment being 'an assessment fashion' that has lost its popularity, whereas in some schools, self-assessment in CLIL was an integrated part of the school year reporting system.

The least used assessment method by teachers was the *language portfolio* with three mentions, although the concept of the European Language Portfolio as an offspring of the CEFR (2001) has been promoted by both European and national steering instances. Among the less used methods were also *simulations, peer assessment*, and *graded presentations*. A curiosity and common feature in less preferred assessment methods is their alternative nature as opposed to what are seen as traditional assessment methods (see e.g. Dochy 2001; Brown & Hudson 1998; Wewer 2014a). This calls the serious question of whether or not teacher-based, formative assessment is genuinely used for the purpose of enhancing learning.

6.5.1.2 Means of Giving Feedback

The study revealed that not all teachers offer learners feedback on their TL proficiency or its development in CLIL circumstances. One reason to this may be that they do not associate assessment of language to CLIL but rather to EFL, another reason their being oblivious of or indifferent about the NCC requirements. According to teachers (79%), *oral feedback* was the most common means of feedback. It was most often provided through passing comments and praises during lessons, or it was corrective in nature. The second most frequent method was *written feedback* (52%)

in relation to tests, assignment papers, or Wilma.[4] *Development discussions* (teacher-parent-pupil conferences), typically held once a year, were the third most common feedback method chosen by teachers (45%). The survey indicated that regardless of this, the TL in CLIL is not always on the conference agenda, as the allocated time is spent on more urgent issues (e.g. behaviour), and the format of the development discussion fluctuates. Other non-formal one-to-one discussions with parents or pupils were clustered together with development discussions. The next most often used feedback method by teachers (33%) were general *whole-class discussions* during which relevant or recurrent language issues are addressed.

6.5.1.3 Frequency of Feedback

The results indicated that teachers generally did not further convey the assessment information they gathered to pupils or their parents: 62% of 39 teachers stated giving linguistic feedback rarely or not very often. This is in line with the finding of learners' perception of not receiving feedback on their TL proficiency and development: only 8% of the participating pupils ($n = 109$) felt that they obtained frequently feedback on their language proficiency and its development; 63% of them would have liked to receive more feedback on their language development and skills, especially directly from the teacher in form of oral feedback. The same wish applied to parents ($n = 99$) of which 76% would have liked to receive more information on their children's language proficiency and its development particularly in teacher-parent conferences.

Teachers' own temporal descriptions of their feedback practices were categorised into three different time spans: feedback given (1) fairly or very often, (2) regularly but not often, and (3) rarely. Additionally, miscellaneous answers were differentiated. Examples of teachers' characterisations in Table 6.1 also elucidate examples of assessment methods rather than explanations on how exactly feedback is given to pupils. Furthermore, various stances toward assessment in CLIL can be detected and interpreted in the comments, ranging from negative to nonchalant, and from more rigorous practices to equalisation with EFL. Section 6.5.2 concerning challenges in assessment will address the outlook of CLIL teachers to CLIL and assessment in it.

The situation remained rather constant concerning parents, as a fifth (9/42) of teachers stated not providing parents with any kind of feedback on their children's coping with English in CLIL studies, five did that regularly, and the rest (18/42) occasionally, or rarely (10/42). A collection of teachers' portrayals of their feedback practices catering for parents are in Table 6.2.

Interviews and teachers' free wordings of feedback practices disclosed approaches designated as *indirect and direct feedback* in this study. *Direct feedback*

[4] Wilma is a commonly adopted digital teacher-learner-parent user interface in Finnish schools. It also includes a feedback system.

Table 6.1 Frequency of teacher-given feedback on pupils' TL proficiency and its development

Rarely (n = 15)	Regularly but not often (n = 9)	Fairly or very often (n = 12)	Miscellaneous (n = 3)
Rarely, because in my opinion, the English subject teacher takes more care of assessment	Several times during the semester: always after a composition, performance or test. Also, in connection with reading, listening comprehension is checked.	I give oral feedback on a daily basis. Quarterly, a written assessment form also including a self-assessment section.	I give feedback on English language proficiency almost solely in connection with formal language instruction. In my opinion, the most important thing in CLIL instruction is the mastery of contents, not language. The language is only a tool which each pupil uses according to his/her skill level.
When thinking of my first-graders, not at all. We're just in the vocabulary input stage, so actual testing is a restricted procedure only. In the 6th grade the tested language proficiency was graded with numbers and also partly by verbal representation, and I conducted the development discussions with pupils.	Mainly when returning tests or during the lessons when you can see that pupils understood the subject matter. In other words, regularly irregularly. In connection with returning tests, the feedback on language proficiency level or skills is on a general level, nobody's individual performance is under scrutiny.	Every week spelling tests, 2nd graders write 10 sentences of spelled words. Non-formal almost daily, approximately every second month another written test. On everyone's weekly turn of show–and–tell type of oral situation. Once a year an oral situation one at a time with the teacher during the break.	It depends on the pupil and his/her level of development. When a weak pupil is involved, I give feedback more often and on a general basis. In the case of an advanced pupil, I give general feedback more rarely and concentrate on the elaboration of language proficiency and giving more detailed feedback.
Not often and not regularly	when we finish each section in a textbook	oral feedback on weekly basis, in written form related to returning tests and monthly briefings	It is not relevant to my teaching.
A couple of times per school year, not systematically	at least every fourth week	daily, weekly feedback in the classroom	

Based on Wewer (2014a, p. 147)

involves one-to-one communication and decoding of the assessment information, normally initiated and gathered by the teacher, whereas *indirect feedback* shifts the responsibility of data interpretation to the parents who rarely are linguistically trained and not always capable to understand the TL enough to draw educated inferences from the materials presented. Furthermore, the language objectives and

Table 6.2 Teachers' descriptions of feedback practices serving parents

Regularly (n = 5)	Occasionally (n = 18)	Rarely (n = 10)
Once a year a development discussion; in connection with tests, projects and other productions self-assessment section and teacher's comment.	Every result on tests/ exercises goes to an electric system which parents can check. At least once a month we have assessed something.	In connection with development discussion I mentioned to the parents of competent pupils that it is going well – 1st grade in question.
Once a week, information on pupils' school work goes home in individual pupil diaries. English assessment is only a part of a larger weekly assessment.	weekly word tests, larger tests after each study unit and term report cards twice a year	Always in developmental discussion, i.e. at least once a year. When applicable, I enter a mention of positive feedback into Wilma
They see what I have written in children's production and homework, what I have written when checking their notebooks and such, spelling tests every week. Then weekly bulletins through Wilma and reminders in case someone forgets homework etc. Sometimes in their 'home-school communication notebook', but nowadays Wilma is so handy that almost all communication back and forth goes through it electronically and at least once a week. Additionally, I write general things via the school website.	Twice a year within formal report cards, but tests (also those of other subjects carried out in English) are always sent home and parents have the possibility to follow the development. I often ask parents to sign compositions etc.	Parents have the possibility to attend their child's development discussion twice a year. Additionally, we assess the language proficiency of the child very closely together with the parents before enrolling into the first grade.

Based on Wewer (2014a, p. 148)

contents may be unfamiliar to the parents. Some teachers appeared to think that the production and provision of evidence for parents qualifies as feedback; they seemed to leave the interpretation of the child's linguistic development and attainment level to the parents. The quotes below mirror such a view.

> I surely expect that parents monitor it [language proficiency] and make observations. If the parents are active and want to get information, they can open those notebooks [written in English] and investigate where we are going and how [proficient] the child is. I think that activity from the parents' part is important. I don't see that teachers' work load is added with reporting accountability on English language proficiency towards parents. I'd rather pass the ball to parents so that they can monitor. And if English has been used in tests, so there it comes as well.

> We are trying to bring forth this assessment of language proficiency, but there's no going around the fact that parents see already in the [test] answers what's the level and what the situation is. […] But my opinion is that if the parents would bother to read the notebooks and look what the children are doing, that would tell them what the level of language proficiency is.

> I don't know how much they get [information] elsewhere than in the development discussions and seeing how successful the tests were. I have to admit that it is one thing I should do more, to inform parents, but I can't do everything.

One teacher assumed that parents would not consider feedback as very important or relevant. Making assumptions about parents' capabilities and preferences may be deceptive, for in this study, parents and children alike clearly indicated a wish to receive more information on the linguistic level and development in CLIL subjects.

6.5.2 Challenges in Assessment

The study bared several fundamental problems and challenges related to assessment in CLIL. The most crucial of them, affecting the successful implementation of CLIL and also assessment in CLIL, is the absence or haphazard compilation of a local CLIL curriculum, as already shown above in the previous Section and the following quote exemplify.

> When I don't know the objectives quite clearly myself either, then also the assessment is alike; sometimes I think to myself, what's the point in this. It is obligatory to assess; we [teachers] should talk it over together and figure out what we are going to require of them [learners]. […] It would clarify the situation when we could show that 'these are the minimum objectives; these should be covered'. Then, assessment would be much easier. Now it is really hard in my opinion.

The lack of learning objectives leads to teachers drawing and following their own shadow curriculum. The following comments depict such a situation.

> It is challenging that there are no common criteria in setting the objectives. The teacher sets the objectives him/herself – comparison with the level of EFL pupils is hardly helpful.

The lack of the CLIL curriculum, equalling the lack of a solid foundation for CLIL, manifests itself in the ambiguity, fluidity, and multiplicity of assessment beliefs, practices, and approaches. However, ensuring that a CLIL curriculum is drafted is an administrative task. Teachers are not to blame for lacking working tools. The following two passages are from teacher interviews which allowed dwelling on the topic area in more depth. One teacher felt distressed in the interview, and she voiced concerns quite forcefully, advocating for a shared assessment plan.

> Agonising! There are no clear instructions on how to assess [the language]. We don't actually have any objectives either. I don't know what, in which subject or how much. But I can't know that when I don't know what I'm assessing. And the assessment varies; we have talked with one another. It is like night and day.

Another teacher defended the right for individual decisions and deviating practices equally forcefully.

> I would say that it [assessment in CLIL] varies tremendously from teacher to teacher. We have such a great freedom in acting as a teacher, also in assessment. […] [A]lso in these CLIL issues the teacher has to have the freedom [to assess] according to how s/he perceives

> it, what kind of an understanding s/he has of learning and assessment and so forth. It [assessment in CLIL] is based on that.

Another pivotal problem is that teachers do not always seem to be aware of the requirements of the NCC, or they have not familiarised themselves with the document. For example, the expectation of monitoring and assessing language in CLIL as well as provision of feedback was distinctly announced in the NCC 2004 in force at the time of the study, which did not come across in some teachers' responses. When teaching and assessment are not directed by the national norm, it is more likely to become guided by one's own beliefs and preferences, as exemplified in the teacher quote below. This teacher taught both EFL and CLIL to her class.

> I have to say that when I answered the questionnaire, it occurred to me for the first time that [the language] could also be assessed, the actual CLIL performance within every subject. […] In that sense, I have not assessed the language in different subjects. The language use has been assessed in general in the formal English grade.

Teachers raised questions about the proper placing of the TL assessment when grading. Should the TL be assessed separately from content or together with content as integrated? Or should the TL in CLIL be included in the EFL grading? These questions will be addressed in Sect. 6.6. In addition to the issues covered above, teachers named a number of complications hindering or challenging the assessment of the TL in CLIL: lack of time, diversity of pupils, laboriousness of assessment, and lack of assessment tools. Assessment of oral skills was often regarded as particularly difficult. The data also revealed unexpected results explicated in the following section.

6.5.2.1 Three Foci for the Role of Language

In addition to the above-mentioned challenge of the non-existing CLIL curriculum, objectives, common assessment scheme, and the use of indirect feedback (see Sect. 6.5.1), the analysis of the data disclosed three different foci for and perceptions of the role of the TL in CLIL which ultimately have an effect on how teachers approach the assessment of the TL: (1) *instrumental*, (2) *dual,* and (3) *eclectic focus*. The *instrumental focus* refers to the view according to which teachers see the TL as an instrument or tool that is implicitly used to achieve the goal of learning content, and language is an incidental side product in the learning process. Language is not explicitly taught and therefore not assessed. The instrumental focus also entailed parallelisation of EFL and English in CLIL. The two quotes below represent the instrumental focus on target language.

> In my view, it is not necessary to assess language skills in, for example, mathematics, environmental sciences etc. Assessment of language proficiency takes place in formal [English] instruction. I don't want that learning school subjects transforms into pure language study. Language is only a means of studying, and every pupil uses it according to his/her own proficiency level.

> Understanding the content matter is always the number one thing for me; language is a bonus which is learnt through the content.

The *dual focus* on the TL follows the current and most common CLIL definition, also promoted by the NCC, in which the double role of language is highlighted: language is both the medium and target of learning, an end itself in a similar vein as content. Formal EFL instruction is seen as different and separate from CLIL English; the former advances more social, communicative language proficiency, whereas the latter more academic, disciplinary language skills. Language-awareness is associated with the dual focus and showcased in the following quote.

> In other words, I may stop in the middle of a lesson and just remark that now we have used this kind of expression or something else. Sometimes, in subject lessons, that may generate little streams so that they form a language flash.

The *eclectic focus* encompasses the remaining approaches to language, most often stressing affective factors such as motivation and individualism in learning languages, or courage in using the TL which is clearly displayed in the two excerpts from the questionnaire. Educators in the very first grades represented this focus more often than teachers instructing older learners.

> Personally, I would like emphasis to be put on brave use of even elementary language. Language is nonetheless used only as a tool for learning in which case it is not, in my view, even meaningful to assess the level of language proficiency in the first place, not at least in elementary instruction. It would be more important to assess whether the child has the courage to use even that little amount of language s/he knows.

> It is more essential to awaken interest and keep up enthusiasm. That is something you can assess with gut feeling when you've had the same group a long time.

A moderate connection can be discovered between the three foci identified in the study and the four quadrants matrix by Leung and Morton (2016) and the types of CLIL learning drawn from it discussed in Chap. 1 of this volume. The instrumental focus resembles the competence-based approach to language, while the dual focus would refer to the higher disciplinary orientation coupled with subject literacies taken more visibly into account in instruction. The eclectic focus in turn reminds one of situated learning with lower orientation toward both subject and language learning. As CLIL in Finland is more often than not guided by content, an equivalent to the fourth quadrant was not present in the data.

6.5.3 Development Points and Future Visions

Logically, most ideas teachers presented as visions and ideas for improvements were directly connected to the problems they had faced as CLIL teachers. Some teachers, however, were pleased with the operational model and practices as they were, and felt no need for any particular assessment methods or tools. Such teachers often represented the instrumental focus to the role of language in CLIL (See

previous Sect. 6.5.2.1.). The majority was apt to consider assessment in CLIL from a new perspective and propose ways to develop it. Particularly the following developmental themes were the most frequent.

1. Uniform foundation – Teachers were asking for a CLIL curriculum with more specific objectives from which it would be possible to draw assessment criteria.
2. Coherence – More pedagogical coherence can be achieved through ensuring a uniform foundation. Coherence is generated through adapting common assessment tools, methods and practices within the school, yet leaving room for individual ideation and creativity. Coherence carries obvious implications to learner equality and equity.
3. Basic vocabulary – Vocabulary commonly taught for all CLIL pupils also contributes to coherence and was seen as advantageous in so far that every student would master the key glossaries of CLIL disciplines on which it is more convenient to build new vocabulary and content knowledge through the TL.
4. Establishment of various tests – The accumulation of a test bank was suggested in hopes of creating and sharing testing tools. Despite the fact that Finland employs non-accountability assessment practices only, some teachers were prepared to adapt some form of standardised testing in CLIL. One needs to bear in mind that standardising tests has a different connotation in Finland. It merely refers to improving test reliability, comparability, and validity in larger sense; standardised tests are seen as applicable to the whole cohort of CLIL learners in an age group and certain subject.
5. Use of the CEFR proficiency levels – The CEFR levels and proficiency descriptors were mentioned by a few teachers only which implies that they are not widely used or known among CLIL teachers. Adaptation of the CEFR taxonomy, as is recommended in the NCC 2014, would also afford more congruence for assessment and provide a firmer, Pan-European baseline not only for proficiency level evaluations, but also for linguistic goal setting and pupil self-assessments. The problem is, however, that the CEFR is not discipline-specific (see Chap. 2, this volume, for an adaptation proposition).

A possible future assessment trend that emerged notably from one teacher's comments and visions were participatory assessment methods (e.g. group tests, collaborative testing) perceived as meaningful in reducing stress, incorporating various working skills, harnessing each group member's strengths, and compensating weaknesses. After all, an assessment situation can also be an opportunity for learning. Additionally, technology-driven methods for assessing oral communicative competence were innovated by teachers and revealed in the study.

6.6 Discussion

The chief aim of the study was to gain an understanding of how TL assessment is organised in primary-level bilingual content instruction in Finland, verify what kind of assessment and feedback practices and methods are used, and what is considered to be challenging or desirable in language assessment. Also, suggestions for improvement were gathered. The results demonstrated that target language assessment in CLIL is not an established practice and the field is scattered both ideologically and methodologically. The generality of particularly the 2004 National Core Curriculum for Basic Education, provision of various types of CLIL in Finland as a local decision combined with almost untouchable pedagogical and methodological independence of Finnish teachers contribute to miscellaneous implementations of CLIL which in turn leads into various assessment practices and views on assessment. However, teachers seem to strive for congruence. It appears that there are several structural factors causing uncertainties and even transgression in teachers which is why the foundation of CLIL needs to be strengthened.

These structural factors, called CLIL fundamentals, must be urgently and profoundly elaborated to ensure a thriving CLIL provision in Finland. The 10 fundamentals are: (1) existence of a CLIL curriculum and language objectives, (2) solid teachers' CLIL knowledge base, (3) adaptation to the NCC 2014 reform, (4) reinforcement of teachers' assessment literacy in CLIL, (5) clarification of the role of the TL and its connection to assessment, (6) shifting from implicit language learning to dual focus thus giving more attention to language besides content, (7) seeing CLIL English widely as academic and complementary to or inclusive of EFL, (8) clarifying the assessment approach, (9) ensuring linguistically qualified teacher force, and (10) move from implicit to evidence-based assessment and feedback. One of most important factors likely to create a positive impact on assessment in CLIL is teachers' proper linguistic training, language assessment included. The most crucial basic factor is the proper CLIL curriculum.

The absence of a proper language curriculum in CLIL is inarguably one of the most significant shortcomings among those detected in the study, as it is contrary to educational principles and the instructions of national steering. In many cases, there were no language objectives for CLIL, or they were loosely formulated. Following from the lack of curricular foundation, there was no reference for language assessment available. One could indeed question the rationality of CLIL provision without any language goals present (Dalton-Puffer 2007, p. 295). It may be tempting to replace objective setting by following the NCC definition of large-scale bilingual instruction and generally defining that, for example, 25% of instruction takes place in the TL, but this is not the spirit of the NCCs. A language curriculum is necessary, as discipline-specificity is one of the key aspects of CLIL. The academic nature of language in CLIL study needs to be taken into account already from the very beginning, although the language is still simple and mainly social.

General frameworks such as the CEFR, although helpful in other ways (see Chap. 2, this volume), lend little support for developing discipline-based language

objectives for young learners. The work made in the USA by the WIDA consortium[5] might help CLIL curriculum designers write academic language objectives. WIDA's Can Do descriptors[6] provide examples of language usages at six different levels for basic academic language functions of recounting, discussing, explaining, and arguing. Such descriptors can also be adapted to different school subjects and their content areas with relative ease. The WIDA English Language Development Standards and Resource Guide for levels from Kindergarten to 12th grade (2014) published for the international audience is even more detailed a presentation of English language development in different disciplines. It is as such worth exploring for the compilation or editing of the CLIL curriculum and improvement of assessment.

Assessment of the target language in CLIL seemed to evoke distress and confusion in many teachers mostly due to the lacking CLIL curriculum. Teachers cannot be criticised of such a situation. However, being oblivious of the regulations of the NCC or even deliberately disregarding them cannot be excused. Assessment was mostly seen as important and relevant by teachers. Regardless of the perceived importance of assessment and the NCC guidance, not all teachers carried out any kind of language assessment, and the assessment information gathered by teachers did not readily become translated into actual feedback. Due to the lack of language objectives, common understanding, or more extensive knowledge of CLIL, teachers seem to base their assessment practices, or the lack thereof, on their own beliefs and preferences.

Teachers' three different approaches to the TL, instrumental (language as a tool only), dual (language both tool and target), and eclectic (other, e.g. affective factors foregrounded), identified in the study appear to have an effect on their approach to assessment in CLIL, and their reluctance or willingness to focus on the TL in instruction and assessment. Teachers with instrumental and eclectic focus to language were less inclined to assess language than those recognising the role of language also as a target of learning (see also Chap. 9, this volume, on secondary CLIL teachers' outlooks on assessment). Hüttner, Dalton-Puffer, and Smit (2013) have also reported on partially resembling findings related to teacher beliefs on explicit and implicit language learning in Austrian tertiary CLIL. Additionally, Serragiotto (2007) in Italy has concluded similarly in respect of CLIL team teaching.

The instrumental focus is likely to have its origins in the dawn of the European CLIL, taking influences from the Canadian immersion, and avid experimentations of bilingual instruction without proper knowledge of its tenets in the first decade of Finnish CLIL. Whilst the scholarly and national view of the role of language in bilingual education has since moved toward dual focus, the 'language is a tool only' mantra has persisted among Finnish CLIL teachers and occasionally in schools' official documents. As a consequence, particularly many pioneer teachers in CLIL do not see the relevance of paying attention to TL or assessing it. Language is seen

[5] See https://wida.wisc.edu/ (accessed in March 2019).

[6] See an example for the first grade: https://wida.wisc.edu/sites/default/files/resource/CanDo-KeyUses-Gr-1.pdf (accessed in March 2019).

as an addition to content study, not as a purpose in itself. One reason for this may be the fact that a part of teachers in the informant group had no formal linguistic or CLIL training.

It is not therefore surprising that in this research context, another discovery was that also teachers' linguistic background, i.e. the totally lacking, partial, or entire fulfilment of CLIL teacher requirements, has an impact on how not only language but also assessment is viewed and feedback conveyed to stakeholders. There is no escaping that the municipal recruitment officials bear an enormous responsibility in making long-term decisions pertaining to the future effective implementation of CLIL. Regardless of the FNAE Ordinance (25/011/2005) on CLIL teachers' language proficiency, there are teachers in the field teaching content through a foreign language without any specific linguistic studies which clearly contributes to the quality of input, overall language teaching and assessment. Enthusiasm to teach through English is a convenient starting point for a CLIL teacher, but successful and professional CLIL pedagogy entails knowledge of second language learning/acquisition theories, didactics, teaching methodologies, and assessment literacy. Hasselgreen (2005) aptly underlines that even ordinary class teachers should minimally gain basic knowledge of assessment methodology and criteria when in position of assessing any language.

In this study, teachers reported using diverse assessment methods ranging from the most used teacher observation to the least common language portfolio. Whether or not teachers used those assessment methods to give pupils and parents feedback remained mostly unclear, indicating aborted or partial assessment function. Implicit assessment and feedback refer to 'gut feeling' judgements made without proper evidence or situations in which the teacher collects assessment information, evidence, and data, but leaves their interpretation to pupils and their guardians and assumes that they can infer the linguistic progress and state of art from this evidence. Such a practice must be rooted out in a similar manner as externalised collection of assessment information. Without collaboration, information exchange, or intense analysis the teacher will not be able to fully capitalise on the language samples someone else gathered in which case it becomes a feedback opportunity wasted. A linkage to learning should always be present. Such a practice resembles pseudo-assessment, even, since superficially and in appearance assessment activities take place, but, in fact, there is no continuation in form of feedback.

Teachers' assessment intentions differed from their self-reported practices especially when there was no school-specific assessment scheme available. In the data, a few schools were distinguished due to their assessment scheme, i.e. systematic feedback and report practices including learner self-assessment. The new Finnish assessment culture sketched in the NCC 2014 should be an asset in further stressing assessment *for* learning. The cornerstone of any approach to assessment promoting learning in CLIL is to make the learners aware of the dual learning objectives (content and language), their own learning processes, what is already learnt, and how they themselves can further promote and advance the attainment. Such an action necessitates communication and feedback. One means to this end could be the least used assessment method reported by teachers, the language portfolio.

The language portfolio, be it the European Language Portfolio concept or a modified version of it, has been found functional, informative, and invigorating, even, for young learners both in CLIL and EFL circumstances (Wewer 2015). According to the investigation, primary-aged learners and their parents alike viewed the portfolio very positively, as it was considered to be a tool to showcase even the most modest language proficiency in connection with content study. Another assessment method worth experimenting with is collaborative assessment, i.e. group tests which allow co-construction of knowledge and building on the group members' strengths rather than weaknesses both in language proficiency and content knowledge. Individual or group tests based on authentic materials provided in turn reveal how well learners apply the language and working skills already acquired.

Multimodality is essential to gather a variety of assessment information. A topical assessment method, for example, would be a modification of an escape room challenge; or flipped assessment in which tests are individualised for learners according to their own mark or grade goals, and testing entails, among other, phases, learners revisiting and correcting the tests for enhanced learning (Toivola 2020). Research has shown that learners benefit more of verbal feedback and assessments than marks or grades (e.g. Mäensivu 1999; Pulfrey, Buchs, & Butera 2011), which advocates for one-on-one conferences and feedback in which also non-linguistic, affective factors such as motivation, effort, and resiliency included. Even the most modest language skills deserve recognition in the beginning stages.

The pivotal question of adopting either integrated or separated content-language assessment cannot be unambiguously and all-inclusively answered in the Finnish context, since there is no 'one size fits all' approach to bilingual content teaching. CLIL is prone to vary from one municipal or school to another. Every teacher, or the given CLIL teacher community, must consider their specific circumstances and cater to the assessment needs as a response both to the objectives and circumstances defined in the local CLIL curriculum and learners' needs in learning. This is the essence of teacher-based assessment. Either an integrated content-language curriculum (Marsh 2013, p. 137) or a separate language curriculum (Dalton-Puffer 2007, p. 295) could be a solution to solve language assessment reference issues, depending on the local conditions.

Samples of language use and content knowledge can be, for the sake of convenience, collected simultaneously (one through the other), but it might be more sensible to make separate inferences based on the samples collected. In other words, one could use separate criteria for language and content, since it is possible that one can interfere the other by creating an illusion of better substance knowledge through eloquent delivery (Hönig 2010). For the same reason, in bilingual content instruction, particularly in cases of less eloquent delivery, it is important to resort to bilingual assessment, as the developing learner language may ruin the performance and portray weaker content mastery than it actually is. It is fair to employ integrated assessment only when the given content was taught and learned through the target language in which the skills are already considerable.

Another dilemma pinpointed in the study was the positioning of EFL and English in CLIL in respect to one another. A few teachers uttered their fear of CLIL lessons

turning into pure language lessons, which signals that the two forms of language instruction were seen as somewhat similar. This fear should be ungrounded, as language in CLIL is driven by disciplinary content and traditional EFL lessons with entirely different objectives are held separately, often by different teachers. The related assessment practices were also problematic, as some teachers left the language assessment in CLIL to the EFL teacher who taught the children two or three lessons a week. Letting the EFL teacher assess the CLIL language is not an applicable practice, since the social language focus and so also the assessment criteria applied should be different from CLIL. Moreover, the more academic proficiency demonstrated in CLIL lessons does not necessarily come across in EFL lessons. If all CLIL class teachers possessed formal double qualifications to teach EFL, linguistic support could be lent bilaterally, as the two could more readily complement one another. The view would ideally entail some alignment of language objectives in CLIL and EFL to reinforce language study and provide common benchmarks for assessment.

Conscious effort to upgrade CLIL education have recently been made in Finland. More precise boundaries for bilingual content instruction have been set by national steering, as the NCC 2014 now prescribes what should be included in the local CLIL curriculum. Moreover, teachers should nowadays be more aware of the texting of the new NCC than the previous one due to the highly comprehensive way it was drafted and introduced with supporting materials and local trainings. Funds for CLIL development have been allocated by the FNAE to local CLIL actors. It seems that Finnish CLIL is on the move toward more visible and tangible CLIL pedagogy with higher disciplinary orientation to language. The fourth decade of CLIL in Finland, the 2020s, will hopefully be the one in which assessment issues become the focal point in the CLIL development trajectory (see Marsh 2013). The final section in this chapter, a set of recommendations for assessment in primary CLIL, can be perceived as primary CLIL teacher's Code of Assessment. This set of recommendations is largely adopted from Wewer (2014a, p. 240), and the points represent aspects for development in the Finnish CLIL landscape eligible for adoption also elsewhere.

6.7 A Set of Assessment Recommendations for Primary CLIL Teachers

- Familiarise yourself with the National Core Curriculum and local CLIL curriculum. If there is no local curriculum, demand one to be drafted.
- Use the CLIL curriculum and CEFR or an equivalent framework as a reference.
- Assess both content and language using pre-defined criteria. This can happen either integrated or separately, interchangeably even, depending on the circumstances.
- Monitor both the learning process and language use.

- Remember all four basic language skills of listening, speaking, reading and writing as well as cultural knowledge and affective factors in various assessments.
- Acknowledge effort – even minor advancements can be celebrated.
- Let the learner know what s/he can do with language in CLIL contexts rather than what s/he cannot yet.
- Gathering of evidence is the starting point. Constructive feedback is an integral part of the assessment process, as the primary task of assessment is to enhance learning.
- Do not settle for observation only, and do not keep it to yourself only. Rely on various evidence as basis for your inferences about learners' language proficiency.
- Be consistent: assessment methods should be familiar to the learners: they should not significantly differ from the teaching methods used – nor should different language be used in assessment than in the classroom.
- Inform the parents about the principles of CLIL study and convey explicit assessment information also to them.
- Favour personal contact but also provide written information. A single mark or grade is very one-dimensional.
- Note that language in CLIL is disciplinary and more academic than in EFL study. The two are not directly comparable and cannot replace one another in assessment. Ideally, the two support and complement each other.
- Set the bar high for language use – go beyond word and single sentence level already at early stages. Build academic language needed in meaning making and conveying.
- Organise plenty of opportunities for pupils to use the target language, as it clearly is easier and more meaningful to assess active than passive language performance.
- Approach assessment from a positive perspective; there are multiple ways to praise a child.

Further Reading

Gottlieb, M., & Ernst-Slavit, G. (2014). *Academic language in diverse classrooms. Definitions and contexts*. Thousand Oaks: Corwin.

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge: Cambridge University Press.

Swinney, R., & Velasco, P. (2011). *Connecting content and academic language for english learners and struggling students. Grades 2–6*. Thousand Oaks: Corwin.

References

Barbero, T., & Järvinen, H.- M. (2009). Assessment in CLIL. In H.- M. Järvinen (Ed.), *Language in content instruction* (pp. 108–118). Turku: University of Turku and European Commission.

Basic Education Act. (628/1998). https://www.finlex.fi/en/laki/kaannokset/1998/en19980628.pdf. Accessed 16 Mar 2019.

Basic Education Decree. (852/1998). https://www.finlex.fi/fi/laki/ajantasa/1998/19980852. Accessed 16 Mar 2019.

Bentley, K. (2010). *The TKT teaching knowledge test course CLIL module*. Cambridge: Cambridge University Press.

Bovellan, E. (2014). *Teachers' beliefs about learning and language as reflected in their views of teaching materials for content and language integrated learning (CLIL)*. Jyväskylä: University of Jyväskylä.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653–675.

CEFR. (2001). *Common European framework of reference for languages: Learning, teaching, assessment* (3rd ed.). Cambridge: Cambridge University Press.

Czura, A., & Papaja, K. (2010). Classroom-based and external assessment in CLIL. *Anglica Wratislaviensia, XLVIII*, 163–170.

Dalton-Puffer, C. (2007). *The discourse of CLIL classrooms*. Amsterdam: John Benjamins.

Dochy, F. (2001). A new assessment era: Different needs, new challenges. *Research Dialogue in Learning and Instruction, 2*, 11–20.

Davison, C. (2008). *Assessment for learning: Building inquiry-oriented assessment communities*. Paper presented at Annual TESOL Convention, International TESOL, New York, March 2008.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415.

Eurydice. (2017). *Key data on teaching languages at school in Europe – 2017 Edition* (Eurydice Report). Luxembourg: Publications Office of the European Union.

FNAE Ordinance. (25/011/2005). *Opettajalta edellytettävä kielen hallinta muulla kuin koulun opetuskielellä annettavassa esi- ja perusopetuksessa* (Requisites of Teacher's Language Mastery in Preschool and Basic Education Given through a Language Other than the Language of Schooling). Finnish National Board of Education. https://www.finlex.fi/data/normit/26301/oph250112005su.pdf. Accessed 16 Mar 2019.

Gottlieb, M. (2012). *Common language assessment for English learners*. Bloomington: Solution Tree.

Hasselgreen, A. (2005). Assessing the language of young learners. *Language Testing, 22*(3), 337–354. https://doi.org/10.1191/0265532205lt312oa.

Hirsjärvi, S., & Hurme, H. (2010). *Tutkimushaastattelu: teemahaastattelun teoria ja käytäntö* [Research interview: Theory and practice of theme interview]. Helsinki: Gaudeamus.

Hönig, I. (2010). *Assessment in CLIL: Theoretical and Empirical Research*. Saarbrucken: VDM Verlag.

Hüttner, J., Dalton-Puffer, C., & Smit, U. (2013). The power of beliefs: lay theories and their influence in the implementation of CLIL programmes. *International Journal of Bilingual Education and Bilingualism, 16*(3), 267–284. https://doi.org/10.1080/13670050.2013.777385.

Ioannou-Georgiou, S., & Pavlou, P. (2003). *Assessing young learners*. Oxford: Oxford University Press.

Jäppinen, A.-K. (2003). *Ajattelu ja sisältöjen oppiminen vieraskielisessä opetuksessa.* [Thinking and content learning in CLIL]. (Research Report 2/3). Jyväskylä: Centre for Applied Language Studies.

Järvinen, H.-M. (1999). *Acquisition of English in content and language integrated learning at elementary level in the Finnish comprehensive school*. Turku: University of Turku.

Järvinen, H.-M. (2012). What does a CLIL teacher need to know about language – The linguistic competence of a CLIL teacher. In A. Koskensalo, J. Smeds, R. de Cillia, & Á. Huguet (Eds.), *Language: Competence – Change – Contact* (pp. 47–54). Berlin: LIT.

Kangasvieri, T., Miettinen, E., Palviainen, H., Saarinen, T., & Ala-Vähälä, T. (2012). *Selvitys kotimaisten kielten kielikylpyopetuksen ja vieraskielisen opetuksen tilanteesta Suomessa. Kuntatason tarkastelu* [Survey of immersion of national languages and instruction in a foreign language in Finland at municipal level]. Jyväskylä: University of Jyväskylä, Centre for Applied Language Studies.

Leal, J. P. (2016). Assessment in CLIL: Test development at content and language for teaching natural science in English as a foreign language. *Latin American Journal of Content and Language Integrated Learning LACLIL, 9*(2), 293–317.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol/Buffalo: Multilingual Matters.

Mäensivu, K. (1999). *Opettaja määrittelijänä – oppilas määriteltävänä: sanallisen oppilaan arvioinnin sisällön analyysi* [Teacher as the definer – Pupil being defined: Content analysis of verbal pupil assessment]. Jyväskylä: University of Jyväskylä.

Marsh, D. (2013). *The CLIL trajectory: educational innovation for the 21st century igeneration*. Córdoba: Universidad de Córdoba.

Massler, U. (2011). *Assessment in CLIL learning. Guidelines for CLIL implementation in primary and pre-primary education*. A publication of the PROCLIL project. http://arbeitsplattform.bildung.hessen.de/fach/bilingual/Magazin/mat_aufsaetze/clilimplementation.pdf. Accessed 16 Mar 2019.

Massler, U., Stotz, D., & Queisser, C. (2014). Assessment instruments for primary CLIL: the conceptualisation and evaluation of test tasks. *The Language Learning Journal, 42*(2), 137–150. https://doi.org/10.1080/09571736.2014.891371. Accessed 19 Mar 2019.

McKay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge University Press.

Merisuo-Storm, T. (2007). Pupils' attitudes toward foreign-language learning and the development of literacy skills in bilingual education. *Teaching and Teacher Education, 23*, 226–235.

NCC 2004. (2004). *National core curriculum for basic education*. English translation. Finnish National Board of Education. https://www.oph.fi/english/curricula_and_qualifications/basic_education/curricula_2004. Accessed 16 Mar 2019.

NCC 2014. (2016). *National core curriculum for basic education*. English translation. Finnish National Board of Education Publications 2016:5. Helsinki: Next Print.

Nikula, T. (2005). English as an object and tool of study in classrooms: Interactional effects and pragmatic implications. *Linguistics and Education, 16*, 27–58. https://doi.org/10.1016/j.linged.2005.10.001.

Nikula, T., & Järvinen, H.-M. (2013). Vieraskielinen opetus Suomessa [Instruction in a foreign language in Finland]. In L. Tainio & H. Harju-Luukkainen (Eds.), *Kaksikielinen koulu – tulevaisuuden monikielinen Suomi [Bilingual school – Multilingual Finland of the future]* (pp. 143–166). Jyväskylä: Jyväskylän yliopistopaino.

Papaja, K. (2014). *Focus on CLIL: A qualitative evaluation of content and language integrated learning (CLIL) in Polish secondary education*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Peltoniemi, A., Skinnari, K., Mård-Miettinen, K., & Sjöberg, S. (2018). *Monella kielellä Suomen kunnissa 2017. Selvitys muun laajamittaisen ja suppeamman kaksikielisen varhaiskasvatuksen, esiopetuksen ja perusopetuksen tilanteesta* [Many languages in Finnish municipalities 2017. A report of other large-scale and small-scale bilingual ECEC, preschool and basic education]. Jyväskylä: University of Jyväskylä. https://jyx.jyu.fi/bitstream/handle/123456789/57577/978-951-39-7391-9.pdf?sequence=5&isAllowed=y. Accessed 16 Mar 2019.

Pihko, M.-K. (2010). *Vieras kieli kouluopiskelun välineenä: Oppilaiden kokemuksista vihjeitä CLIL-opetuksen kehittämiseen* [Foreign language as a medium of school study: implications of pupils' experiences for development of CLIL instruction]. Jyväskylä: University of Jyväskylä.

Poisel, E. (2007). Assessment modes in CLIL to enhance language proficiency and interpersonal skills. *VIEWS, 16*(3), 43–46.

Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology, 103*(3), 683–700. https://psycnet.apa.org/doi/10.1037/a0023911.

Rahman, H. (2012). *Finnish pupils' communicative language use of English in interviews in basic education grades 1–6* (Research report 340). Helsinki: University of Helsinki.

Rasinen, T. (2006). *Näkökulmia vieraskieliseen opetukseen. Koulun kehittämishankkeesta koulun toimintakulttuuriksi* [Perspectives on bilingual instruction. From a development project to operational culture]. Jyväskylä Studies in Education, Psychology and Social Research 281. Jyväskylä: University of Jyväskylä.

Seikkula-Leino, J. (2007). CLIL learning: Achievement levels and affective factors. *Language and Education, 21*(4), 328–341. https://doi.org/10.2167/le635.0.

Serragiotto, G. (2007). Assessment and evaluation in CLIL. In D. Marsh & D. Wolff (Eds.), *Diverse contexts – Converging goals. CLIL in Europe* (pp. 271–283). Frankfurt am Main: Peter Lang.

Short, D. J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly, 27*(4), 627–656. https://doi.org/10.2307/3587399.

Toivola, M. (2020). *Flipped Assessment: A Leap towards Assessment for Learning*. Helsinki: Edita.

Wewer, T. (2014a). *Assessment of young learners' English proficiency in bilingual content instruction CLIL.* Doctoral dissertation, University of Turku. http://urn.fi/URN:ISBN:978-951-29-5754-5. Accessed 12 May 2018.

Wewer, T. (2014b). One minute to Michigan: Computer simulations as an assessment method for profiling English language proficiency in bilingual CLIL instruction at primary level. In H. Juuso, A. Lindh, M. Hasari, K. Kumpulainen, A.-P. Lapinoja, P. Pirilä, S. Raappana, & O. Tiainen (Eds.), *Research-based school and teacher education* (Publication of Finnish teacher training schools 2014) (pp. 93–102). Oulu: University of Oulu, Oulu University Teacher Training School.

Wewer, T. (2015). *Portfolio as an indicator of young learners' English proficiency in mainstream English instruction (EFL) and bilingual content instruction (CLIL)*. Turku: University of Turku. http://www.utupub.fi/bitstream/handle/10024/105166/Gradu2015Wewer.pdf?sequence=2&isAllowed=y. Accessed 16 Mar 2019.

WIDA (2014). *The WIDA 2012 English language development standards and resource guide, international edition, kindergarten through grade 12.* https://wida.wisc.edu/sites/default/files/resource/2012-ELD-Standards-International-Edition.pdf. Accessed 16 Mar 2019.

Zafiri, M., & Zouganeli, K. (2017). Toward an understanding of content and language integrated learning assessment (CLILA) in primary school classes: A case study. *Research Papers in Language Teaching and Learning, 8*(1), 88–109.

Chapter 7
Gearing Teaching and Assessment Towards CLIL: Theorizing Assessment *for* Learning in the Junior High School Soft CLIL Classrooms in Japan

Hidetoshi Saito

7.1 Introduction

This chapter describes an emerging trend in implementing principles of the Content and Language Integrated Learning (CLIL) instruction in junior high (or lower secondary in Europe) school English as a Foreign Language (EFL) lessons in Japan. Unlike European contexts, CLIL in Japan, apart from a handful of private schools, is not conducted among primary and secondary schools. Although CLIL in Japan is still in its infancy (Ikeda et al. 2013), there is growing understanding of the value of focusing on content (see Chap. 1, this volume). The teachers who participated in this study have attempted to integrate content in the EFL context, where the acquisition of traditional linguistic and skills categories dominate in the curriculum and content knowledge is trivialised. The teachers of this study have oriented themselves towards emphasis in content, as the study reveals, by engaging in various assessments *for* learning (A*f*L).

The lessons reported here have adopted approaches commonly labelled communicative language teaching (CLT). These courses follow language-centred curricula without clear disciplinary focus which represents the third Quadrant of Leung and Morton's (2016) integration matrix (Chap. 1, this volume). This can be called 'soft CLIL', aimed at attaining linguistic objectives by means of content in the curriculum, yet content is equally or more focused at times in the lessons.

By describing the emerging trends in adopting CLIL approaches in a CLT context, the current study is intended to delve into the ways in which the two teachers teach and assess the students by balancing content with language through various A*f*L or formative assessment techniques. The courses are traditional EFL classes in which content learning is not part of the final grade, although the teachers have

H. Saito (✉)
College of Education, Ibaraki University, Mito, Ibaraki, Japan
e-mail: hidetoshi.saito.cldwtr@vc.ibaraki.ac.jp

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_7

incorporated and valued content in their lessons. The assessment episodes of the soft CLIL lessons examined in this chapter illustrates seamless continuity between CLIL and EFL lessons. The common A*f*L practices along with teaching techniques observed in the lessons partially fills the existing gaps in the literature on the interface between CLIL and other language teaching approaches (see Ortega 2015). This study thus contributes not only to CLIL pedagogy, but also to EFL pedagogies in general.

7.2 Literature Review

Previous research has offered descriptive frameworks of classroom-based assessment that promotes learning in L2 and other subject areas. Some of these frameworks depict complex classroom-based assessment processes or beyond under different banners, such as classroom-based assessment (Hill 2012; Hill and McNamara 2012), learning-oriented assessment (Turner and Purpura 2016), and teacher-based assessment (Davison and Leung 2009). Similarly, Rea-Dickins (2001), Torrance and Pryor (1998), and Wiliam (2010, 2011) offer simpler descriptive models that focuses more on A*f*L. The goals of all these models are to identify and classify classroom-based assessment practices and processes in light of potential factors that influence teachers' assessment-related decisions. Introducing each model is beyond the scope of this chapter. Among the nine key characteristics identified across these models, four factors were chosen for the present investigation: (1) whether assessment is designed initially such that the information is used for grade reporting (functions), (2) whether it is carried out in a planned or spontaneous fashion (planning), (3) who gathers the information (assessors), and (4) what learning aspects are targeted (targets).[1]

In planning assessment promoting learning, the teacher needs to decide, first, whether the information collected will also be used for grade reporting. Functions other than for reporting may include teaching, learning, classroom management, and socialisation (Hill 2012), which comprise formative functions of assessments. Grades are a major consequence of classroom learning and assessment, and have a significant emotional and motivational effect on students. Grades can inform how learning is going (Brookhart 2004; Wiliam and Leahy 2015; see Chap. 1, this volume). They also motivate, and hence can also debilitate, students' continued learning (Crooks 1988). This element here accommodates using assessment activities meant to serve initially summative function for the purpose of promoting learning. This requires thinking how to use the information activities, such as term tests for example, formatively. While various short-cycle assessments, such as worksheets

[1] Note that these key A*f*L characteristics parallel the language assessment literacy competencies that the teacher needs to acquire as a professional, which consist of background/rationale (or purpose), describing the proficiency (or target), and implementing the assessment (or method) (Inbar-Lourie 2008).

and homework, are intrinsically formative by design, one-shot, long-cycle assessments, such as mid-term and final examinations, are considered intrinsically summative. Despite these apparent design features being inherent in assessment, whether it is formative or summative hinges on its purposes (Brookhart 2010). Assessment activities initially designed to have a summative function can be used with a formative function in mind, such as planning the next step in the lesson (ceasing thus to be summative; see Chap. 1, this volume). Likewise, assessments planned initially to have a formative function can acquire a summative function (see Mehisto and Ting (2017) for further examples). However, since simplicity is preferred for practicality, the present A*f*L framework shown below has chosen the dichotomy of grade reporting versus not reporting as initially planned in the lesson. Multiple categories of various other functions were also excluded.

Second, assessment can be incidental or planned (Ellis 2003; Hill 2012). Much of the teacher's action is derived from moment-by-moment decisions contingent upon various on-task student needs that arise in the classroom (Torrance and Pryor 1998). While this may be the case, A*f*L can also exploit prior planning, especially when collecting information that provides feedback for learning targets. For example, using a set of planned criteria, the teacher can offer detailed feedback to the students on what is needed to satisfy the target level. Unreported assessments can thus be either planned or incidental. Incidental assessment for grade reporting is unlikely, and thus excluded here.

Third, all agents in the classroom, teachers, students themselves, and peers, are involved in the A*f*L processes (Wiliam 2011; see also Chap. 1, this volume). They are the gatherers of assessment information—the assessors. They co-construct the learning process and outcome by cooperatively implementing A*f*L in the lessons.

The fourth and final factor in A*f*L is learning targets and standards. The teachers decide what ability, knowledge, and skills are the focus of the lesson. They are the goals for the learners to achieve and for the teachers to observe progress towards. Standards are used as a referential framework that the teacher uses formally (norm- or criterion-referenced) or informally (personal theory/beliefs about language and learning) to judge what is observed. They can also be described as the focus (Davison and Leung 2009), evidence interpretation (Mavrommatis 1997), and 'what theory or standard to use and what to look for' (Hill 2012). In the present study, learning targets are dichotomised as whether they are language-focused or content-focused, not to mean that the acquisition of both is not the goal, as integration of the two, of course, is underscored in CLIL. The learning targets can be more specific linguistic or content goals, depending on the focus in the lesson. The resulting matrix of 18 possible combinations of the four variables appears in Table 7.1. The numbers in the grids or types represent the 18 possible combinations, and I will use these numbers throughout this paper for reference. Each type will be explained further in the Results Section.

The present study employs the theoretically generated A*f*L framework to understand how assessment and instruction in soft CLIL lessons are implemented. That is, one of the purposes of this qualitative inquiry is to identify assessment types represented in the framework and to discuss reasons for their use. In so doing, the

Table 7.1 The four-factor CLIL A*f*L framework

Grade Report	Planning	Assessor	Activity Focus	
			Language	Content
Reported	Planned	Teacher	1	2
		Peer	3	4
		Self	5	6
Unreported	Planned	Teacher	7	8
		Peer	9	10
		Self	11	12
	Incidental	Teacher	13	14
		Peer	15	16
		Self	17	18

Note. The number is given for referential use

present study demonstrates that the two focal lessons (and other lessons not detailed in this chapter) have displayed the CLIL-orientation and that students' content and language learning has been variously assessed through the A*f*L framework. A corollary of this investigation is the evaluation of the proposed framework of A*f*L, which contains a set of simple, seemingly mutually exclusive categories. One caveat here, however, is that this model focuses more on the first 2 stages (planning and assessment methods) of Davison and Leung's model (see Chap. 1, this volume) rather than on the latter 2 stages (interpretation and feedback), although in some cases, they are touched on in discussion.

The research question posed was:

1. How do the two EFL teachers in their respective soft CLIL lessons use A*f*L?

7.3 Methods

7.3.1 Design

This study used a case study research design with qualitative data to describe the two teachers' assessment practices in their lessons. The present study utilised Yin's (2009) four components of case study research design: purposes, unit of analysis, logic that links the data to the purposes, and criteria for interpreting the findings.[2] First, the purpose of this study was to understand the characteristics of two experienced teachers that distinguished them from others in teaching and assessment, particularly, by balancing content and language. Their professional excellence was guaranteed by the fact that local educational boards commissioned the two teachers on more than one occasion to join professional task forces and requested that they

[2] Because the study is exploratory, the purposes, instead of propositions, are stated here (Yin 2009).

deliver teacher workshops. Given the likelihood that they were highly competent teachers, examining their classroom practice was therefore valuable. Next, the unit of analysis in this study consisted of, following Mavrommatis (1997), each 'assessment episode' in the lessons, like a slice of the planned or incidental moment that provides information for those who are involved in the interpretation of student learning.

The third case research design component, that is, the logic that links the data to the purposes, involves a choice of analytic techniques. The present study used a technique called 'pattern matching' (Yin 2009), whereby empirical patterns are matched with predicted ones. An assumption was made that the teachers demonstrate distinguished characteristics in teaching and assessment. One prediction is that they use various assessment techniques identified in the A*f*L framework because their approach brings with it awareness for the necessity of diverse assessment tasks eliciting content or language or both. The patterns predicted from the framework are matched with actual assessment episodes.

The final component, the criteria for interpreting the findings, concerns rival explanations, which provide alternative accounts of a given phenomenon. Because this study was exploratory, stating a rival hypothesis deemed, if not impossible, difficult. However, one might suppose that even though the lessons observed contain various A*f*L activities integrating content and language, they might lack assessment episodes that require student deep processing of *content*. In other words, students may engage in a content-focused activity without employing higher order thinking skills.

7.3.2 Participants

Two experienced Japanese junior high school teachers of English in their late forties participated in the study. The male teacher, Sota, had taught English at his school for 8 years at the time of the study (all names are pseudonyms). The school is located in a mid-sized city. Most students were strongly academically oriented, and their parents had high expectations of the school's acclaimed curriculum. Sota was teaching ninth graders who had to take a competitive exam to qualify for an entry into reputable high schools. Sota admitted, as do many ninth grade teachers, to feeling stakeholders' pressure about the entrance rates. The 35 students in his class, however, appeared very friendly and cooperative as well as talkative.

The female teacher, Aki, had taught English at her school for 7 years at the time of the study. The school is located in a small city. The students seemed to vary in their academic motivation than those from Sota's. In general, however, the 33 students in the class appeared cooperative, willing to learn, and interested in English.

In the lessons that were the focus of this study, or *focal lessons*, Aki was team-teaching with Donny, a then-new UK-born assistant language teacher. Donny asked questions and organised the lessons a fair amount of the time, although Aki controlled the lesson plans, including worksheets, activities, questions to be asked, and

even who to speak when and who to be called on sometimes. The data in the Results Section thus include Donny's utterances. Concerning their background knowledge about CLIL, Sota knew what CLIL was and was aware of the fact that he was utilising a soft CLIL approach while Aki had no background knowledge of CLIL. Years of experience seemed to lead her to employing her soft CLIL approach as a natural next step from common CLT.

To assess the speaking proficiency of the students, 10 students from each class were randomly chosen to take a telephone-based standardised speaking test (TSST). Their average scores, on a scale of 1–9, were 3.00 (*SD* = .67) for Aki's class and 3.40 (*SD* = 1.26) for Sota's class (see ALC Press (2018) for level descriptors). Their levels of speaking proficiency were thus regarded as beginning to pre-intermediate. This is in line with the national survey results (MEXT 2017b), and most students in both classes were believed to be at CEFR A1 or A2 levels across the four skills.

7.3.3 Procedure

The researcher visited both schools on approximately a monthly basis and observed both teachers' lessons for 1 year. A total of 16 lessons (8 administered in each class) were observed and video-recorded, and each was immediately followed by a one-on-one interview. The researcher chose one soft CLIL focal lesson from each teacher where content played an equal or greater role than language and implemented stimulated recalls of the lessons, where the teachers individually watched a video clip of their focal lesson. The researcher paused the recording at points when asking questions concerning the teachers' intention, self-evaluation of the activities, and students' reactions. The teachers were also free to pause and repeat the video anytime they wished. The chapter draws mainly on these lessons for analysis. In addition, other lessons are considered complementary.

In post-lesson interviews and stimulated recalls, questions revolved around diverse aspects of the lesson, including previous lessons, planning, and implementation of the lesson, teaching/learning intentions, feedback, and plans for the next lessons. Information was also gleaned from the interviews regarding both the teachers' feelings about the success of the lesson and assessment as well as the gap between their ideal conception of the lesson and how the actual lesson went. In addition to these main text data, a range of data was collected from observation notes, worksheets, student interviews, and email. Sota also conducted audio-recorded group discussions in the focal lesson that were made available for the analysis while those of Aki's came from other lessons.

7.3.4 Analysis

The qualitative analysis drew on both the stimulated recall data of the focal lesson and all other collected data. The present study employed concept-driven (Gibbs 2007) or deductive coding based on theoretically developed codes presented in Table 7.1. The 18 types listed in Table 7.1 thus guided the coding of the main data. The principal researcher initially completed all coding of transcripts of stimulated recalls, post-interviews, and teacher-student exchanges of the focal lesson. Other interviews, lesson videos, and student discussions were all transcribed but selectively coded and interpreted. A research assistant confirmed whether the coding of each piece of data fit the 18 types. No disagreement between the two ensued. A few ambiguous interpretations of the data were clarified by the two teachers in the follow-up interviews and email correspondence.

7.3.5 Lesson Goals in Focal Lessons

Lesson goals in these focal lessons are one of the reasons why they are considered to be soft CLIL. In both lessons, the teachers clearly set out their content goals in the beginning, although neither of them declared language goals, which is not typical for junior high EFL lessons. Both teachers wrote their goals on the blackboard. Aki's goal for the lessons was to explore the questions of 'What is friendship? Justice? Promise? Kindness?' The focal lesson was devoted to characterising the protagonists' behaviour in light of these four terms. The lesson used a graded version of O. Henry's classic *After 20 Years* in which two old friends meet for the first time in 20 years without knowing the fact that one is a police officer and the other is a wanted man. Thus, the lesson involved moral dilemma questions that prompted the students to analyse the minds of the protagonists critically, to compare them with a set of criteria, and subsequently to engage in group discussions. This, therefore, introduced elements of a literature lesson to the EFL lessons.

Sota's lesson was based on a unit about fair trade, and the goal he wrote on the board was: 'What can we do to help the poor children on a cacao farm?' He set up repeated role-play discussions for the students to deepen their understanding of the situation in Ghana, which was briefly described in the text. In the two previous lessons, he showed two short video clips about child labour in Ghana dubbed in L1 (mother tongue) and one clip about Kenya's case in L2. To meet the lesson goals, the students needed at least to understand the main text and integrate it with video information in order to form their own opinions. This brought social studies content in an EFL lesson.

Although both teachers admitted in the post interviews that they had language goals, they did not explicitly share those with the students at the beginning of the lesson. Because *After 20 Years* was a supplementary unit in the textbook, it did not contain any explicit target grammars. Aki's hidden linguistic goal in the lesson,

however, was to improve oral fluency and the functional (genre) goal of stating opinions and reasons. In Sota's lesson, although undeclared at the outset, specific linguistic forms were still in focus, particularly the present perfect of *have* constructions. Sota repeatedly encouraged the use of the target structures and backchannels during the lesson. Thus, in Sota's focal lesson, both language and content were the foci.

7.4 Results

The following section examines how the two teachers incorporated A*f*L into their lessons. By linking the assessment episodes with the 18 types in the A*f*L framework (Table 7.1), the results show whether each episode fits into each category.

7.4.1 Reported, Planned Assessment (Types 1–6)

Types 1 and 2 are teacher-planned assessments for grade reporting that include summative performance assessments as well as other supporting documents, such as homework, worksheets, quizzes, and essays, some of which may also belong to Types 7 and 8 (ungraded but planned). Thus, Types 1 and 2 are included to identify A*f*L by design, yet also have a summative function. Some evidence (Kokatsu 2018) suggests that Types 3 and 5, planned peer or self-assessments, seem to be less commonly practised and used for grading in Japan. In addition, it needs to be mentioned that Types 2, 4, and 6 are targeted at content learning, hence they were not explicitly included in grade reporting in the present soft CLIL situation.

During my visit, there was no opportunity to observe any planned formal test-like event. I asked if any part of the observed lessons went into the final grade, and teachers mentioned at least the following:

Worksheets: The teachers used various types of worksheets almost every lesson, some of which were content-focus, but not all. They were usually collected, checked, and returned to the students throughout the course. See below for examples.

Recitation and Skit: In Sota's class, a few students in lessons recited Malala Yousafzai's UN speech or a short story called *Mother's Lullaby*. Similarly, students in Aki's class practised their own skit of celebrity interviews and performed it in front of the class. In both cases, only the language aspect was marked for grading.

All of these can be considered to serve both summative and formative functions because they were factored into the final grade but also were regular follow-up feedback whose goal was to improve learning. For example, in one of Aki's lessons, students practised a self-made role-play skit, whose language goal included the use of interrogatives. One of the groups was asked to perform theirs as a model. Immediately after this demonstration, Aki engaged all students in a group reflection on the sociolinguistic appropriateness of the interview skit by suggesting "Leaders

(of each triad group), ask group members whether the interview questions sounded appropriate as the ones for the school principal, and give feedback to your members" (video transcript, translated by the author). This instruction itself functioned as a feedback to each group to reconsider interviewee's status and suitable linguistic politeness levels in their own skit, which might end up being revised.

Both teachers reported that they based final grade reports mainly on individual interview tests, homework, speeches, quizzes, performance tests, and mid-term and final paper tests. Language contributed to the final grade the most. Aki admitted in the interviews that she did not use Types 3–6 for final grade reports, while Sota said students chose the best speaker from among their own group members for discussion activities, and those who were chosen by members earned extra marks. Except for this, the student-based A*f*L rarely contributed to grade reports although both teachers frequently used peer and self-assessments in their lessons (see below). The primary reason for the absence of Types 3–6, as both admitted, was because grading is the responsibility of the teachers.

7.4.2 Unreported, Planned Teacher Assessment of Language and Content (Types 7 & 8)

For both language and content aspects, observation is probably the most commonly used teacher-planned A*f*L activity. The teachers milled around and monitored students' work in terms of both content and language during pair or group activities. Feedback from observation was given immediately or shared later with the entire class. Another example of these types seen in the data was teachers' planned questions or questioning. Sota's linguistic goal included the appropriate use of the present perfect *have* construction. He asked a series of metalinguistic questions after the first group discussion on fair trade.[3]

By asking these questions in Extract 7.1, Sota was able to confirm the number of students who successfully achieved linguistic goals in the discussion. He explained in the interview that the fluency goal of sustaining a conversation for one minute (line 1) was derived from a goal depicted in the regional ninth grade can-do statement list. Sota's checks on the use of the target form occurred once or twice in one lesson and across lessons. His oral checks frequently come with encouragement for the students to try out the target phrases/grammar in conversation, suggesting additional information of when to use them and sharing peers' good use he has overheard. The primary function of this simple oral check thus seemed, first, to reconfirm the students of the teacher's linguistic agenda, which was sometimes not clearly stated in the beginning of the lesson, and second, to prompt students to use it and

[3] Extracts and Table 7.4 use the following notations: (()) = non-verbal responses; () = omissions for clarification; [] = the English translation of Japanese utterances.

Extract 7.1 Sota's Focal Lesson video

1	Sota: 一分間持ちましたか? [Did you sustain conversation for one min.?] Did you use *genzai*
2	*kanryo* [present perfect]? '*Have you ever*' をきけた人? [Who could use '*Have you ever*'?]
3	Students: ((raised hands))
4	Sota: '*How many times*' をきけた人? [Who could use '*How many times?"*]
5	Students: ((raised hands))
6	Sota: 三つの用法全て言えた人? [Who could use all three types of the present perfect?]
7	Students: ((raised hands))

Note. Numbers in the left column are used for reference purposes

Extract 7.2 Aki's focal lesson video

1	Aki: I want you to think why. Please raise hands. 丸の人 [Those who chose it?].
2	Bob has justice?
3	Students: ((raised hands))
4	Donny: Promise. Bob has promise?
5	Students: ((raised hands))
6	Donny: Bob has kindness?
7	Students: ((raised hands))
8	Donny: How about Jimmy? Jimmy has friendship?
9	Students: ((raised hands))

Table 7.2 Results of teacher survey on students' evaluation of the main characters' behaviours

	Friendship	Justice	Promise	Kindness
Bob	14	0	26	2
Jimmy	22	27	24	22

Note. The numbers were actual student votes

reflect on their use. Hence, this facilitated students' attempt to use them on another occasion.

For content-focus A*f*L, Aki and Donny surveyed the main lesson goal question first by asking the students to choose and check off suitable labels on the worksheet, from among friendship, justice, promise, and kindness, to best characterise the behaviours of Jimmy and Bob. Then, the teachers asked the same questions orally as in Extract 7.2.

The teachers counted the number of hands and wrote the numbers on the board. The final tallies appear in Table 7.2. This required the students not only to understand the four descriptive terms, but also to infer the meanings of the protagonists' behaviours in the story, and to determine whether they matched (see Sect. 7.5 for a follow-up).

7.4.3 *Unreported, Planned Peer Assessment of Language and Content (Types 9 & 10)*

Prior research has found that on average, more than a half of junior and senior high English teachers used planned peer/self-assessment (Types 9–12) (Saito and Inoi 2017), probably mostly for language targets. To assess language, Sota asked students to complete peer assessments. While a peer assessment sheet (Table 7.5) was used for the focal lesson, he switched to a simple tick slip (Table 7.3) in later lessons. The students in these lessons were put into groups of four in which three students talked about the topic and one observed the members as an assessor. They took turns playing the assessor's role during the repeated discussions. Using the tick slip, the assessor tallied the number of discourse features observed by simply drawing circles in the right column and returned the slip to each speaker. Sota believed that this method prompted students to become aware of their own and others' linguistic limitations and needs to hone their skills.

To assess content learning, Aki used tabletop whiteboards in the focal lesson. Group members shared ideas and decided which ideas went on the board to be evaluated by peers. The board was about the size of a laptop computer; hence, information written down needed to be selected carefully. Here, they discussed the second question of 'What would you do if you were Jimmy (police officer)?' The ethical dilemma depicted in the story sparked heated group discussions, and they wrote their groups' ideas on the whiteboard. Each group presented its ideas to the entire class on the tabletop whiteboards stuck to the big blackboard in front, as seen in Table 7.4.

This question urged them to think like Jimmy, which presumably ignited their imagination and hence surpassed their current language levels. Much of the discussion was thus conducted in a mixture of Japanese and English, and the comments reflected this translanguaging process. Aki praised all ideas even though the comments on the board contained L1. Aki agreed in the interview that this group discussion and whole class sharing of diverse ideas fulfilled the lesson goal of the day.

Table 7.3 A sample tick slip peer assessment

Peer's Name	Counts
Short utterance	○○
Long utterance (5 sentences)	○
Question	○○
Response	○○○○○
Backchanneling	○○○
Support	

Note. The original version was in Japanese

Table 7.4 Groups' responses on tabletop whiteboards from Aki's focal lesson

Group 3: (I'd arrest him) Because I'm a policeman.
Group 4: 何もしない。BobのlifeはBobのfree [I wouldn't do anything, because Bob is free to do anything in his life].
Group 7: 指名手配されるぐらいの悪人をつかまえて昇進する [I'd get promoted by arresting the wanted man].
Group 9: (I'd) Arrest Bob by myself. For New York. For Bob. 自分の良心のためにも [for my conscience]. 自分の仕事が首にならないために [in order for me not to get fired].
Group 10: If I'm Bob, I say 'you have to go to police yourself.'
Group 11: 友達を逮捕できない [I can't arrest my friend]. Bobと話したい [I want to talk to Bob].

7.4.4 Unreported, Planned Self-Assessment of Language and Content (Types 11 & 12)

Self-assessment is commonly done using worksheets and self-reflection sheets. Both teachers asked students in the focal as well as other lessons to complete the sheets to reflect on their own learning. These sheets were collected and usually returned with teachers' comments on. Thus, they can also be regarded as unreported, planned teacher's assessments. Both Sota and Aki used simple forms that students could complete in a few minutes. After group discussions, students in Sota's class used Table 7.5 in the focal lesson.

This sheet was part of a larger form that included rubrics. As can be seen, this also functioned as a peer assessment. The content descriptor attached to this form appears in Table 7.6. Although this simple descriptor could identify the presence of opinions and support, actual ideas presented were left unreported in this A*f*L form.

7.4.5 Unreported, Incidental Teacher Assessment of Language and Content (Types 13 & 14)

Incidental A*f*L activities (Types 13 and 14) are probably one of the most common types of assessments. Teachers observe students' work during lessons and ask questions (i.e. questioning) to make decisions about the next steps. They are mentally or physically noted for immediate or later use. Various moments of unreported incidental teacher assessment have been found in the data. In Extract 7.3, where the language goal of the task was to 'give opinions,' Nana, a student in need of support, tried to explain to two other members how they could help the poor children. Nana sought help in two places, lines 2 and 5 in Extract 7.3. In line 2, she wanted to know whether the word *product* was OK to use, and a group member confirmed it in line 3. In line 5, she had another problem, so she used a Japanese word for *transport*. The teacher, Sota, who just happened to stop by, offered a linguistic scaffold (or support to reach out a higher-level of performance) in line 6 so that Nana could complete the

Table 7.5 Self and peer assessment form used in Sota's class

Date		Partners		
Topic				
	Contribution	Content	Language	Eye contact/ voice
Myself	3 2 1	3 2 1	3 2 1	3 2 1
Member ( )	3 2 1	3 2 1	3 2 1	3 2 1
Member ( )	3 2 1	3 2 1	3 2 1	3 2 1

Table 7.6 Descriptors of content from Sota's reflection sheet

Levels	3	2	1
内容 [Content]	理由や例をあげて説得力のある意見を述べる [Gives convincing opinions with reasons and examples]	短い説明や意見交換をする [Gives short explanations and exchanges information]	最小限の情報提供をする [Gives minimum information]

Note. Translated by the author

Extract 7.3 Sota's group recording (10_06)

1	Nana: Cacao, cacao 何だっけ [what is it?] made into chocolate.
2	Fair trade 商品ってなんていうの?[how do you say *product*?] product?
3	Chika: うん [Yes].
4	Asa: *Goods*でもいいんじゃね? [*Goods* are OK too?]
5	Nana: (If we) Buy, buy, we buy this, money 輸送 [transport/transfer].
6	Sota: Money goes to…
7	Nana: Goes to Ghana, Ghana, Ghana children.
8	Chika: おお! [Wow!].
9	Sota: 完璧 [perfect].

sentence in line 7. This ended with Chika and Sota's praise. However, it is not clear whether their praise was targeted at successful performance (language) or at her idea suggesting how to help the poor children (content). In this case, it could be both.

A more obvious example of unreported, incidental teacher assessment of content is Aki's seeking for details and confirmation of the student responses to the survey.

Extract 7.4 describes a follow up exchange to Table 7.2 above, in which the teachers surveyed how the students evaluated the two protagonists' behaviours. After the votes were counted, Aki called on the two students who voted for 'Bob's kindness' to justify their choice. She recalled in the interview that she just wanted to know why the students regarded the wanted Bob (the thief) as being kind. In fact, she employed 'questioning' to explore the quality of students' opinions, building on the previous assessment of 'survey,' which was limited to capturing their opinions quantitatively. After line 9, she candidly disagreed with majority students' votes (22 in total) for Jimmy's friendship in order to stimulate students' critical thinking. The students and teachers thus exchanged their opinions to explore the reasoning

Extract 7.4 Aki's focal lesson video

1	Aki: Bob is wanted, Bob's wanted, so no justice. Taku, why do you think Bob is
2	kind? In Japanese, OK. どういう部分が優しい? [What part of his
3	behaviour is kind?]
4	Taku: 何をやったかにもよりますけど [It depends on what he did]. ちゃんと
5	Jimmyが遅れても待っていた [He has been waiting for Jimmy even though he is late].
6	Aki: ああ、wantedなのにね。[I see, even though he's wanted].
7	もう一人 [another person], Another opinion?
8	Kay: 同じです [the same].
9	Aki: Same opinion? OK. Jimmy has kindness…Everyone, I think, I
10	don't think Jimmy has friendship because, if I were Jimmy, I'd arrest Bob myself.

for their choice. Although the survey here was a planned A*f*L, this type of further probe into mutual understanding ensued incidentally. This suggests that incidental A*f*L can be embedded in planned A*f*L. More importantly, Aki's move from quantitative assessment to qualitative assessment illustrates a potential strategic A*f*L process CLIL teachers can take advantage of.

7.4.6 Unreported, Incidental Peer Assessment of Language and Content (Types 15 & 16)

The last four types, 15–18, are difficult to capture. Unreported incidental peer feedback would be found in pair/group interactions that take place during or after the completion of a task, whereas unreported incidental self-reflection would remain in their mind or perhaps could be heard in private speech. Both types are ephemeral in nature; yet they are likely to promote learning as they are essentially immediate feedback and reflection.

Working in groups allows students to exchange feedback regarding the language, which is an important and widely supported aspect of second language acquisition (SLA) (e.g. Philip et al. 2014). An example in the data is shown in Extract 7.5, in which the language goal of the task was to 'explain what fair trade is.'

In line 4 of Extract 7.5, Emi could not recall *sweet* and used the Japanese equivalent. Others in the group quickly provided scaffolding in line 6.

The group recordings of the lessons contained only a few occasions of unreported, incidental peer A*f*L types with the focus on content. Before discussing the main question about how to support poor children, Sota asked students in groups to define fair trade.

In Extract 7.6, Atsu's question in line 4 was a factual check. He probably knew what fair trade was, but he seemed to challenge Toki to explain it in English. This question was most likely a display question—commonly found in CLIL lessons (Dalton-Puffer 2007). Toki offered a suggestion for helping the children in line 10, although lines 6–8 indicated, despite its brilliant attempt, an insufficient explanation

Extract 7.5 Sota's group recording (5_03)

1	Ken: Is chocolate delicious?
2	Emi: Yes, it is.
3	Ken: Do you like chocolate?
4	Emi: Yes, I do. Chocolate is so 甘い[sweet].
5	Three students (including Ken): Sweet. Sweet.
6	Emi: Sweet. Do you like chocolate?

Extract 7.6 Sota's group recording (1_02)

1	Toki: Have you seen this trademark?
2	Atsu: No.
3	Toki: Oh, really? It's a fair trade mark.
4	Atsu: フェアトレードって何?[What is fair trade?]
5	Toki: それを今から説明します。[I'll explain it now] This is a cacao tree.
6	It's made into chocolate…えー [well] These children work hard. Some of
7	them, them have never been to school. They, they have never seen chocolate.
8	Fairtrade can solve these problems.
9	Atsu: Right?
10	Toki: Right. Your shopping choice will help them too.

of fair trade. Atsu did not give any peer feedback here to improve the speaker's content understanding.

7.4.7 Unreported, Incidental Self-Assessment of Language and Content (Types 17 & 18)

Students' unreported incidental self-assessment may include immediate self-correction involving noticing where they could improve and self-reflective comments on how well they have done on a task. For example, in line 4 (Extract 7.7, where language goal included the use of interrogatives), Daiki responded to Miz's question. He formulated the sentence correctly in the beginning, but ended up making an ill-formed sentence via self-correction, presumably due to overgeneralization from recently learned present perfect constructions.

Incidental self-assessment of content was not clearly captured in the data. The methodological limitations of the present study precluded reporting substantive findings for this category. Capturing this aspect of A*f*L would require students to wear a microphone, as in some SLA studies (e.g. Saville-Troike 1988). Alternatively, student stimulated recalls may also facilitate the gathering of information about these two types.

Finally, the empty Table 7.1 (p. 170) is filled in with those A*f*L activities, in which the teachers and students engaged in various assessment episodes, as displayed in Table 7.7. As indicated by dashes, six types were still unused. It is not

Extract 7.7 Aki's group recording (226_10_02)

1	Miz: How many CDs do you have?
2	Mina: Many CDs.
3	Miz: How about you, Daiki?
4	Daiki: I don't, I've not, I have not CDs.
5	Miz: Oh, no.

Table 7.7 Summary of the sample A*f*L observed in the data

Grade Report	Planning	Assessor	Activity Focus	
			Language	Content
Reported	Planned	Teacher	*Worksheets, Skit	–
		Peer	–	–
		Self	–	–
Unreported	Planned	Teacher	*Vote counting, Observation	*Oral Survey, Worksheet
		Peer	*Tick slip	*Tabletop whiteboard
		Self	*Reflection sheet	*?Reflection sheet
	Incidental	Teacher	*Feedback	*Questioning, Feedback
		Peer	*Feedback	*Questioning
		Self	*Self-correction	–

Notes. Sample A*f*L activities in the study are listed for each category. Hyphens indicate assessment opportunities missing in the lessons. A question mark ('?') indicates that assessment of 'content' seemed not to be strongly focused. An asterisk ('*') indicates a category (not a particular activity in the slot) used in both Aki's and Sota's lessons

surprising, since the main goal here was to describe A*f*L, that most of reported, planned types, which are typically used for summative purposes, were found unseen. Overall, however, both teachers engaged students in A*f*L activities that required deeper processing of content, which were often beyond their language levels.

7.4.8 L1 and Translanguaging in AfL

An additional feature of the soft CLIL lessons that is worth noting here is L1 use and translanguaging, which seemed to be prevalent in various A*f*L in the focal lessons. In CLIL, where language support is usually in great need for content learning, teachers harbour mixed emotions about translanguaging or purposeful multilingual use for content learning (Lasagabaster 2017). Several CLIL guidebooks (e.g. Mehisto and Ting 2017) suggest limiting the use of L1 so as not to miss out on language learning opportunities. In regular CLT lessons, some teachers discourage or even prohibit the use of L1. In Japan, the Course of Study for junior high school (MEXT 2017a) declares the monolingualism to be a general rule of L2 instruction. Both teachers in the present study seem to ignore this rule and allowed the use of L1 to facilitate content learning, which resulted in learners translanguaging at times.

In Aki's lesson, for example, group 4's opinions in Table 7.4 (Sect. 7.3) were a unique mix of English and Japanese. In this inter-lexical translanguaging sentence, the student amalgamated Japanese functional words (particles) with English content words (subject and complement) to express the reason why the student, if they were Jimmy, would not arrest their friend, Bob. This complex mixture of the two languages appearing in one sentence in alternating fashion seems to be the students' risky attempt to avoid the easy way out of L1-only sentence, such as the one by group 7 in Table 7.4. In fact, this attempt manifests "the best of their creativity" (Li 2018, p. 22) by drawing on their available repertoire or a path to developing into multilinguals by thinking truly multilingually, rather than flouting the English-only rule.

Permitting the use of L1 allows the slow students to find their own voices or else to be marginalised due to their limited proficiency. Aki contended that she "want(s) them (slow learners) to maintain their motivation to manage to communicate by mixing Japanese and English or by using only keywords (not sentences)… if they lose motivation, they won't grow…" (stimulated recall, translated by the author). Sota abandoned his monolingual approach 2 years ago, since he had been convinced by its theoretical rationales in the literature. He observed "slow learners' positive change in their look—especially those who tried to shun English" (stimulated recall, translated by the author). Given the fact that most students fell into CEFR A2 level or below, L1 use likely facilitated democratic participation by giving slow learners opportunities to voice their opinions (Creese and Blackledge 2010) even in the challenging soft CLIL tasks.

Another aspect that might promote translanguaging in the Sota's lesson was mediation from L1 input text (video clips) to discussion in L2. He recognised that the topic, fair trade and child labour, overlapped with peace education, the goal of another subject, the Integrated Study. This cross-curricular connection allowed him to feel justified in using the L1 input text to enhance student absorption of the issue. Aki also reported, in another unit, her cross-curricular treatment of Severn Suzuki's UN speech, used both in the English and Ethics lessons on environmental issues. She brought a Japanese translation of the whole speech to promote student grasp of the issue, since the one appearing in the English textbook was abbreviated. In this context, bilingual input thus seems to enhance content understanding, while translanguaging output, i.e., participating in the discourse, brings about opportunities to take the issues to deeper levels. Aki felt "since we've done it this way (L1 use/translanguaging), the ideas presented have been deepened. (Without L1 support) their responses would have been shallow" (stimulated recall, translated by the author).

Both teachers seemed to be aware that the challenging soft CLIL activities—labelling the protagonists' behaviours by text analysis, reasoning in a hypothetical situation, defining fair trade, and discussing ways to help poor children—were beyond the majority of students' language levels. This might make translanguaging necessary. In the various A*f*L episodes, translanguaging may function to enhance student content understanding, democratising participation, and hence maintaining motivation to learn. No less important is it demonstrating a different and

complementary picture about learners' abilities at the moment and its developmental stages in content and language.

7.5 Discussion and Conclusion

This study described the ways in which the two experienced teachers and students employed a number of A*f*L activities in junior high lessons, in which CLIL is not a formal part of the curriculum, for the purpose of promoting and gaining insights into both content and language learning. They have run soft CLIL lessons by employing various unreported, planned, and unplanned A*f*L of content and language from careful observation of the on-task performance of pair/group work, student responses to questions, and various worksheets filled out by peers and learners themselves. These different A*f*L types, identified in the simple A*f*L framework, complementarily support students and teachers in continuously monitoring and improving language and content learning. The data suggest that regardless of planning, all unreported A*f*Ls except incidental self-assessment of content are observed. Methodological limitations seemed to impede observation of the last category of the framework. While this may be unfortunate, the data fit into and thus support the proposed framework for classroom use.

One strength of the framework lies in the inclusion of perspectives from multiple participants. Multiple information provided from multiple assessors supports the feasibility of A*f*L. If all the participants—teachers, students themselves, and peers—co-construct learning, they, too, should co-construct the assessment process, in which each piece of information builds on interactive sharing of the continuous, be it planned or unplanned, learning records.

However, the data from the present study indicated the need for the two teachers to enhance content peer assessment. For example, the content peer assessment rubric the students used in Table 7.6 contained the descriptors of functional aspects but not the specific target content. The teachers seemed unaware of this feature of the rubrics. In Extract 7.6, the peer could not point out insufficiency in the speaker's definition of fair trade. While this may have to do with his language proficiency level, the students seemed unaccustomed to giving feedback about the content. While both examples may indicate that content is of secondary importance in the EFL context, they underscore the premise that successful A*f*L requires assessors' assessment literacy (Mehisto and Ting 2017).

Another issue concerns the use of L1 in teaching and assessment. The content focus seemed to make teachers compromise to permit the use of L1. In this soft CLIL context, L1/translanguaging seems to function to promote a good balance of both content and language learning. L2 discussion based on L1 source text—unusable as a linguistic resource yet beneficial as a conceptual resource—created challenges as well as scaffolds for learners. One should therefore not overlook potential positive aspects of L1/translanguaging working in the lessons such as content enhancement, democratic participation, resultant motivation maintenance (as

discussed above), and gaining unique insights into learning process (e.g. Antón and DiCamilla 1998). This challenges the monolingual philosophy in EFL instruction, and a good balance between them needs to be sought after.

Two further issues concern the design of the framework. First, as seen in Sect. 7.4.5 above, demarcations between planned and incidental A*f*L are not as distinct as they were originally conceived to be. In fact, incidental A*f*L can be frequently embedded in planned A*f*L (Wang 2017). Second, methodological difficulty in observing types 17 and 18 may suggest a need for modification of the framework. These issues need to be addressed to improve the framework.

Among several limitations of the present study, a notable one could be the lack of a control case to which the selected teachers could be compared. Such data are critical for identifying whether the soft CLIL classes use content and language A*f*L in a balanced manner. Another limitation of the present study is the deductive nature of the analysis. While the 'theory first' approach of the present study allows hypothesis testing, potential A*f*L use outside of the framework could have gone unrecognised.

Description of the teachers' soft CLIL approaches to A*f*L in this contribution offers food for thought for both CLIL and non-CLIL practitioners. If CLIL proved to provide students with greater benefits than regular foreign language courses, the merits and values of CLIL practice should be imported into non-CLIL contexts. The content focus found in the present study exemplifies how regular CLT courses can integrate content and balance with language in various A*f*L activities. Implementing a soft CLIL approach in the CLT lessons would likely advance learning beyond language-only focus lessons.

As shown, A*f*L practices entail many planned and incidental collaborative assessment activities working complementarily for monitoring and promoting learning continuously. If the teacher-fronted lessons still linger in some CLIL courses, the incorporation of A*f*L at the planning stage may change the whole lesson structure. By employing various A*f*L, the CLIL teachers become aware of what is observable and unobservable in lessons. For example, in the planning stage, one can list all activities of the lesson and fill in the grids of the framework with them, as done in Table 7.7. By doing so, the teacher will realise what A*f*L activities are needed, given the lesson purposes. This supports the teachers' metacognitive skills in managing assessment plans. The A*f*L framework has thus the potential to facilitate the use of A*f*L activities by both soft CLIL teachers and hopefully CLIL teachers alike. The framework also helps the teachers reflect on their own lessons by verifying whether what is planned is actually seen.

Overall, the A*f*L practice of the two CLT teachers has highlighted the view that *every language teacher is a content teacher and vice versa* (see Sect. 11.3 of Chap. 11, this volume). Content focus is an indispensable ingredient of language teaching and assessment even in a CLT context, because it fulfils authentic purposes for language use, and thus motivates learners to participate in challenging discourse. Taking advantage of plurilingual input and output, they can seek to achieve the meaningful learning. Multiple assessment provides multiple opportunities for learners to demonstrate their learning in content, language, and trans-language, in which

the assessors can assess and interpret valuable information about where they are and where to go. This chapter has provided a case for expanding the border of CLIL to accommodate CLT (with CLIL-orientation) mapped right next to task-based language teaching (TBLT) on the CLIL continuum (see Ortega 2015 for its relationship with TBLT).

The chapter has also captured a snapshot of the current state of Japanese junior high schools, which are on the verge of moving slowly from CLT to soft CLIL. Close scrutiny of the lessons in the midst of a gradual transition in language teaching policy has shed light on both caveats and learning opportunities for teachers in preparation for the change. The next decade will definitely accelerate the shift towards CLIL in curriculum, instruction, and assessment policy in Japan. This chapter has described an initial attempt of non-CLIL teachers heading in that direction.

Acknowledgements This work was funded by the Japan Society for the Promotion of Science, under Grant number 26381175 (Grants-in-Aid for Scientific Research). My sincere thanks go to our editors, Mark and Dmitri, and the reviewers for their valuable and insightful comments. I am also grateful to all the participants and Sayaka Kokatsu.

Further Reading

Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). New York: Routledge.

Bostwick, M. (2001). English immersion in a Japanese school. In D. Christian & F. Genesee (Eds.), *Bilingual education* (pp. 125–137). Alexandria: Teachers of English to Speakers of Other Languages.

Brown, A. (2013). Uses of language assessments. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5979–5985). Chichester: Wiley-Blackwell.

Garcia, O., & Kleyn, T. (2016). Translanguaging theory in education. In O. Garcia & T. Kleyn (Eds.), *Translanguaging with multilingual students: Learning from classroom moments* (pp. 9–33). New York: Routledge.

Topping, K. J. (2013). Peers as a source of formative and summative assessment. In J. H. McMillan (Ed.), *SAGE Handbook of research on classroom assessment* (pp. 395–412). Thousand Oaks: Sage.

References

ALC Press. (2018). *TSST (telephone standard speaking test)*. https://tsst.alc.co.jp. Accessed 21 Feb 2019.

Antón, M., & DiCamilla, F. (1998). Socio-cognitive functions of L1 collaborative interaction in the L2 classroom. *The Canadian Modern Language Review, 54*(3), 314–342.

Brookhart, S. M. (2004). *Grading*. Upper Saddle River: Pearson.

Brookhart, S. M. (2010). Mixing it up: Combining sources of classroom achievement information for formative and summative purposes. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 279–296). New York: Routledge.

Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal, 94*(1), 103–115. https://doi.org/10.1111/j.1540-4781.2009.00986.x.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438–481. https://doi.org/10.3102/00346543058004438.

Dalton-Puffer, C. (2007). *The discourse of CLIL classrooms*. Amsterdam: John Benjamins.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Gibbs, G. R. (2007). *Analyzing qualitative data*. London: Sage.

Hill, K. (2012). *Classroom-based assessment in the school foreign language classroom*. Frankfurt: Peter Lang.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing, 29*(3), 395–420. https://doi.org/10.1177/2F0265532211428317.

Ikeda, M., Pinner, R. S., Mehisto, P., & Marsh, D. (2013). Editorial: CLIL in Japan. *International CLIL Research Journal, 2*, (1).

Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing, 25*(3), 385–402.

Kokatsu, S. (2018). *How do the Japanese junior high school teachers of English assess students' English speaking skills?* Unpublished Master's Thesis, Ibaraki University.

Lasagabaster, D. (2017). I always speak English in my classes: Reflections on the use of the L1/L2 in English-medium instruction. In A. Llinares & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 251–267). Amsterdam: John Benjamins.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol/Buffalo: Multilingual Matters.

Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30.

Mavrommatis, Y. (1997). Understanding assessment in the classroom: Phases of the assessment process—The assessment episode. *Assessment in Education, 4*(3), 381–399.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge, UK: Cambridge University Press.

MEXT (Ministry of Education, Culture, Sports, Science, and Technology) (2017a). 平成29年度英語力調査結果(中学3年生)の概要 (precis of the 2017 survey results of third-year junior high school students' English ability). https://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2018/04/06/1403470_01_1.pdf. Accessed 6 March 2019.

MEXT. (2017b). 中学校学習指導要領解説: 外国語編 (Commentary on course of study for junior high school: Foreign languages). Tokyo: Kairyudo.

Orgeta, L. (2015). Researching CLIL and TBLT interfaces. *System, 54*, 103–109. https://doi.org/10.1016/j.system.2015.09.002.

Philip, J., Adams, R., & Iwashita, N. (2014). *Peer interaction and second language learning*. New York: Routledge.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462. https://doi.org/10.1177/026553220101800407.

Saito, H., & Inoi, S. (2017). Junior and senior high school EFL teachers' use of formative assessment: A mixed-methods study. *Language Assessment Quarterly, 14*(3), 213–233. https://doi.org/10.1080/15434303.2017.1351975.

Saville-Troike, M. (1988). Private speech: Evidence for second language learning strategies during the "silent" period. *Journal of Child Language, 15*, 567–590. https://doi.org/10.1017/S0305000900012575.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Maidenhead: Open University Press.

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second and foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–274). Berlin/Boston: De Gruyter.

Wang, X. (2017). Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly, 14*(4), 312–327. https://doi.org/10.1080/15434303.2017.1393819.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). New York: Routledge.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington: Solution Tree.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. West Palm Beach: Learning Sciences International.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks: Sage.

Chapter 8
Does A*f*L Promote Discussion in CLIL Classrooms? Exploring A*f*L Techniques and Their Effect on Classroom Communication

Rachel Basse and Irene Pascual Peña

8.1 Introduction

As strongly argued in Chap. 1 of this volume, the role of assessment in CLIL (an umbrella for various educational approaches where an additional language is used to teach subjects other than that language; Wolff 2007) is an under-researched area needing attention (Llinares et al. 2012; Coyle 2010) to investigate its potential in facilitating the learning of language and content in tandem. The aim of this chapter is to investigate the potential of assessment for learning (A*f*L) to promote communication through discussion in Content and Language Integrated (CLIL) classrooms. One of the key aims of CLIL learning which distinguishes it from traditional English as a Foreign Language (EFL) teaching is the focus on oral communication (Lasagabaster 2008). Communication is cited by Coyle (2008) as part of the 4 Cs framework of CLIL, as learners are encouraged to foster language learning by engaging with content through meaningful communication.

Due to the dual focus on learning content and language in CLIL contexts, communication is key in order to provide learners with ongoing learning aims and feedback on their progress toward these aims. Much of this takes place through communicative means, with teachers using a variety of strategies to communicate learning aims, encourage learners to display their knowledge, and deliver feedback. How does this method of assessment affect communication in CLIL classrooms? In this chapter, we consider the strategies used by A*f*L teachers and their effect on classroom discourse in the CLIL context in Spain, which falls into the lower part of Davison (2008) model (see Chap. 1, this volume).

R. Basse (✉) · I. P. Peña
Universidad Autónoma de Madrid, Madrid, Spain

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_8

8.2 Assessment for Learning (A*f*L)

8.2.1 The AfL Approach

We refer the reader to Chap. 1 of this volume when considering the conceptualization of A*f*L as it relates to classroom pedagogy and teaching in various contexts. Much of the time, A*f*L approaches require changes in classroom pedagogy (Black and Wiliam 2003), calling for more active, student-centred learning (Ross et al. 2006) in which teachers deviate from the traditional role as an authority figure and instead act as a mediator (McCallum et al. 2000). This, oftentimes, involves a greater involvement from learners, as they are given more autonomy as learners in the classroom which opens a dialogue in which teachers share learning aims with learners and work together to set achievable goals (Black and Wiliam 1998a; Assessment Reform Group 2002; Bell and Cowie 2001; Wiliam 2006). The nature of A*f*L involves integrating learners in the learning process and co-constructing classroom discourse in order to generate evidence, which is later used by the teacher to determine where gaps in learning lie.

Congruent with Davison's classroom assessment cycle, A*f*L encompasses several key components such as establishing learning aims, using effective questioning techniques and providing learners with continuous and appropriate feedback when necessary (Black and Wiliam 1998a). Some strategies used by A*f*L teachers include:

- Sharing learning aims and establishing clear learning objectives with learners by displaying WALT (What are we learning today…) and WILF (What I'm looking for…) posters to make objectives transparent from the outset of the lesson
- Establishing assessment criteria through the use of 'Can do' statements to make apparent what learners need to achieve to earn a certain grade (see also Chap. 2, 4, and 5, this volume) and identifying short term and long-term learning goals (Davison 2008)
- Using effective questioning techniques, such as open or display questions, to generate communication from learners to display their knowledge
- Continuous feedback to learners—feedback targeting areas for improvement
- Peer and self-assessment sessions
- Analysing assessment information and looking for overall patterns (Davison 2008)

Teachers use these techniques as means of giving learners more responsibility and agency in the learning process and creating and interpreting evidence to see where the learners are in their learning and the best way to get them where they need to be to close the learning gap (Black and Wiliam 1998a). Self and peer assessment, for example, may promote students' self-reflection of their own learning and lead to them gaining more agency in the classroom (see Chap. 3, this volume). This, in turn, may require the need for meta-language which students can use to promote this reflection. Some of this language may be learned through the continuous feedback delivered by teachers.

According to Black and Wiliam (1998a, p. 52), formative assessment can only be useful and effective if it happens in interaction. Meaning is constructed through interaction and all the interlocutors have a joint responsibility for the creation of meaning, identities, and events. In interaction, the teacher can respond to and reorient learner's thinking while building on what is known to the learner, all the while adjusting teaching so as to ensure that the learners' needs are being met. Ways to respond include opportunities for improvement in learners' knowledge or inhibiting learners' opportunities to learn if the teacher is constantly looking for the right answer (Black and Wiliam 1998b, p. 143). A*f*L allows learners to be involved in interaction and to develop skills for meta-cognitive reflection about their learning (Rea-Dickins 2001, pp. 452–453). It may be assumed that opportunities for informal assessment are embedded in good classroom practice (e.g., questioning, interaction, feedback) (Rea-Dickins 2001, p. 457). However, putting A*f*L into practice also requires teachers to be conscious of utilizing assessment opportunities in the classroom and creating learning opportunities for students, either as the lesson unfolds or once an assessment has been given on a learners' performance (Hill 2017; see also Chap. 9, this volume).

As touched upon in Chap. 1 of this volume, recently, there has been some criticism of A*f*L by researchers in the field of education, questioning its practice and aims. With its prolific implementation in the UK in order to meet certain educational standards and the influence of governmental educational bodies, researchers wonder whether the focus is too performance-driven rather than rooted in helping students' learning. Oftentimes, the packaging of assessment initially designed as summative for formative purposes as formative assessment may create confusion (Ninomiya 2016). Others have raised theories regarding the "authentic" A*f*L and the Assessment for Learning Strategy (A*f*LS), which is the version of A*f*L adopted by larger initiatives (Ninomiya 2016) which may alter the original intended purpose of formative assessment or adapt it to serve a larger purpose of meeting national standards rather than focusing on individual learning.

Christodoulou (2017) raises questions about formative assessment and its implementation in the mainstream media and cites several factors in her critique. These include an exploration into assessment practices implemented in the classroom, reliability and validity of these practices and comparative judgement (Christodoulou 2017). The importance of raising these issues cannot be understated: it is, in fact necessary in order to call attention to shortcomings in A*f*L, and to further define it and develop good practice. While it is true that A*f*L may have risen in popularity based on the promise of bettering educational outcomes, its roots lie in classroom practice; taking into account the core pillars, educators may use A*f*L as a tool to inform their teaching to facilitate learning.

8.2.2 *The Importance of Classroom Interaction for Learning Opportunities and AfL*

The study of classroom discourse is the study of the language that teachers and learners use to communicate with each other in the classroom (Zhang Waring 2008). It is through spoken discourse that teaching and learning take place, making interaction and learning inextricably linked (van Lier 1996; Vygotsky 1978). The most common pattern in classroom discourse is Initiation-Response-Feedback (IRF) (Sinclair and Coulthard 1975). Hall and Walsh (2002, p. 196–197) distinguished between IRF and IRE (Initiation-Response-Evaluation). Either the learners have to show what they have learned and the teacher assesses it (IRE, linked to viewing teaching as a process of transmission), or the teacher looks for the learners to be actively engaged and participating in the discussion, so that IRFs can be opened up to more "mutually contingent interactional formats", linked to viewing teaching as a process of inquiry (van Lier 1996, p. 154).

According to Barnes (1975), different types of interaction lead to different types of learning (e.g. memorising, reasoning, creating). One type of interaction that has been acknowledged to be crucial to learners' learning is *dialogic interaction* (Wells 1999; Alexander 2004). In this type of interaction (close to Hall and Walsh's IRFs), discourse is jointly constructed by teachers and learners, helping the learners develop reasoning and inquiry skills. According to Mortimer and Scott (2003), in this kind of interaction teacher and learners explore ideas and viewpoints, creating new meanings, listening to each other and posing questions. Another characteristic of dialogic interaction is *contingency*, which means that the teacher's interpretation of learning evidence and teaching adjustments are context-dependent. Contingency increases learning opportunities and depth of learning (van Lier 1996; see also Black and Wiliam 2009), as it relates new and known material, sets up expectancies for what comes next, validates preceding and next utterances, is never entirely predictable or unpredictable, promotes intersubjectivity, and ensures continued attention (van Lier 1996, p. 193).

However, classrooms in which the teacher speaks most of the time and asks questions are more common, and learners are limited to answering those questions (Mortimer and Scott 2003; though see Chap. 10, this volume). This type of interaction is similar to Hall and Walsh's (2002) IREs, which make classroom interaction unequal and asymmetrical (Applebee et al. 2003; Hardman et al. 2003). Researchers do not, however, argue that a more authoritative presentational kind of interaction should never occur in classrooms; rather, both types of interaction should alternate depending on the teachers' goals at different moments of the learning process (Barnes 1975; van Lier 1996; Mortimer and Scott 2003; Alexander 2004; Wragg and Brown 2001).

Dialogic interactive teaching is fundamental in any type of classroom, but even more so in a CLIL classroom, due to its dual focus (Haneda and Wells 2008; Moore 2011): it is an opportunity for both language development and content learning. In the same way, this type of teaching is also crucial for the realization of A*f*L, as Bell

and Cowie nicely explain (2001, p. 539): "It is through the teacher-student interactions during learning activities (Newman et al. 1989) that formative assessment is done and that learners receive feedback on what they know, understand, and can do. It is also in these student-teacher interactions during learning activities that teachers and learners are able to generate opportunities for furthering the learners' understanding." Considering the important role that dialogic interaction plays in A*f*L, the present chapter studies the how interaction changes when principles of A*f*L are explicitly introduced into it and how then assessment becomes a part of this interaction.

8.3 Research Context and Methodology

8.3.1 Research Context: CLIL in Madrid, Spain

The implementation of CLIL in Spain, as mentioned in Chap. 1 of this volume, depends on the autonomous community in which the school is situated. Due to the integration of the bilingual educational system, teachers trained in content areas have been required to impart their subjects in English or other target languages, requiring veteran teachers to undergo additional in-service training to improve their own language skills. In general, many teachers—trained in subject area expertise rather than language teaching—see themselves as content experts and regard the changes to the system as challenging due to the dual focus of CLIL (Muñoz Lahoz and Navés 2007; Chap. 4, this volume). The perceptions of teachers, in this case, fall into Leung and Morton's (2016) "higher disciplinary orientation and less visible language pedagogy Quadrant" (see Chap. 1, this volume). In terms of the learning process in CLIL schools in Spain, the approach embodies a more traditional learning environment in which the roles of teacher and student hierarchies are clearly defined. This approach leads to a more teacher led classroom structure, which can inhibit student agency and result in decreased opportunities for discussion. Nevertheless, in schools in which CLIL is well-established, such as the ones in which the present study is based, teachers are beginning to move to the strategies outlined in Quadrant 3 in which content serves as a medium for teaching and learning language (see Chap. 1, this volume).

8.3.2 Data

The data used for this chapter is part of a bigger corpus collected in the 2010–2011 academic year. The corpus consists of 500,000 words over a total of 80 class sessions in 5 bilingual primary schools. For the purpose of this chapter, 22 sessions were selected from 2 different schools. The criteria for selection were based on the

two teachers who had the most experience implementing A*f*L. Incidentally, these teachers were native English speakers who received A*f*L training in their countries of origin (England and the United States). These two schools belong to the MEC/British Council Project, which started in 1996 in Spain and includes a CLIL curriculum with half of the content-based classes taught in English. The training of these teachers included A*f*L strategies and techniques. This kind of training, although frequent in other countries like England, Hong Kong, or Australia, is very recent in Spain and is limited to a number of pilot CLIL schools within the MEC/British Council Project.

The two schools were state schools located in middle-class districts, one of them in the centre of the city, the other one in the outskirts. Table 8.1 below illustrates the whole data set. From each school, one teacher collaborated in this project. The data used for this study were collected in 5th and 6th year of Primary Education, in which learners are, correspondingly, 10–11 and 11–12 years old.

In the *Subject* column, there are all the school subjects represented in the corpus: science, citizenship, and drama. The variety of subjects arises from the fact that different subjects are offered in English in the CLIL programme in different schools. In School 1, the subjects were citizenship and science whereas in school 2, the subjects were citizenship and drama.

Following *school*, *teacher*, and *subject*, Table 8.1 presents the school years which learners were in at the time of data collection: 5th and 6th year of Primary Education. As shown in the table, in each school there is one group in 5th year of Primary and another one in the 6th and last year of Primary Education. Next, the number of sessions that were recorded in each group and subject are given. As two complete didactic units were recorded in each group (each corresponding to a different subject), the number of sessions recorded in each subject varied depending on the length of the didactic unit. The next column of the table shows the total number of hours recorded in each group and subject (session length varied from 45 to 60 min), and the last column describes the content covered in each didactic unit.

Regarding the teachers, the teacher in School 1 had 21 years of teaching experience, both at private and state schools, and in different countries (United States and Spain). The teacher from School 2 had been teaching in state schools for 10 years, half of them in England, the other half in bilingual schools in Spain. Both teachers were native speakers of English.

Table 8.1 Data set

		Subject	Year	Session	Time	Didactic units
School 1	Teacher 1	Citizenship	5	8	6 h40	A circle of smiles The Giving Tree
		Science	6	8	6 h35	Sound Bones and Muscles
School 2	Teacher 2	Citizen-ship	5	4	3 h20	Emotions Being healthy
		Drama	6	2	1 h30	Word association Improvisation

Complete didactic units were video recorded at these two different times in each of the subjects and in each of the schools. During the recordings of the units, teachers were asked to maintain normal classroom activity, which was recorded for the purpose of the analysis.

8.3.3 Methodology

This chapter aims to explore the potential of A*f*L to enhance communication in CLIL lessons by providing learners with the opportunity to engage in more discussion-based learning facilitated by concrete A*f*L techniques. These techniques include:

- establishing clear learning objectives
- high-order questions
- delivering meta-feedback
- peer and self-correction sessions

We define meta-feedback as feedback that focuses on learners' learning (weak areas/improvements) and assessment (marks and assessment criteria) (Pascual Peña 2017). By saying students' learning, we mean that it focuses on students' weak areas and/or improvements, what they can do to improve learning. By saying students' assessment, this type of feedback may also concentrate on marks and the criteria for reaching each mark, thus being helpful for self- and peer-assessment. The research question that will be considered is as follows:

1. *How does the incorporation of AfL in CLIL settings promote interaction between teachers and learners?*

To consider this question, a selection of interactions between teachers and learners from A*f*L classrooms was chosen from the larger A*f*L corpus. These interactions were chosen based on the exemplar nature of A*f*L techniques being put into use by teachers. A qualitative analysis was then completed considering the dialogic characteristics being illustrated in the interaction and the role of A*f*L techniques in inciting classroom discussion.

8.4 Results

This section will illustrate different classroom extracts in which the A*f*L techniques mentioned above are prominent, starting with establishing clear learning objectives. When stating the objectives for the lesson, *WALT* and *WILF* (see Sect. 8.2.1) are two important elements of A*f*L. These concepts are displayed as posters to the learners and represent, respectively, what *We Are Learning Today* and *What I'm Looking For.* They are used by A*f*L teachers to make learning aims transparent to learners. The

former refers specifically to learning objectives whereas the latter is more related to success criteria.

In Extract 8.1, the teacher is sharing both aspects with the learners, so that it is clear what is expected of them. WALT and WILF are not only shared and discussed with learners, they are also accessible to learners at any time, since they are posted on the walls throughout the unit and referenced periodically by the teacher.

Instead of presenting the objectives in the form of teacher monologue, these are discussed in interaction. Quality interaction does not seem to be promoted, as learners are just reading aloud what is displayed on the posters and the practice is led by the teacher. Nonetheless, clarification of learning goals is an important element of classroom practice; students who understand what they are being expected to learn are more likely to make learning gains (Young 2005). In A*f*L learners are asked to reflect on the achievement of those goals, both internally and orally (in interaction), therefore an understanding of these goals should be achieved at the outset. If any student had a doubt about any learning goal, s/he would have the chance to discuss it in interaction with the teacher and the rest of his/her classmates.

In Extract 8.2, the first teacher's turn is an introduction of the lesson to remind learners of the topic of the unit.

In turn 5, the teacher asks a question for reason, which is a high order question, as it helps learners develop reasoning skills (Alexander 2004; Dalton-Puffer 2007; Pascual Peña 2017). She could have just started with the content of the lesson, but instead, she is making sure learners understood the importance of the content learned so far, thus validating previous content already explained and also next learners' turns. At the same time that validation is set, in turns 7 and 9, the teacher uses meta-cognitive questions. These questions ask learners to articulate viewpoints so that they can be explored and thinking skills required. These questions also help

Extract 8.1 Stating learning objectives using WALT and WILF

TCH:	I'm going to show you WALT and WILF like this [*unintelligible*] isn't it? I don't have to keep writing it. Can you see that? Is that big enough or not?
STU:	Yes
TCH:	Yes? Ok, who would like to read the WALT for me? Lucas? Could you please?
STU1:	((reading)) Understand what being healthy means. That there are many things we need to think about to be healthy. What these are and why they are important. Understand the reason why we celebrate World Health Day. Know how to tell people effectively ((pronounced incorrectly)) about health.
	[…]
TCH:	WILF, who'd like to read the WILF out? María? Would you [*unintelligible*]
STU2:	((reading WILF)) I can tell you about my healthy ((should be "health")) and what being healthy means. I can look at somebody else…else…he…health ((pronounced incorrectly))
TCH:	Health ((correction))
STU2:	Health and anal…analy…analyse ((pronounced incorrectly))
TCH:	Analyse ((correction))
STU2:	Analyse it.

Extract 8.2 Introduction to the daily lesson

1	TCH:	What are we going to look at today in Citizenship? Adriana?
2	STU:	Adriana: About the [*unintelligible*]
3	TCH:	About…excuse me?
4	STU:	Adriana: About the health
5	TCH:	About being healthy, good. Um, why are we looking at being healthy? Marcos?
6	STU:	Carlos: To have a good diet and…to be good.
7	TCH:	((correction))To be good, to feel good. Good, why else do you think, Bea?
8	STU:	Ivan: to know how eh, we are eh, how is the level of our health.
9	TCH:	Good, to know about ourselves and our own health, good, that's a good suggestion, Layla?
10	STU:	Layla: To make our longer life
11	TCH:	To make our lives longer, brilliant.
12		Um, who can remember who…what happened about two or three weeks ago? You had a special visitor to come in and talk to you.
13		Do you remember that?
14	STU:	Yes
15	TCH:	[*unintelligible*] people it's always things. Layla, can you tell me about that please?
16	STU:	Layla: Eh, a woman, a woman come to six, to year six, for the blackboard
17	TCH:	Yeah, the whiteboard, whiteboard ((correction))
18	STU:	Layla: Whiteboard. And they and she eh, talked to us about how we can do our life longer like [*unintelligible*] says and to have a healthy diet and and to don't em, maybe eh [*fumar*] ((mimes smoking))
19	STU:	Smoke
20	TCH:	No smoking
21	STU:	Layla: Or eh, for being better than [*unintelligible*]
22	STU:	And don't drink! ((interrupting))
23	TCH:	And don't drink and put your hand up as well, yes? Yes, ok, Andrés?
24	STU5:	And they talk about don't eat a lot of meat
25	TCH:	Not eating a lot of meat, ok, why do you think that we should not eat a lot of meat?
26	STU5:	Eh, because, it's bad for the heart ((pronounced incorrectly))
27	TCH:	Bad for your health, yeah, who can expand on that? Who can tell me a bit more about that? Antonio?
28	STU6:	For to not have [*grasa*]

learners to relate content with real life experiences. In fact, it can be seen that learners' responses are made up of at least one clause, even with subordinate clauses, as in turn 8. Especially long and complex is turn 18, in which the student is explaining the visit they had a couple of days back. Actually, that explanation and student contribution stretch across various turns (16, 18, and 21). Again, contingency is present (as throughout the whole extract), since those turns are both validating previous turns and setting up expectancies of what comes next. Contingency appears again when another student interrupts in 22. Even though he interrupts, he is listening to his classmate as well as paying attention to the lesson in general, which makes his contribution possible. The interruption can also be seen as a sign of desire to

participate in the discourse. It can be observed through the whole extract that discourse is constructed by both teacher and different learners.

Furthermore, learners are eager to contribute to this co-construction. Once again, in turns 25 and 27, the teacher uses meta-cognitive questions, their function being to promote learners' reasoning and thinking skills. Such questions, as we discussed earlier, are also a valuable help for learners to relate academic content with their everyday lives. All in all, several characteristics of dialogic interaction are illustrated in this extract. First, discourse is jointly constructed by teacher and learners. The teacher is not the only participant, it is not a monologue, but rather student contributions are equally important. Secondly, the teacher asks high order questions (questions for reasons and meta-cognitive questions), which make learners' responses relevant and help them develop reasoning skills and understanding. In addition, these questions do not normally require a simple and straightforward answer, but rather they encourage learners to think and be active and engaged in the interaction. Thirdly, contingency is clearly appreciated at different points throughout the extract. All these characteristics working together make interaction dialogic and learning-oriented, and learners, active participants and agents of their own learning processes.

While there is not an explicit language focus during these interactions, the teacher uses the learners' spoken production to discover gaps and help learners with their language skills. This is apparent in turns 18–20 when the teacher helps learners search for the correct vocabulary for smoking, and also in turns 24–25 and 26–27 when the teacher gives a recast to correct learners' errors in grammar and pronunciation. Without this opportunity for spoken production through classroom discussion, the teacher may not have been aware of or in the position to correct these linguistic discrepancies. Furthermore, the discrepancy in Lines 26–27 in which the teacher interprets Student 5's pronunciation error ('heart') as an error in vocabulary ('heart for health') serves as a reminder that content and language are constantly intertwined. The presence of these two layers is at the heart, so to speak, of each CLIL interaction.

Extract 8.3 below is an example of the teacher offering feedback in real time. Learners were required to self-assess in a class activity, but not only did they have to assess content learning but also behaviour. In this way, they had to evaluate themselves for 'respect'. Before learners' self-assessment the teacher helped them by

Extract 8.3 Connecting marks to lesson aims

1	TCH:	No, it's very easy to say no, no, [*notable, no or sobresaliente*[a]], no you have to ((points to poster)) look at the paper. So the first thing is respect, and respect to get a bien is to "work respectfully in class all of the time" and some people can't say that. The maximum for some people here is a sufi.
2		((class silent))
3	TCH:	Right. Right or wrong?
4	STU:	Right
5	TCH:	So let's try to get a sufi plus I'm going to try more in 30 min

[a]Notable and sobresaliente are the two highest grades learners may reach

reminding them of the criteria they need to follow to do the self-assessment successfully. In turns 1 and 5, the teacher is offering meta-feedback, which in this case focuses on the assessment process. The use of this type of feedback is also exemplified in Extract 8.4. Unlike in Extract 8.3, which focused on setting up criteria as it relates achieving a certain grade, in this extract the meta-feedback (in turns 1 and 6) is language-based and focused on the learning process: the teacher wants to move learners' learning forward. The use of this type of feedback is of vital importance, since it focuses on the learning and assessment processes through the identification of learners' strengths and weaknesses.

Extract 8.5 illustrates the A*f*L peer assessment technique of two stars and a wish, a strategy in which the teacher encourages learners to comment on their peers' work by providing two pieces of positive feedback (stars) and one area of improvement (wish).

In the extract above, learners are asked to assess their classmates' sentences. To do that, they had clear directions they had to follow. Learners' awareness of these directions is shown in different learners' interventions (see turns 7, 9, 22). Nonetheless, they still need the teacher's guidance in their assessment, as they are not only asked to assess in terms of formal requirements (whether they followed the instructions or not) but also in terms of content (see turns 8, 12, 23). Even though learners on their own would not have been able to assess their classmates at all levels required, they are capable of doing it with the teacher's assistance and scaffolding. The goal will be to withdraw this assistance with time, so that learners become more and more autonomous (Vygotsky 1978), meaning the ability to internalise mediation provided by the teacher and apply it in a different context. It is also worth noting how this episode of peer-assessment triggers a long exchange in which learners have the opportunity for discussing with the teacher and among themselves about each other's work. The teacher is not simply correcting the learners and providing the assessment herself, but rather she is engaging learners into interaction and discussion, creating a more dialogic interaction in which learners can acquire the necessary strategies to be able to provide a sound assessment. In other words, the interaction created by the teacher helps learners to assess themselves fairly and

Extract 8.4 Giving feedback in real time

1	TCH:	Your level is very high and here you are just changing the same idea and write it in a different way, ok?
2		Eh.. another sentence, Laura
3	STU:	Laura: Eh.. bones can fracture
4	TCH:	Bones can fracture
5	STU:	Laura: And
6	TCH:	One minute, if you say bones can fracture and then you can write "break" here ((writing the sentence and writing "break" under "fracture")), that is good because maybe somebody doesn't remember this word for the examination and they will put "break". So if you put both words, it's better to study, and then on the examination you could put the most difficult words that you want.

Extract 8.5 Peer assessment using two stars and a wish

1	TCH:	Ok? Now on your whiteboard, I'd like you to predict what's going to happen in the story. And these are some sentence starters for you. You can choose any of these, or if you have your own idea, you can write another idea starting with (*unintelligible*). Can you talk to the people next to you for ideas?
2		((something clatters on the floor)) three, two, one…read!
3	STU:	All: ((reading)) My opinion is that the boy is very selfish. I think that the tree is very generous.
4	TCH:	Ok, one positive comment about this…about these sentences because there are two, Ivan?
5	STU:	Ivan: em, is good for the [*unintelligible*] good for but, I don't know.
6	TCH:	Ok, he did it well, Carlos?
7	STU:	Carlos: He, em, underlined the verbs.
8	TCH:	OK, so he followed directions, but we're talking about the content. What about the… the sentence? Sonia?
9	STU:	Sonia: That he followed directions because eh, he wrote about em, em, his opinion and it's a very good sentence.
10	TCH:	Ok, so we've said many good things about his sentences but I don't agree with Sonia
11	STU:	No
12	TCH:	Because he didn't tell us how the story is going to end, so the English is good, he underlined the verbs, but he didn't tell us how he thinks the story is going to end. Let's see if this one followed the directions. Three, two, one, read!
13	STU:	All: ((reading)) the boy is going to go to the tree and say he wants something and the tree give it to he. The tree will die.
14	TCH:	Did this person follow directions?
15	STU:	Yes
16	TCH:	Very well. Does this person have a very good idea about what's going to happen in the story?
17	STU:	Yes
18	TCH:	It's very sad!
19	STU:	Because…
20	STU:	Yeah
21	TCH:	It's very, very sad. But this person followed directions. Another positive comment about this sentence, this idea? Eh, Julia?
22	STU:	Lucia: He also underlined the verb
23	TCH:	But about the idea because we see the full stop, we see the capital letter, but what about the idea? What did this person write? Carla.
24	STU:	Carla: that he write about…
25	TCH:	He or she, because I don't know
26	STU:	He did
27	TCH:	Whose is this?
28	STU:	He
29	STU:	Laura!
30	TCH:	It's Laura's. So it's a she.
31	STU:	Isabella: She, eh write about what is going to happen. And... that you, we know that the tree died.

completely. These strategies can be used at the moment of the interaction, but they will also be of help in the future and for internal student reflection.

8.5 Discussion and Conclusions

The purpose of this study was to investigate the potential of A*f*L strategies to enhance communication in CLIL classrooms. By promoting two-way communication between teachers and learners, the learning process shifts from one of teacher centred content unveiling to the co-construction of knowledge. Through analysis of these interactions, it becomes apparent that the integration of certain A*f*L strategies allows the teacher to shift from a role of authority to that of a moderator. Analysis of the classroom extracts presented shows how A*f*L techniques enable teachers to transform traditionally monologic classroom interactions into an opportunity for discussion with learners.

When considering these techniques in the greater picture of CLIL—namely the Spanish context—the use of A*f*L provided teachers with an opportunity to open up learning to their learners as an experience of co-constructing discourse. This was especially apparent in Extract 8.2 in which teachers used higher order questions to set up a process of discovery for learners, using a variety of questioning techniques to promote student involvement. These included:

- questions for explanations
- for reasons
- meta-cognitive questions

Integration of these question types led to complex responses on the learners' part, triggering their active participation in classroom discourse (Dalton-Puffer 2007; Pascual Peña 2017). This shift offered learners more agency, setting up the learning process as discovery. The use of meta-questions used by the A*f*L teachers encouraged learners to reflect on their learning throughout these interactions. This supports the claims made by Pascual Peña and Basse (2017) that metacognitive questions are seen frequently in CLIL classrooms with A*f*L trained teachers and require the use of interaction to support learners' learning. The use of feedback—namely meta feedback—was also seen in the data. The integration of this feedback in a dialogic interaction with learners allowed the teacher to link learning objectives to student performance in real time. This feedback encompassed both content and language related learning, demonstrating the versatility of A*f*L in catering to the CLIL mission.

Described in Chap. 1 of this volume as a process of discovering content and language, with learners co-constructing knowledge through dialogic interaction with others, the approach of fostering more dialogic expression breaks with the Spanish concept of CLIL—and other educational contexts, such as Finland (see Chap. 9, this volume)—which favours the teacher-led approach due to the fact that many teachers consider themselves to be content area specialists. This coupled with

the more traditional Spanish setting in which teacher is seen as authority limits opportunities for learners to showcase their language abilities and content knowledge. The integration of A*f*L provides opportunities for learners to increase their spoken production and offered the teacher a way to assess learners' speaking skills. For example, the use of peer assessment in Extract 8.5 allowed the A*f*L teacher to shift classroom culture and promote student autonomy in learning. By giving learners the opportunity to formulate and give feedback in real time to their classmates, the teacher simultaneously promotes autonomy in the learners and provides an opportunity to promote the communicative exchange of feedback, giving learners an active role in the classroom (Flórez and Sammons 2013). The strategies used by A*f*L teachers—in this case, asking learners to provide positive commentary and areas for improvement in their peers' work—encourages them to evaluate the work in a communicative way using the lens of the learning objectives presented to them at the outset of the lesson. The teacher integrated learners in a whole class discussion of student work, and created a continuous cycle of IRF patterns to generate meaningful classroom interaction.

The integration of A*f*L in CLIL classrooms has the potential to be a valuable tool for teachers who may have a more content-based focus to provide opportunities for language focus, as well, as noted in Chap. 1 of this volume, as content is discovered in a dialogue with others, learners' linguistic competencies (both L1, L2, and translanguaging, a process in which bilingual or multilingual speakers access different features of language in order to maximise communication potential; see Chap. 7, this volume) and allow content knowledge to emerge and develop. Throughout the exchanges, the learners' spoken production of content allows teachers to make note of and correct any weaknesses in language skills that learners may exhibit. Without these opportunities for learners to produce language, these gaps might have gone unnoticed. While language may not have been the explicit focus of these interactions, it is impossible for the teacher to ignore areas in which learners lack the vocabulary or the grammatical means of expressing their message. Similarly, these interactions allow learners to fill gaps in language in naturally occurring conversations, in real time. This organic learning of language is one of the key factors separating CLIL classrooms from traditional EFL classrooms. The encouragement of this spoken expression with gentle correcting by the teacher during classroom discussion along is preferable to learners and may account for the reasons CLIL learners are more creative and risk taking in their own language abilities.

This focus on spoken communication between teachers and learners as well as the learners themselves has been shown to enhance language development (Lasagabaster 2008) but also allows content to be discovered in dialogue. The data in this study reflect the potential for A*f*L to be used in CLIL classrooms as a tool to create meaningful interaction patterns which support both language and content-based learning aims.

Further Reading

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

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics, 1*(2), 1–38.

Leung, C. (2004). Developing formative assessment: Knowledge, practice and change. *Language Assessment Quarterly, 1*(1), 19–41.

Leung, C. (2005). Language and content in bilingual education. *Linguistics and Education, 16*, 238–252.

Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge: Cambridge University Press.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing, 18*(4), 429–462.

References

Alexander, R. J. (2004). *Towards dialogic teaching: Rethinking classroom talk*. Cambridge: Dialogos.

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730.

Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. https://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf. Accessed 4 Feb 2019.

Barnes, D. (1975). *From communication to curriculum*. London: Penguin books.

Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536–553.

Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa*, October, 1–13.

Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal, 29*(5), 623–637.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5.

Christodoulou, D. (2017). *Making good progress?: The future of assessment for learning*. Oxford University Press-Children.

Coyle, D. (2008). CLIL – a pedagogical approach. In N. Van Deusen-Scholl & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., pp. 97–111). Berlin: Springer Science+Business Media LLC.

Coyle, D. (2010). Assessment issues in CLIL. In D. Coyle, P. Hood, & D. Marsh (Eds.), *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated (CLIL) classrooms*. Amsterdam: John Benjamins.

Davison, C. (2008, March). *Assessment for learning: Building inquiry-oriented assessment communities*. Paper presented at 42nd Annual TESOL Convention and Exhibit, New York.

Flórez, M. T., & Sammons, P. (2013). *Assessment for learning: Effects and impact*. Evaluative Report. CfBT Education Trust, University of Oxford (England), Department of Education.

Hall, J. K., & Walsh, M. (2002). Teacher-student interaction and language learning. *Annual Review of Applied Linguistics, 22*, 186–203.

Haneda, M., & Wells, G. (2008). Learning an additional language through dialogic inquiry. *Language and Education, 22*(2), 114–136.

Hardman, F., Smith, F., & Wall, K. (2003). Interactive whole class teaching' in the National Literacy Strategy. *Cambridge Journal of Education, 33*(2), 197–215.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment, 6*(1), 1–17. http://www.altaanz.org/uploads/5/9/0/8/5908292/3.si1hill_final_formatted_proofed.pdf. Accessed 26 April 2018.

Lasagabaster, D. (2008). Foreign language competence in content and language integrated learning courses. *The Open Applied Linguistics Journal, 1*, 30–41.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol/Buffalo: Multilingual Matters.

McCallum, B., Hargreaves, E., & Gipps, C. (2000). Learning: The pupil's voice. *Cambridge Journal of Education, 30*(2), 275–289.

Moore, P. (2011). Collaborative interaction in turn-taking: A comparative study of European bilingual (CLIL) and mainstream (MS) foreign language learners in early secondary education. *International Journal of Bilingual Education and Bilingualism, 14*(5), 531–549.

Mortimer, E. F., & Scott, P. H. (2003). *Meaning making in secondary science classrooms*. Maidenhead: Open University Press.

Muñoz, C., & Navés, T. (2007). CLIL in Spain. In D. Marsh & D. Wolff (Eds.), *Windows on CLIL. Content and language integrated learning in the European spotlight* (pp. 160–165). Alkmaar: European Centre for Modern Languages.

Ninomiya, S. (2016). The possibilities and limitations of assessment for learning: Exploring the theory of formative assessment and the notion of "closing the learning gap". *Educational Studies in Japan, 10*, 79–91.

Pascual Peña, I. (2017). *Assessment for learning in primary CLIL Classrooms and its co-construction in classroom discourse* (Doctoral dissertation, Universidad Autónoma de Madrid). https://repositorio.uam.es/bitstream/handle/10486/680101/pascual_penna_irene.pdf?sequence=1. Accessed 2 Mar 2018.

Pascual Peña, I., & Basse, R. (2017). Assessment for learning in CLIL classroom discourse: The case of metacognitive questions. In A. Llinares & T. Morton (Eds.), *Applied linguistics perspectives on CLIL* (pp. 221–235). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.47.13pas.

Ross, P. M., Siegenthaler, S. L., & Tronson, D. A. (2006). Assessment for learning and motivation. In *Proceedings of the UniServe science assessment symposium* (pp. 120–125). Sydney: University of Sydney.

Sinclair, J. M., & Coulthard, M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. London: Oxford University Press.

van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. New York: Longman.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge: Cambridge University Press.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment, 11*(3/4), 283–289. https://doi.org/10.1080/10627197.2006.9652993.

Wolff, D. (2007). CLIL: Bridging the gap between school and working life. In D. Marsh & D. Wolff (Eds.), *Diverse contexts- converging goals. CLIL in Europe* (pp. 15–25). Frankfurt: Peter Lang.

Wragg, E. C., & Brown, G. A. (2001). *Questioning in the primary school*. London: Routledge.

Young, E. (2005). *Assessment for learning: Embedding and extending*. Retrieved May, 20, 2008.

Zhang Waring, H. (2008). Using explicit positive assessment in the language classroom: IRF, feedback and learning opportunities. *The Modern Language Journal, 92*(4), 577–594.

Chapter 9
Assessing (for) Understanding in the CLIL Classroom

Dmitri Leontjev, Teppo Jakonen, and Kristiina Skinnari

9.1 Introduction

With the social-constructivist turn in 1970s, the notion of educational assessment expanded from summative measures of learning outcomes to include a different purpose—promoting learners' abilities. Our conceptual starting point was Black and Wiliam's (1998, pp. 7–8) definition of assessment *for* learning (A*f*L) 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." This definition views assessment as a key part of instructional practices expanding the understanding of what constitutes assessment (see Chap. 1, this volume, for a detailed discussion). As Wyner (2014) notes, every teacher employs assessment *for* learning in some sense in order to promote their learners' abilities. The question still remains whether teachers always employ these consciously and build on these to adjust teaching and learning, thus making them a part of the classroom assessment cycle (Davison 2008; see Chap. 1, this volume).

From the practical point of view, this study started from our realisation that all teachers are capable of using different kinds of assessment instruments and approaches. The kinds of challenges that assessment in CLIL contexts involves were discussed in Chap. 1 of this volume. Here, we want to emphasise that a lack of shared assessment criteria and goals can result in incidental, unsystematic, and impressionistic classroom assessment practices. In addition, lack of awareness of assessment opportunities is likely to prevent teachers from using these opportunities systematically. Hence, for example, Wewer (2014, p. 150) has argued that assessment in CLIL is not an established practice, at least in Finland, in primary education, and with regard to assessing language.

D. Leontjev (✉) · T. Jakonen · K. Skinnari
Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland
e-mail: dmitri.leontjev@jyu.fi; teppo.jakonen@jyu.fi; kristiina.skinnari@jyu.fi

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_9

As CLIL involves the learning of content and language, student development in both areas should be assessed. The exact way this assessment can be organised depends on the way content and language are integrated in the particular CLIL classroom or classroom activities (Leung and Morton 2016; see also Chap. 1, this volume). That said, while the central idea of CLIL is that it is about integration of content and language, this does not mean that a particular assessment episode or procedure cannot have a focus on either the language or the content. We emphasise that, for example, teachers' content-orientation can result in that the focus on language can become incidental and as a 'side effect' of content teaching, with teachers being unsure about how to address learners' linguistic problems (Leung and Morton 2016; Wewer 2014). The reverse is true, as well. Thus, promoting teacher understanding of how assessment can be a tool informing teaching and learning of both content and language is necessary.

In this chapter, we explore how assessment and its challenges are part of Black and Wiliam's (1998, pp. 7–8) "activities undertaken by teachers" that aim at promoting students' understanding, which can be seen as a prerequisite for learning. We do this by studying first how secondary school CLIL teachers report on their classroom assessment practices in interviews collected in Finland and Austria. We then examine an extended episode of classroom interaction in a primary school CLIL physics class from a socioculturally informed Conversation Analytic (CA) perspective (for prior similar work, see, e.g. van Compernolle 2013, 2017) to illustrate the occasioned and context-embedded work of assessing and promoting a student's subject-specific understanding. One reason for studying classroom interaction is that teacher interviews sketch but a part of the picture of teachers' practices. Moreover, research into how practices are actually implemented in the classroom can expand insights gained from interviewing teachers (see, e.g. Tsagari and Vogt 2017, for a similar argument). Zooming in on micro-level interactional detail shows how the teacher can attend to a student's talk for what kind of conceptual understanding and knowledge it implies and attempt to advance that understanding as the interaction proceeds. Before the analysis, we will introduce the notion of teacher assessment literacy, suggest why it is an important part of assessment *for* learning, and outline a theoretical basis for viewing assessment as an integral part of the classroom interaction.

9.1.1 Teacher Assessment Literacy

Applying assessment *for* learning in the classroom contexts requires teachers to expand their understanding of the role and purposes of assessment in the classroom. With the shift of focus from assessment *of* learning to assessment *for* learning culture (see Davison and Leung 2009; Chap. 1, this volume, for a discussion), the understanding of assessment literacy has expanded to include knowledge of assessment processes and the ability to evaluate the impact of assessment (Fulcher 2012; see Chap. 1, this volume). Many teachers, however, admit that they are not prepared to meet these demands, and often their assessment practices are limited to those

with which they feel the most comfortable, i.e. assessment *of* learning (e.g. Tsagari and Vogt 2017). One challenge is that the expansion of what constitutes assessment leads to teachers having to reconceptualise their understanding of assessment (Scarino 2017), which can be difficult. Hence, teacher training that involves mere knowledge transmission may not result in tangible changes unless teacher beliefs and understandings are also taken into account and engaged (Livingston and Hutchinson 2017). This suggests that a mere awareness of different kinds of assessment tools and approaches beyond more traditional tests and essays is not enough for teachers to make these a part of their practices.

Teacher assessment literacy can be raised in different ways. Pascual Peña (2017) found that CLIL teachers who were trained in assessment *for* learning (A*f*L) strategies and techniques used significantly more guiding questions and feedback than those who had no such training, thereby attending to assessment opportunities in a more systematic way (see also Chap. 8, this volume). Even when no formal training is available to teachers, directing teachers' understanding of assessment is important in the process of developing teacher assessment literacy. Hill (2017) argues that a crucial aspect in developing one's assessment literacy is to realise that classroom interaction makes opportunities for spontaneous assessment available. She further describes a framework that can help teachers act upon such interactional assessment opportunities and make them a part of one's classroom practice. This involves considering such self-reflective questions as *What do you do in classroom-based assessment*? *What do you look for*? *What beliefs and understandings do you have*? and *How does the context shape your assessment practices*?

In the present chapter, we aim at raising teacher awareness of how classroom interaction affords opportunities for assessment *for* learning by illustrating how this kind of assessment can unfold. Namely, we will explore how assessment is embedded in interaction between a teacher and a learner in the CLIL classroom, thus demonstrating how exactly such assessment serves the two purposes delineated above—aiding teachers' understanding of their learners' abilities and promoting these abilities. The need for this exploration emerged not only from the previous research, as will be outlined in the following section, but also from a teacher interview study that is also discussed in this chapter and which informed our exploration of assessment as a part of classroom interaction.

9.1.2 Social Interaction and Assessment Practices

Assessment *for* learning can be organised and conducted in numerous ways (as is evidenced by the present volume). In this chapter, we focus on unplanned assessment occurring as a part of classroom interaction. Unlike previous research (e.g. Pascual Peña 2017), we are not aiming here at proposing an inventory of specific strategies teachers utilise when making use of these assessment opportunities. Rather, we will study the centrifugal process of interaction (see Chap. 1, this volume) in which teachers are sensitive to and build upon learners' contributions in such interaction in order to both make insights about learners' mediated abilities and to promote these

same abilities. To elaborate, building on Hill (2017), we argue that classroom interaction affords opportunities for assessing what learners can do or know, what lacks in their understanding, knowledge, or performance, and how learners' performance can be developed. It, therefore, becomes important for teachers to recognise these assessment opportunities and systematically apply what they have learned about learners in subsequent instruction and assessment. In the present chapter, we focus on the process in which assessment and teaching are integral and indivisible parts.

Social interaction can be seen as both a context of and a tool for classroom-based assessment activities. Even when teachers assess students by merely observing their work in the classroom, such work typically involves interaction with others. On the other hand, routine instructional practices have an evaluative/assessing dimension, which is clearly visible in what the interactional literature refers to as 'known information questions' (e.g. Mehan 1979). Commonplace in many educational contexts, these refer to questions whose correctness the teacher assesses via so called initiation-response-feedback/evaluation (IRF/E) sequences, not only to make students' existing knowledge visible but also to co-produce new knowledge (e.g. Koole 2010; Nystrand 1997; Rusk et al. 2017).

In L2 contexts, the close relationship between assessment, instruction, and interaction has been theoretically perhaps most elaborately developed within research informed by the sociocultural theory (SCT). The SCT understanding of learning as a mediated process that emerges in interaction underscores the importance of talk between the teacher and learners (and among learners). Interaction is a resource that can yield insights into learners' abilities and their areas of struggle, as well as help learners overcome them, provided that teachers notice and act upon such opportunities for assessment.

The notions of reciprocity and transcendence (e.g. Feuerstein et al. 2010) often inform assessment conducted within the sociocultural paradigm, allowing for structuring assessment as a dialogical and contingent process. *Reciprocity* has to do with how the learner responds when particular forms of support are available (see also Lidz 1991, p.100). The notion builds on the understanding that a mediator's (teacher or peer) support both constrains and creates a number of possibilities for a learner to respond. The mediator's insights obtained about learners' abilities, knowledge, or understanding help structure the following assistance. This is not to say that guiding learners is not already a part of teaching practices. However, seeing this guidance as assessment shifts the focus from completing the task and considering learners' responses in a binary way ('correct' or 'incorrect') to learning what learners are capable of and promoting learners' abilities.

Transcendence refers to how learners and teachers apply the prior knowledge and understanding that emerge while they together work on a similar task. For the learner, transcendence is about the extent that she/he is able to recontextualise the knowledge, abilities, and understandings to new contexts. From the perspective of the teacher, transcendence refers to the teacher building on what is learned about the learner previously, allowing for checking to what extent the learner is able to apply the previous knowledge in a new situation. In other words, achieving transcendence presupposes that teachers are systematic in tracing how learners connect prior insights to new contexts.

Reciprocity and transcendence compel us to conceptualise classroom assessment and instruction not just as complementing one another but as parts of the same process in which the teacher (or any other mediator) cooperates with the learner. This joint functioning, then, directs the learner's future independent performance. That is, as other contributors to the present volume also discuss (Chaps. 8 and 10, this volume), assessment as a part of classroom interaction becomes inseparable from teaching and learning.

9.2 Research Questions, Data, and Method

In this chapter, we set out to answer the following research questions:

1. How do CLIL teachers describe their practices for assessing and supporting students' understanding?
2. How is assessment involved in instructional activities in CLIL classroom interaction?

The second research question emerged from the data analysis we conducted to answer the first research question (as elaborated below). That is, the insights that we gained into teachers' assessment practices and behaviours, led us to explore how assessment is actually a part of classroom interaction and how this assessment can address some issues that were raised by the interviewed teachers.

To address the first research question, we draw on interviews of CLIL teachers from lower secondary schools which are part of larger datasets collected between 2013 and 2016 as part of two different research projects. The first set of interviews, conducted in Austria, Finland and Spain, focused on the role of language in language and content integration and the second one investigated teachers' beliefs on the new language-related curriculum issues in Finland. All of the interviews were analysed, but for the purposes of this study, parts of interviews from four teachers will be employed. Interviews conducted in Finnish have been translated for the present chapter. In Table 9.1 below, we briefly introduce the four teacher participants whose reports we analyse in the present chapter.

The classroom interaction data, used to respond to the second research question, comes from a sixth grade physics and chemistry lesson taught in English, recorded in 2016 as part of a larger research project in a Finnish school where CLIL instruction was a relatively recent development. The class was co-taught by two teachers

Table 9.1 Teachers in the interview excerpts

Teacher code	Description
T1	Austria, physics content teacher
T2	Finland, biology and geography content teacher
T3	Finland, physics content teacher
T4	Finland, physics and chemistry content teacher

(neither of whom participated in the interviews), a Finnish-speaking class teacher (CoT) and a native English-speaking content teacher (T). The lesson involved a group-based learning activity during which teachers mostly interacted with individual students and groups. In the analysis, we focus on one interactional episode between the content teacher and a student. The aim was to illustrate how assessment can take place through interaction between the teacher and a learner.

We began the interview data analysis by studying themes that emerged with regard to initially *assessment* and related terms as the keywords, and, as our data-driven understanding developed, also *(mis-)understanding*. We then scrutinised the interview data in order to find out how assessment was involved in teacher classroom practices as reported by the teachers themselves. In other words, we did not limit our analysis to cases in which teachers explicitly talked about assessing their learners, e.g. using words such as *assessment* or *evaluation*.

In order to examine the classroom data, we draw on the analytical perspective of conversation analysis, CA (for an introduction, see, e.g. Gardner 2004). In very broad terms, this means investigating (inter)action in its sequential context and analysing things like knowing, understanding, and learning as publicly observable phenomena (e.g. Maynard 2006; Kasper 2009). The analysis of what turns-at-talk 'mean' builds on the participants' *emic* perspective, i.e. on how they observably and publicly treat each other's turns and conduct. Moreover, along the lines of the sociocultural paradigm, we also view learner knowledge and understanding as not only displayed in the context of interaction but also directed and (potentially) promoted by that interaction. That is, in our analysis to follow, we consider that the way that the learner responds to the teacher's queries is developed by these same queries. Within the CA literature, there is at the moment no uniform stance on what actually counts as 'learning' or whether it can even be investigated without a specific learning theory such as the SCT (see, e.g. Hauser 2011; Jakonen 2018). Here, we do not wish to take a definitive stance on these matters but we aim to bring together the sociocultural theory and the micro-analytical perspective of CA (for prior similar work, see, e.g. van Compernolle 2013, 2017) in order to illuminate how assessment takes place procedurally via specific interactional practices. Specifically, we will, by way of sequential analysis, trace how teacher and learner turns in the interaction occasion and create a context for one another.

We note that the transcript of the teacher interviews is less detailed than that of the classroom interaction. The reason for this is that the purposes of the analyses were different. While the analysis of the former aimed at gaining insights into teachers' self-reported practices, so we did not need to look into such details as changes in intonation, gestures, etc., in the latter, we aimed at understanding how knowledge was co-constructed, so these became an essential part of our analysis.

9.3 Analysis and Results

9.3.1 Teacher Interviews

Considering that the interview data were collected for a different purpose (see Sect. 2) than discussing how teachers perceive assessment, we did not expect to find many instances of teachers explicitly discussing assessing their learners. Explicitly, the word *assessment* emerged rarely in the interviews, but when it did, the participants talked about assessment *of* learning, such as grading or exams. Excerpt (1) illustrates how T1 reacts when the topic of assessment was raised by one (I1) of two interviewers.

(1) Understanding assessment.

1	I1:	is the assessment of CLIL students that of course needs some cooperation so
2		would you say that you need to work hard on that?
3	T1:	d'you mean the grading of?
4	I1:	I mean grading, assessment
5	T1:	the students who are already in our school or?
6	I1:	yes those who are in your school
7	T1:	mm well this is the thing with the grading is, is kinda complicated because
8		I mean officially I'm not allowed to like grade their English, of course not
9		I even don't want that and for example if there's a like physics test or quiz and
10		there are questions in English, if they answer in German I'm fine

It is also notable that T1 understood assessment as the product of learning. This is expected considering the discussion of teacher assessment literacy in Chap. 1 of this volume, especially that teachers often are more comfortable in using assessment *of* learning. Furthermore, they do not necessarily consider other types of assessment as assessment, be it assessment of content, language, or both. In Line 3, it transpires that the teacher in her response limits assessment to grading, and although the interviewer's reaction rather sounds as enumeration, the teacher continues to talk about grading her learners. There seems to be a clear separation of responsibilities between the CLIL teacher and the language teachers in the school. It is interesting that the separation is also maintained/reinforced by the teacher's 'want' to steer clear from language assessment and her procedures to make sure that language troubles do not get in the way of assessing learner understanding by allowing the use of German in tests.

In addition to explicit assessment talk, the teachers talked about assessment-relevant issues when discussing *understanding*. Many teachers in the interview corpus raised the problem of interpreting how well their students understand a topic when they are copying the teacher's words or definitions from the course book. One example of this is provided in the following Excerpt (2), in which T2 responds to the interviewer's question about structures typical to texts that learners produce in assessment situations. Specifically, she is telling how she makes sure learners' understanding of difficult concepts transpires in their texts.

(2) Learners' understanding in essays.

1 T2: it would become in a way scientific kind of writing a scientific (.) answer or a text
2 how to start and use the concepts of that field and open them up and what is meant
3 by what and that (.) you answer in a logical and consistent manner opening up the
4 concepts […] I don't want any textbook answers how something has been explained in
5 the textbook but you should always (.) it's enough for me, or I want that it's explained
6 to me like she would explain to her best friend who doesn't know something or
7 doesn't understand.

T2's response implies that she finds it problematic that learners use somebody else's words, such as definitions from the textbook, without displaying understanding of the concepts. Specifically, she requires learners to open these up (Line 2). In other words, she wants learners to use everyday language rather than scientific terms. Yet, she does not state how she enables the use of easier language in her classroom, for example teaching learners such everyday language to explain concepts.

To be clear, teachers do not necessarily have an inventory of teaching practices for making sure that learners understand and are able to express their understanding. In the following two excerpts, teachers T3 and T4 tell how they make sure that learners understand complex definitions and concepts. These practices, we argue, are based on their experiences of what learners' needs are, in other words, on assessment of their learners, even if incidental and implicit. In Excerpt 3, T3 is asked to opine how scientific language is different from everyday language and share whether and how he teaches his learners to write scientifically. This leads T3 to describe his classroom practices and explain how he habitually deals with situations of non-understanding in the classroom.

(3) Teacher practices: making definitions understandable.

1 T3: I often talk about that if there are some definitions for, or some phenomena or quantities,
2 the definition might, might be very long and hard to understand and I'm trying to explain that,
3 that this is the way it has to be said because if it's said in a shorter way or in a somehow easier way
4 it might change the point or the meaning, but after we have written down that the, the
5 definition, then I'm trying to sort of make the definition to be more understandable so
6 that they will understand the definition when they, when they understand the phenomenon
7 I2: yeah, and does that that happen in, both the Finnish and the
8 T3: Yes
9 I2: in English
10 T3: both I think

In Lines 1–2, the teacher reports his talk in the classroom, from which it transpires that he can sense which topics can present problems to his students. In spite of initially assessing these situations as problematic for his students, he does not want to compromise the exactness of science (Lines 3–4). He knows, however, that explanations have to follow, so he makes sure that these definitions and phenomena they describe are understandable. The teacher goes on to say that for this purpose, he uses both English and Finnish.

In Excerpt 4 below, T4 was asked to explain what the language of physics is like and how she explains concepts to learners and whether she uses everyday language for this. T4 responds that she aids understanding and memorising in her lessons with everyday language and then proceeds to more academic language. According to her, understanding is hampered already in the beginning if the students do not understand what they are asked, so she tries to scaffold their understanding already in the beginning when introducing a new topic.

(4) Scaffolding learner understanding.

1 T4: in science the text is very, like we try to say everything in a clear and short way […] a text
2 may be really short, but every sentence has like a great meaning […], for
3 example when you talk about how atoms bind ((a scientific term in Finnish)) to each other, this kind
4 of things, then often the kind of fun or even funny examples from everyday language aid
5 understanding and memorising it's very hard to remember something that is difficult
6 that, for example electrons attract protons, but if you formulate it in another way,
7 you say that opposites are drawn to each other then it's much easier to memorise […]

This teacher describes how she predicts when learner non-understanding can arise in classroom situations and responds to it by using catchy and fun examples in everyday language (Lines 4–5) and further explanations (Lines 6–7).

Taken together, Excerpts 3 and 4 suggest that these teachers have procedures to help learners understand scientific phenomena and better remember them. These teachers' practices also suggest that they should have at least an implicit understanding of their learners' abilities, which, we assume, they base on their teaching experience accumulated in the interaction with their learners. Clearly, the teachers reported on their practices based on experiences rather than on actions that emerge in the immediate interactions with the learners. These practices are based on inferences that accumulated over time. However, these inferences are still based on learner assessment, even if implicit, which, we assume, were recurrent to the extent that they became generalisable for the teacher to base their teaching practices upon. That said, how exactly these practices emerged and how the interactions leading to developing these practices happened is not reflected in the teachers' reports nor it is to be expected, as teachers often report making assessment-related decisions in the classrooms based on their gut feeling (Wewer 2014). This necessitates an exploration of how teachers' assessment happens as a part of their interaction with learners.

Although the two examples above illustrate that teachers can, to some extent, predict and detect their learners' problems, this is not to imply that they always know how to address them. In Excerpt 5, T1 responds to the interviewer's question "So what do you think integration is? Is it like you said … you teach passive along teaching physics?" She reveals that she considers it important to know where her learners' problems lie even though she thinks these are problematic to address. This is interesting as the same teacher earlier claimed that she did not assess/grade learner language (Excerpt 1).

(5) Merging content and language.

1 T1: I mean I think that's kind of a utopic thing to do, that would be the best thing ever
2 if you really, you know, if you would be able to pick that up if there's a certain
3 problem and then go into the language part and then go back to the science part
4 that would be really perfect, I tried to do it a little bit but there are definitely much
5 more opportunities to implement that…

Here, T1, having earlier in the interview strongly positioned herself not as a language teacher, reports that she finds it problematic to pick up where her learners' problems lie and is not sure how to address these problems. What is less clear based on what she reported earlier is whether here she refers to only learners' problems with the language or both the content and the language problems, but it appears that she finds it difficult to address learner problems in language and content separately. The way she formulates the response still suggests that it is both. Namely, perhaps, owing to her insecurity about her ability to teach the language, she finds it difficult to see exactly where her learners' problems lie, i.e. whether they are due to the lack of linguistic proficiency or misunderstanding of the content or both, and then address it respectively. It is, however, clear that she finds it highly problematic due to the choice of words ('utopic' and 'the best thing ever'; Line 1), and this uncertainty comes from her own experience (Line 4), and her report suggests that this kind of assessment was not an established practice for her at the time of the interview.

9.3.2 Classroom Analysis

As we mentioned earlier, the interviews that we studied led us to explore how teachers' understanding of their learners can develop in classroom interaction. In this section, we illustrate how teacher-learner interaction in a co-taught CLIL classroom provides assessment opportunities which allow a teacher (T) to both build his understanding of a learner's (Simo) abilities and simultaneously promote those abilities. We investigate one nearly 3-min episode of desk interaction (Excerpts 6–8) where the teacher is observably 'cluing' a learner, i.e. "leading [him] to correct answers by small steps" (McHoul 1990, p. 355) as a way to promote self-discovery (Waring 2015) instead of providing an explanation. The overall interactional organisation of the episode resembles a chain of IRE/F sequences so that instead of closing down a sequence in the 'evaluation move' with an explicit positive assessment (Waring 2008), the teacher repeatedly reframes his question in different ways and, sometimes, acknowledges partial correctness of the student's prior answer at the same time. From a sociocultural point of view, this gradually gives the learner responsibility for his own performance (even though this performance is still regulated by the teacher) and goes beyond leading to correct answers to promoting the learner's abilities. In essence, it is not just the learner's performance or the teacher's direction of this performance, but a joint construction of the learner's understanding, which also allows the teacher to understand this learner's abilities in greater detail.

The interaction takes place as the teacher circulates in the classroom during a task in which small groups summarise previously studied textbook chapters on a poster. The groups move from one poster/desk to the next, so that by the end of the task, all groups are meant to have covered all chapters, and each poster should include one summary per chapter. However, because of a task-organisational mishap, the focal group of four learners is supposed to summarise a chapter (about energy saving) which they have already done at a previous stage. Despite this, the teacher still asks them to re-summarise the chapter—perhaps in order not to complicate the task organisation any further. Instead of moving to another group, he remains by the learners' desk and begins to probe their understanding of the topic. In the course of doing this, he gradually directs their attention to the role of thermal insulation, a topic that the Finnish-language physics textbook did not address in much detail. We enter the exchange as the teacher (T) presents his first question to Simo.

(6) Working with a learner's everyday experiences.

1.	T	we spoke about that the last day what's the
2.		↑easiest way to save energy, (.) in a house for example?
3.		(1.3)
4.	Simo	( )
5.		(0.6)
6.	T	say aga- ↑no:: (0.5) I mean that saves some energy
7.		but how would you save energy in a house better °than°
8.		(1.0)
9.	T	where do you <lose> most of the energy in a house
10.		(0.5)
11.	Salla	in the °(toilet)° ((SMILES AND FACES AWAY FROM TEACHER))
12.		(1.6)
13.	T	no
14.		(1.7)
15.	Simo	for me it's actually in my room
16.	T	in your room where in your room °(do) you lose a lot of energy°
17.	Simo	(on the)° computer
18.	T	'kay so you have all your computers plugged in an'
19.		stuff on standby? (0.3) yes
20.		(2.1)
21.	Simo	actually:: (.) not
22.	T	not (.) so you (un)plug them (over) at night (0.6) everything
23.	Simo	no
24.		(0.8)
25.	T	no so you're using some energy there. (0.7)

The framing of the question in Lines 1–2 connects the topic to a previous lesson and conveys that the teacher is asking the learners to produce knowledge that they should already have. While Simo's response is not audible on the recording, the teacher treats it as only a partially correct answer by first rejecting it but then softening the evaluation (Line 6). He then reformulates the question by specifying that he is after an efficient energy saving method instead of the 'easiest'. As no learner response is forthcoming, the teacher reformulates the question yet again (Line 9), so that it now hints on a strategy that learners can use to answer it: the easiest way to save energy can be identified by thinking where it is lost the most. This signifies a momentary shift of emphasis from probing the learner's understanding to mediating the learner's thinking. Here, the teacher segments the question so that the learner is first invited to identify a place or a device instead of offering immediately what could require a lengthy explanation of how to save energy. However, the expectation is that the response will then be used to address the 'main problem' of what are most efficient ways to save energy. As we shall see, this becomes the focus of the whole 3-min interactional episode.

Following a short side sequence where the teacher deals with a jokey interruption (Lines 11–13), Simo's response (Line 15) suggests most of the energy is lost in his own room. As an answer, such a personal account is not only a way to avoid giving an 'objective' response in the form of averages, percentages, or the like, but it also displays an understanding that the most significant part of the energy consumption of a house depends on individual habits. As Simo, prompted by the teacher, specifies that his energy use is down to the 'computer' (Line 17), the teacher provides a candidate understanding of what kind of energy use habit Simo's answer would mean (Lines 18–19). Having 'all computers' plugged in is an extreme scenario (see Pomerantz 1986), which, here, hints to the learner that there is essentially one possibility—to say no. Simo does just that (Line 21), and the teacher responds with an opposite candidate understanding that Simo unplugs 'everything' (Line 22). Explicating such contrasting digital appliance use habits allows the teacher to convey a sense that the electricity taken by personal appliances does not represent a significant part of total consumption, that appliances only eat 'some energy' (Line 25). Such an understanding then forwards the main interactional project of identifying how one can best save energy in a house.

From the perspective of gaining insights into what Simo knows about saving energy in the house, this implies that he still lacks understanding of how energy in a house can be saved. Thus, he has to be directed rather explicitly by the teacher, as will also transpire in the following interaction. It is, perhaps, because of this understanding that the teacher begins to steer the attention away from electrical appliances and individual variation in their use to less visible, yet more significant, forms of energy consumption (Excerpt 7).

(7) Hinting at insulation.

```
25. T       no so you're using some energy there. (0.7)
26.         what are the things in a- (.) house that (0.3)
27.         take the most energy that consume °the most energy°
```

28. Simo fridge and [freezer]
29. Anneli [but we had-] (0.4) <done> °this°=
30. T =I <know,> (0.3) now you [have to use] your brains to dig deeper=
31. Simo [( )] ((TO ANNELI))
32. T =where else could you save energy in a house
33. (2.0)
34. Simo (depends for) the people in
35. (13.0) ((T WALKS TO THE WINDOWS, KNOCKS ON ONE OF THEM. ON THE WAY BACK, HE STOPS, POINTS AT THE CEILING AND FLOOR, AND THEN RETURNS TO THE GROUP))
36. T what would you put there, ((POINTS AT THE CEILING))
37. (1.0) ((POINTS AT THE WINDOWS))
38. why are those windows like that,
39. (1.5)
40. T why do we not just have one- one (0.4) sheet of glass (0.4) ((MOVES HAND DOWNWARDS SEVERAL TIMES))
41. why do they have three sheets of glass? ((BOTH HANDS PARALLEL))
42. (1.3)
43. T what's (happening)
44. (2.5) ((CROUCHES IN FRONT OF SIMO))
45. T what's happening there
46. (1.2)
47. Simo ( ) they use some ( ) (additional) light source (0.5) ( )
48. T oka:y so you could be doing that (and) I have seen very many
49. houses new houses with <big> glass (fronts on them) (0.5)
50. so you're right they're trying to get in (0.4) extra heat there
51. (1.6) but why else do you need three (0.6) sheets of glass
52. (0.9)
53. Simo it keeps out the cold (more)
54. T keeps out the cold ↑more (.) so you don't have to do what
55. (2.0)
56. Simo use heat
57. T (okay) good (0.3) so- (0.7) on the (paper) ((WALKS AWAY))

Here, we focus our observations on what Simo's turn in Line 34—that ways of saving energy 'depend' on people—signals for the teacher in the current task context, and how it occasions a change in the teacher's instructional strategy. Before that, we briefly note that already the teacher's prior question in Lines 26–27 is formulated so as to avoid the possibility of receiving a personal account, unlike the questions in Excerpt 6 (e.g. Lines 6–9). First, the grammatical structure no longer employs the generic you, which is vulnerable to being interpreted as specific to its

recipient (as Simo's earlier response in Line 15 showed). Second, by asking for 'the things' that consume 'the most energy', the teacher conveys that such things are commonly found in many houses and thus invites a generalised or an 'objective' answer. 'Things' is a somewhat ambiguous reference term, but Simo's response—'fridge and freezer' (Line 28)—shows that he continues to search for significant energy consumers among electrical appliances.

Line 34 responds to yet another question reformulation (Lines 30, 32), which addresses another learner's complaint about the repetitive nature of the task (Line 29). Simo's turn is a nonconforming response (Raymond 2003) in the sense that it does not specify a location, device, or activity with an energy-saving potential in response to the teacher's 'where' question, and from the way the teacher treats Simo's turn it is clear that it represents an inadequate answer in this context. Thus, instead of ratifying (even a part of) the answer, the teacher decides to change his strategy. Namely, a long silence ensues during which the teacher walks to the side of the classroom, knocks on windows, and points at the ceiling and the floor. These embodied actions serve to highlight parts of building as significant for the answer (and thereby for energy saving). They and the series of two questions in Lines 36–38 direct attention to structural elements that insulate buildings against cold weather (double or triple glazing; roof structure), which are particularly important in climates such as that of Finland.

Simo's answer, due in Line 39, does not arrive, and the teacher specifies the question about windows in Lines 40–41, depicting the multiple glazing with a gesture. As Simo does not take a turn even then (Line 42), the teacher rephrases the question once more (Line 43) and pursues an answer (Line 45) before Simo finally responds. While the response (Line 47) is partly inaudible, the way the teacher builds upon it is visible in the structure of his turn in Lines 48–51: he acknowledges the relevance of Simo's answer and accepts it as partially correct ("you could be doing that") and elaborates on it by way of a real-life example ("I have seen very many houses"). This, again, is an example of the teacher's assistance building on the learner's responsiveness to previous assistance. The teacher acknowledges that one function of big windows is to save energy and only then readjusts the question ("why else"), also eliciting the word 'heat' and inquiring once again about the function of three sheets of glass in windows. Such guidance enables Simo to produce an explanation for triple glazing ("keeps out the cold", Line 53). The teacher ratifies the answer by repeating it and links the answer to energy saving with a designedly incomplete question (Koshik 2002) that explicitly directs Simo to complete the teacher's utterance and thereby use the terminology ('heat') introduced by the teacher.

It is not known what the teacher's intention was in leading Simo's thinking from the idea of 'keeping out the cold' to 'using heat' instead of accepting Simo's response. Still, the preceding interaction resulted in that Simo produced a scientific, rather than an everyday, explanation, which also contributed to the main interactional project of the exchange—ways of saving energy in the house. That is to say,

Simo's knowledge was co-constructed together with the teacher allowing the former to eventually verbalise in a scientifically appropriate way how energy can be saved by using triple-glazed windows and why. As Simo does this (Line 56), the teacher accepts the answer and leaves the group. Note that Simo's talk in the whole of this exchange is only minimal, and he requires substantial directing in order to produce this response. Nevertheless, it is Simo in the end who produces the response albeit with explicit support from the teacher. Should the teacher have not used the opportunity to assess the learner's understanding and instead have resorted to directly telling the learner that 'triple-glazed windows help us use less heat', these insights into the learner's abilities would not have been available. Furthermore, it is likely that Simo's understanding of the phenomenon of insulation would not have been developed with explicit support as it was with gradual directing to the idea of saving heat by having several sheets of glass in windows.

Excerpt 8 shows how the teacher walks back to the group within a matter of seconds to continue the discussion. His question (Line 60) invites Simo to make sense of the teacher's earlier pointing at the ceiling (Lines 35–36 of Excerpt 7) and seems to project talk about different physical materials and their insulation properties. In other words, even though this happens immediately after the previous episode, this can be regarded as transcendence. The teacher attempts, building on the insights that he obtained about what the learner knows, to help the learner to connect his emerging understanding of saving the energy to the roof, assisting thus the learner in reconstructing his knowledge to use in a new context.

(8) Coming back for elaboration.

```
58.          T AND SIMO
59.          (0.9)
60.          why would I be pointing to the ↑ceiling ((POINTS AND
             WALKS BACK))
61.          (2.4)
62. T        where do you think most of the energy (0.3) is lost
             in a house
63.          (1.8)
64. Simo     if there is no ( ) (1.2) solar panels °(on the roof)°
65. T        ↑no::: I wouldn't say tha:t (.) ((SHAKES HEAD))
66.          (1.4) what stops the heat getting out (0.3)
67.          what keeps the heat in your (↑room)
68. Simo:    (the) roof
69.          (1.3)
70. T        the roof okay (.) and if I just put up a roof made
71.          <of> (0.5) straw, (1.5) will that house be warm
72. Simo     no
```

```
73.        (0.8) ((T SHAKES HEAD))
74. T      ↓no:: (0.3) so what would I need to ↑do (0.5) to my
75.        roof to keep the heat in? ((KALLE RAISES HAND))
76.        (0.8) ((T POINTS AT KALLE))
77. Kalle  (wood) (.) maybe?
78.        (3.9)
79. Simo   (to buy) (0.7) very very soft ( ) ((KNOCKS ON THE DESK))
80. T      okay s- (1.2) no::?= ((LOOKS AT CO-TEACHER))
81. CoT    =okay ti:me
82.        (2.2) ((CO-TEACHER'S TIMER BEEPS))
83. T      I'm gonna follow you to the next table and >we('ll) do
           it< o↑KAY ((TEACHER ANNOUNCES CHANGE OF DESKS))
```

A 'why' question such as that in Line 60 makes relevant a subject-specific explanation for the teacher's action. From the perspective of assessment, the function of this would be to find out whether the learner is able to extend the understanding of how windows can save energy to how it is achieved in other parts of the house, such as the ceiling/roof. Simo's answer to Line 60 is delayed, and the teacher's follow-up question in Line 62 is hearable as a hint for finding the answer to the 'why' question; at the same time, Line 60 also offers a resource for answering the 'where' question in Line 62. While these relevancies are visible in Simo's answer in the sense that 'solar panels' (Line 64) are indeed in the pointed direction, the discovery that he makes is not the one that is being elicited from him in this situation. It fails to perceive that the teacher is inviting an explanation (for *ceiling*) that is similar to what was 'happening' with triple-glazed windows, something that links these two structural elements and energy saving.[1]

After rejecting Simo's answer, the teacher offers a new hint in Lines 66–67, directing him to find the relevant response in energy that is lost by heat transfer as opposed to energy that one can gain from solar panels. At the same time, it reinforces a more scientific conceptualisation of insulation (compare: "keeps out the cold" in Excerpt 7 with "stops the heat getting out"). The teacher affirms and builds on Simo's answer (Line 68) by asking another leading yes/no question, in which the word 'just' makes quite clear that 'no' is the invited answer (Lines 70–71): as such, it serves as explicit mediation of Simo's thinking, as the teacher's formulation does not leave room for any other response from the learner. Furthermore, now that a negative response has been produced by Simo, an opportunity to ask further guiding questions is created for the teacher.

[1] It is worth pointing out that in everyday Finnish language usage, one word ('katto') can refer to both the roof and the ceiling. Given this, Simo is also facing the (linguistic) task of distinguishing between these two structures as he is being instructed by the teacher, in addition to the subject-content task that deals with insulation. This may also explain, at least partly, why he responds to a question about the ceiling by mentioning solar panels.

The teacher affirms Simo's answer by shaking his head and repeating the answer. His follow-up question (Lines 74–75) builds on Simo's response (by way of the turn-initial *so*) and makes relevant an explanation of what needs to be 'done'. Kalle, another student in the group who bids for (Line 75) and is given a turn (Line 76) does not provide such an explanation but instead suggests in a hedged manner that a different material could be used (Line 77). The teacher does not seem to respond to Kalle in any way (although his facial expressions are not entirely visible on the camera at this moment) during the ensuing silence. Simo's subsequent turn (Line 79) is more sensitive towards the teacher's question in that it appears to describe what one needs to do (*buy*). Parts of the turn are inaudible, but turn-design features such as the intra-turn silence, description of an object, and knocking on the desk suggest that a word search is underway. "Very very soft" (material) could refer to thermal insulation wool, which would be a content-wise relevant answer in this context. That is, it could be that Simo's problem is a linguistic rather than a conceptual one at this point. Perhaps due to the somewhat ambiguous formulation, the teacher does not treat it as a (fully) correct answer but as one that still needs work (Line 80). However, at the same time, the co-teacher announces a change of desks, and the focal teacher tells Simo that he will resume the discussion in the next stage (though he does not do it during the lesson). Above all, such a closing remark conveys that Simo has not yet displayed sufficient understanding of insulation, a topic that as such has not been named but only alluded to during the exchange.

Unfortunately, the following lessons were not recorded. Thus, we cannot trace whether and to what extent there was transcendence in terms of Simo's internalisation of the knowledge/understanding that emerged in the interaction and was co-constructed together with the teacher. However, what we are able to trace in this interactional episode is the practical work that a teacher does to gain insights into the learner's abilities and to simultaneously direct the learner's understanding through the use of guiding questions and other mediational means, such as gestures. To summarise, by formulating his questions in different ways, the teacher is able to both open up opportunities for Simo to display and elaborate on his understanding (e.g. as in Line 54) and limit this range by way of yes/no questions (e.g. Lines 71–72). We still wish to underscore that it was not the grammatical structure of the questions per se that mattered but how these questions manifested an attempt to calibrate assistance on the basis of the learner's responsiveness to previous assistance and in turn, to direct the learner's responsiveness and emerging understanding.

From the point of view of simultaneously teaching and assessing, it transpires that Simo was still other-regulated regarding his understanding of the concept of insulation, meaning that even though learners wrote down the key information from the unit onto the poster, he still struggled to understand the material without external assistance. We also learn that with enough guidance, Simo was able to understand this concept (even if the word itself did not emerge once in the interaction)—insulation is there so that heat, that is energy, is not lost. However, Simo was probably not yet able to transfer this knowledge to other contexts.

We argue that it is namely this assistance-reciprocity cycle that yielded insights into the learner's abilities, helping the teacher to build his understanding of what Simo knew about insulation and promoting Simo's understanding of the concept of insulation. In the discussion, we will return to these two functions of assessment as a part of classroom interaction: assessment and instruction, connecting them to the insights that we gained from teacher interviews.

9.4 Discussion

In the present chapter, we aimed at gaining insights into how CLIL teachers understand assessment and how it is involved in their practices. Our major goal was to raise awareness about assessment as a dialogical process in which the teacher's and the learner's understandings are jointly constructed.

Gaining teachers' perspectives is an important first step in developing their assessment practices (Livingston and Hutchinson 2017). Overall, several observations can be made based on our analysis of the interview data. Bearing in mind that due to the nature of the data, these cannot be generalised, we summarise these in following:

- CLIL teachers may associate assessment with assessing learning outcomes more readily than with assessing their learners as a part of their classroom interaction (Tsagari and Vogt 2017; Chap. 1, this volume);
- when assessing content-related knowledge in assessment *of* learning, CLIL teachers have ways to minimise the role that learners' language skills/competence plays in the assessment; this can happen by allowing the use of learners' L1 or asking learners to explain concepts and definitions in everyday language (as transpired in the interviews we analysed); or by scaffolding the language in assessment activities (Lin 2016);
- teachers' concern for learners showing understanding in assessment of learning outcomes often emerged in the interviews;
- with regards to assessment other than that of the product, teachers tend to refer to it as understanding of their learners' abilities;
- some teachers' reports on their instructional practices imply that they have implicit understanding of what learners can do, what they find difficult, and how to make sure that learners understand; other teachers can find it problematic to find whether the source for learner problems is language or content and how to address these problems.

All in all, these observations led us to study how teachers' understanding of their learners and their learners' understanding can develop during classroom assessment when assessment opportunities in classroom interaction (Hill 2017) are recognised and utilised by the teacher.

The interaction that we analysed exemplified how instructional decisions can be made during the assessment of the process and immediately after it. Namely, as the result of the assessment, the teacher decided to follow the learner in order to guide

him further. It is interactions like this that can give teachers insights into what presents a difficulty to learners and thereby inform teacher practices, e.g. opening up complex definitions for learners, such as those described by T3 and T4. They can provide a foundation for continuing instruction with the same learner at some later time (with reference to the notion of transcendence) as well as for teaching the whole cohort. In the case of the activity during which the teacher-learner interaction we studied took place, this could involve checking learners' understanding of the concept of insulation while they present their posters, for example, asking guiding questions as those in the classroom interaction we analysed.

Both the interview and the classroom interaction data show that teachers strive to find out if there is understanding behind learner responses in assessment, be it a test, an essay, or an interactional turn. In T2's interview, it transpired that she pushed learners to express their understanding using everyday language even though, as is seen in the classroom interaction episode, learners can be directed towards using scientific language. However, it is then not surprising that learners use this language in their writing and subject-specific discussions. Furthermore, if learners are allowed to use their L1 or everyday language, do they always possess the required resources to do so? As Dalton-Puffer (2016) argues, teachers should make content, including scientific vocabulary, accessible to learners. Indeed, as far as T3 and T4 practices are concerned, it appears that some teachers have procedures for making sure that learners understand the content. The classroom interaction data that we analysed is then an example for how practices such as using everyday language to explain complex concepts (T4) can happen procedurally. In fact, we cautiously propose that teachers' use of the knowledge about what learners understand accumulated through A*f*L can help them recognise if there is understanding behind scientific definitions and explanations of phenomena in learners' essays or tests. This speaks in favour of using Leung and Morton's (2016) integration matrix as a way to inform classroom-based assessment cycle (see Chap. 1, this volume), that is, to understand the types of knowledge elicited by different assessment activities and ways that the following teaching, learning, and assessment can be informed based on the insights emerging in assessment.

We emphasise that in this process, both assessment and instruction are important, as it is when both are analysed together that one can gain insights into what the learner knows and understands and how to build upon this understanding. We certainly admit it that as teachers interact with their learners, they may not have the luxury of analysing the data as we did, having access to the video and the transcript and being able to trace in minute detail how the interaction unfolded. Despite this, the interactional episode shows the teacher being sensitive to Simo's emerging understanding and continuously assessing his performance as the interaction unfolds. Neither is it always possible to follow one learner for an extended amount of time in the classroom, as was the case in our data, too. Acknowledging all this, we nevertheless argue that being aware of assessment opportunities and how these assessments could be arranged can be the first step for teachers to develop both assessment practices that work for *them* and understanding what their learners can do, sources for their problems, and how to address these problems, be these language-related, content-related, or both.

We also want to underscore that language and content *were* integrated in the interactional episode that we studied, even if the integration was not planned and did not involve using pre-designed procedures focusing on language (explicit language pedagogy) or content (content pedagogy). Instead, the nature of integration was more situated and 'centrifugal' (see Leung and Morton 2016; see Chap. 1, this volume) and emerged as the teacher and the learner developed their understanding through the teacher's continuous assessment of what the learner could and could not do and the subsequent guidance. This guidance included questioning, gestures, and confirmations from the teacher leading the learner to demonstrate his knowledge through language including shifting to using scientific rather than everyday language. In a way, the integration of content and language as we propose it here is contrary to how T1 (Excerpt 5) attempted to address merging language and content (going first to the linguistic part and then, the content part). Rather, in this perspective on integration, the purpose of assessment is not to focus on either the language or content separately, but to dynamically build learner understanding in interaction. We propose such assessments can be organised with reference to the notion of reciprocity (Feuerstein et al. 2010), meaning that the following assistance should build on the way that the learner reacts to the preceding assistance.

We would like to draw the reader's attention to the fact that the teacher in the interactional episode we analysed did not focus straight-away on learning, but aimed at promoting Simo's understanding. Certainly, the ultimate aim of any instruction is to promote learning. However, in order to assess whether Simo learned from the interaction, a further assessment would be required, perhaps, during the following lesson, informed by the other concept we mentioned at the beginning of the chapter—transcendence. In other words, the teacher would need to assess whether and to what extent Simo could apply the understanding that emerged in the previous interaction in a different context, similarly to how the teacher attempted to guide Simo in applying the understanding of the function of triple-glazed windows to saving energy. In order to maintain a continuous assessment cycle, the teacher would have to build his subsequent teaching and assessment on what he learned about Simo's understanding of insulation in these extracts. We note that the teacher at least promised that this would be the case. We also note that the teachers in our interview data, too, talked about understanding their learners and ways of ensuring learners' understanding, which is not surprising, as understanding is an important part of education and, therefore, is also the object of a body of research (see, e.g. Lindwall and Lymer 2011, for a discussion).

Considering the value of raising teacher awareness in developing teacher assessment literacy and bearing in mind all of the above, we propose that the purpose of assessment as described in this chapter is promoting learners' and teachers' understanding and by doing so effectively introduce a term—*assessment for understanding*. We argue that the idea of *assessment for understanding* can be useful, at least in teacher training, as teachers are informed of what the gains of such assessment are and then, hopefully make it a part of their practice to serve these two purposes: to understand their learners and to promote their learners' understanding. Above all, 'understanding' in these two meanings emerged in the teacher interviews we

analysed when teachers talked about assessment, explicitly or indirectly. Furthermore, understanding aligns with the concept of assessment *for* learning and the classroom assessment cycle discussed in Chap. 1 of the volume. That is, in order to give learners feedback that tells them where they are in relation to the goals, where they are likely to move next in their development, and how to get there, teachers should first themselves understand where the learners are and how to push their development. For CLIL, both content and language goals are needed to be considered together and in relation to one another, though the focus on one or the other can vary (see Chap. 1, this volume). With regard to promoting learner understanding, the present chapter, informed by sociocultural theory, discussed one way this can happen in practice, building the argument for teachers not only using assessment opportunities as they arise in classroom interaction but consciously seeing them as such, which allows for building on them in subsequent teaching and assessment (see the classroom-based assessment cycle in Chap. 1). The classroom data illustrated that understanding in the classroom data is a social and interactional phenomenon that teachers make visible through their practice. In a way, this is similar to our interview data, where we did not observe learner understanding directly but through teachers making it social by talking about it.

Having said that, we underscore that, at present, *assessment for understanding* is a practical rather than theoretical concept. Our goal is, to repeat, to raise teachers' awareness about what and how to assess. As teachers talk about understanding, the discussion and changing teacher assessment practices should build on it, revolving around how they assess (*for*) understanding. To sum up our argument, learning to understand and learning with understanding is "linking ideas one to another in a rich intricate web; applying what we learn to answer new questions; reflecting on knowledge; and expressing ideas in creative ways. With understanding, learning becomes personal." (NCISLA 2005, qtd. in Wilson and Peterson 2006, p. 9). By all means, though, more conceptual and empirical work is required for making the term usable, such as establishing above all, the relationship between *understanding* and *learning*. We tentatively propose that this relation is reciprocal, as one needs to understand to learn, which opens up possibilities for further development of understanding. Furthermore, a conceptual relationship between *assessment for understanding* and *assessment for learning* should then be established. Only then can the term be used to its full potential, be it in (CLIL) teacher education or elsewhere. For example, as we suggested, understanding can be presented as the purpose of one assessment event (such as the classroom interaction episode in this chapter) in the classroom assessment cycle. These events together can be conceptualised as a contingent continuous process having the purpose of promoting learning, assessment *for* understanding being thus a part of assessment *for* learning.

We admit that further conceptualisation of the term is required. We also acknowledge that the introduction of yet another term into a field that has seen a proliferation or terms (see Chap. 1, this volume) can potentially create confusion rather than resolve it. We, however, hope we have sketched some of the directions for the development of conceptualising assessment *for* learning with a practical goal of raising teacher awareness of assessment and its role in CLIL classrooms.

Appendix A: Transcription Symbols

wo::rd prolonged sound
(.) silence less than 0.2 seconds
(2.0) duration of a silence
(word) uncertain transcription
( ) unintelligible talk
wo- cut-off
[ ] overlapping talk
((POINTS)) embodied action
<word> slower pace than in surrounding talk
>word< faster pace than in surrounding talk
word emphasised talk
.hh an audible inbreath
hh outbreath
= latched utterances
°word° quieter than surrounding talk
, continuing intonation
? rising intonation at the end of a prosodic entity
↑ ↓ change in pitch height
italics English translation of a Finnish turn constructional unit

Further Reading

Skinnari, K., & Bovellan, E. (2016). CLIL teachers' beliefs about integration and about their professional roles: Perspectives from a European context. In *Conceptualising integration in CLIL and multilingual education* (pp. 145–167). Bristol: Multilingual Matters.

Tan, O. S. (2003). Mediated learning and pedagogy: Applications of Feuerstein's theory in twenty-first century education. REACT, 22(1), 53–63. https://repository.nie.edu.sg/bitstream/10497/3866/1/REACT-2003-1-53.pdf. Accessed 17 Oct 2018.

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102.

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualizing integration in CLIL in multilingual education* (pp. 29–54). Bristol: Multilingual Matters.

Davison, C. (2008, March). *Assessment for learning: Building inquiry-oriented assessment communities*. Paper presented at 42nd Annual TESOL Convention and Exhibit. New York.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.

Feuerstein, R., Feuerstein, R. S., & Falik, L. H. (2010). *Beyond smarter: Mediated learning and the brain's capacity for change*. New York: Teachers College, Columbia University.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly, 9*(2), 113–132. https://doi.org/10.1080/15434303.2011.642041.

Gardner, R. (2004). Conversation analysis. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 262–284). Oxford: Blackwell. https://doi.org/10.1002/9780470757000.ch10.

Hauser, E. (2011). On the danger of exogenous theory in CA-for-SLA: A response to Hellermann and Cole (2009). *Applied Linguistics, 32*(3), 348–352. https://doi.org/10.1093/applin/amr015.

Hill, K. (2017). Understanding classroom-based assessment practices: A precondition for teacher assessment literacy. *Papers in Language Testing and Assessment*, *6*(1), 1–17. http://www.altaanz.org/uploads/5/9/0/8/5908292/3.si1hill_final_formatted_proofed.pdf. Accessed 26 Apr 2018.

Jakonen, T. (2018). Retrospective orientation to learning activities and achievements as a resource in classroom interaction. *The Modern Language Journal, 102*(4), 758–774.

Kasper, G. (2009). Locating cognition in second language interaction and learning: Inside the skull or in public view? *IRAL – International Review of Applied Linguistics in Language Teaching, 47*(1), 11–36. https://doi.org/10.1515/iral.2009.002.

Koshik, I. (2002). Designedly incomplete utterances: A pedagogical practice for eliciting knowledge displays in error correction sequences. *Research on Language & Social Interaction, 35*(3), 277–309. https://doi.org/10.1207/S15327973RLSI3503_2.

Koole, T. (2010). Displays of epistemic access: Student responses to teacher explanations. *Research on Language and Social Interaction, 43*(2), 183–209. https://doi.org/10.1080/08351811003737846.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education*. Bristol: Multilingual Matters.

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. New York: Guilford.

Lin, A. M. Y. (2016). *Language across the curriculum & CLIL in English as an additional language (EAL) contexts: Theory and practice*. Springer.

Lindwall, O., & Lymer, G. (2011). Uses of "understand" in science education. *Journal of Pragmatics, 43*(2), 452–474. https://doi.org/10.1016/j.pragma.2010.08.021.

Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice, 24*(2), 290–307. https://doi.org/10.1080/0969594X.2016.1223016.

Maynard, D. W. (2006). Cognition on the ground. *Discourse Studies, 8*(1), 105–115. https://doi.org/10.1177/1461445606059560.

McHoul, A. (1990). The organization of repair in classrooms. *Language in Society, 19*(3), 349–377.

Mehan, H. (1979). "What time is it, Denise?": Asking known information questions in classroom discourse. *Theory Into Practice, 18*(4), 285–294. https://doi.org/10.1080/00405847909542846.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.

Pascual Peña, I. (2017). *Assessment for learning in primary CLIL Classrooms and its co-construction in classroom discourse*. Doctoral dissertation, Universidad Autónoma de Madrid. https://repositorio.uam.es/bitstream/handle/10486/680101/pascual_penna_irene.pdf?sequence=1. Accessed 20 Mar 2018.

Pomerantz, A. (1986). Extreme case formulations: A new way of legitimating claims. *Human Studies, 9*(2), 219–230.

Raymond, G. (2003). Grammar and social organization: Yes/no interrogatives and the structure of responding. *American Sociological Review, 68*(6), 939–967.

Rusk, F., Sahlström, F., & Pörn, M. (2017). Initiating and carrying out L2 instruction by asking known-answer questions: Incongruent interrogative practices in bi- and multilingual peer interaction. *Linguistics and Education, 38*, 55–67. https://doi.org/10.1016/j.linged.2017.02.004.

Scarino, A. (2017). Developing assessment literacy of teachers of languages: A conceptual and interpretive challenge. Papers in Language Testing and Assessment, 6(1), 18–40. http://www.altaanz.org/uploads/5/9/0/8/5908292/4.si2scarino_final_formatted_proofed.pdf. Accessed 12 Mar 2018.

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. Papers in Language Testing and Assessment, 6(1), 41–63. http://www.altaanz.org/uploads/5/9/0/8/5908292/5.si3tsagarivogt_final_formatted_proofed.pdf. Accessed 27 Apr 2018.

Van Compernolle, R. A. (2013). Interactional competence and the dynamic assessment of L2 pragmatic abilities. In *Assessing second language pragmatics* (pp. 327–353). London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137003522_13.

Van Compernolle, R. A. (2017). CA-for-SCT: Dialectics and the analysis of cognition on the ground. *Language and Sociocultural Theory, 3*(2), 173–193. https://doi.org/10.1558/lst.v3i2.33171.

Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *The Modern Language Journal, 92*(4), 577–594. https://doi.org/10.1111/j.1540-4781.2008.00788.x.

Waring, H. Z. (2015). Promoting self-discovery in the language classroom. *IRAL – International Review of Applied Linguistics in Language Teaching, 53*(1), 61–85. https://doi.org/10.1515/iral-2015-0003.

Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL*. Doctoral dissertation, University of Turku. http://urn.fi/URN:ISBN:978-951-29-5754-5. Accessed 12 May 2018.

Wilson, S. M. & Peterson, P. L. (2006). *Theories of learning and teaching What Do They Mean for Educators*? National Education Association working paper.

Wyner, L. (2014). Learning-oriented assessment: An introduction. *Papers in TESOL & Applied Linguistics, 14*(2), 38–40. https://doi.org/10.7916/D8H70FDV.

Chapter 10
Teacher-Based Assessment of Learner-led Interactions in CLIL: The Power of Cognitive Discourse Functions

Mark deBoer

10.1 Introduction

Entering the year 2000, Japan began to have a less global economic presence for a number of reasons. One in particular was that less Japanese university graduates were going overseas to engage in business activities. In 2010, the Japanese Ministry of Economy, Trade, and Industry (METI) released a report outlining these issues, also proposing a solution to this problem. Their solution, by putting the onus on the higher education system, was to create "Global Human Resources" (METI 2010, p. 8) with (a) *communication skills* (preferably in English as a foreign language [L2]), (b) *ability to work in teams*, (c) *planning skills*, (d) *thinking skills*, and (e) the *ability to take action*.

Tertiary-level English language education, through the use of daily conversation textbooks (e.g. Richards 2012), does not benefit a Japanese chemical engineer's future of collaborating with foreign researchers or companies. It was recognised by the author that a syllabus could be conceptualised under the umbrella of language integrated with content to develop learners' abilities to discuss scientific concepts in the English language as a starting point. Yet, language teaching faculty are not necessarily capable of teaching content such as chemistry, nor can it be assumed that content faculty have the abilities to teach their field of expertise in the English language (see Chap. 1, this volume). To meet METI mandates, conceptually, the integration of language and content would need to be a collaborative effort. First, the English language learning classroom would be the venue for the integration of language and content. Second, emphasis would be on action and interaction of the learners for an agency-based approach to their education (van Lier 2008). Third, infringing on content teachers' beliefs of content and language integration (Skinnari

M. deBoer (✉)
English for Academic Purposes, Akita International University, Akita, Japan
e-mail: markdb@aiu.ac.jp

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_10

and Bovellan 2016) by inviting content faculty to support learners' language learning process. Taking these all into account, a syllabus was designed and implemented, integrating content with language in the language classroom through assignments that were designed to foster learner agency.

In this chapter, I explore the learner-learner interaction in a course using this syllabus through the lens of assessment practices, focusing on the mediation that occurs between the learners, with the goal to suggest what the teacher, through observation of learner interaction, can learn about their learner abilities. I will then discuss how this can inform the assessment cycle (Davison 2008; see Chap. 1, this volume), drawing on what I as the teacher of these learners learned from this experience.

When the learners were not co-located, they used an online asynchronous forum to collaborate. I will specifically examine their interaction in this environment to determine how they communicated while co-constructing a poster, basing my examination on the notions of language *of*, *through*, and *for* learning (see e.g. Coyle et al. 2010). This poster, henceforth improvable object (IO) (Wells 1999; Bereiter and Scardamalia 1996), was co-constructed in the online forum and shared as an attached file.

How the learners moved through the process will be examined through the language they used and how they mediated their learning as they assessed their IO at each shared iteration. In addition, I will discuss how the online forum can be useful for educators in assessment *for* learning practices (see Fig. 1.1 in Chap. 1, this volume), in particular unplanned assessment of learners during their face-to-face time in the classroom environment.

10.1.1 Designing the Syllabus

A traditional syllabus in L2 educational context in Japan is one that focuses on teaching the grammatical structure of the English language and summative testing being the primary form of assessment (Green 2016). This performance-based approach to learning (Bernstein 2000) is one with pre-programmed knowledge or "inventory of standards" (Leung and Morton 2016, p. 236) in which there is a strong classification and framing (see Chap. 1, this volume).

However, to develop learner abilities to reach the objectives the METI has mandated, a competence-based approach to learning may be a better approach for a number of reasons. First, being able to communicate in a foreign language, preferably English, would require learners to learn to output the language, not through pre-scripted conversations, but through verbal actions about content; cognitive discourse functions (CDFs; see Chap. 1, this volume) to "let others know which cognitive steps they are taking in handling subject content" (Dalton-Puffer 2016, p. 32). Learners have 6 years of English at the secondary level of education, so they do not come to university as empty vessels (Engeström and Sannino 2012). Thus, building

on learners' natural proclivities would reinforce their learning through assignments that elicit centrifugal interaction (Bakhtin 1981).

The remaining METI mandates; *ability to work in teams*, *planning skills*, *thinking skills*, and the *ability to take action*, fall under the umbrella of 'fundamental competencies for working persons' (2010, p.7) in the METI report. To foster those types of abilities in a classroom setting requires to "emphasise participation in rich contexts of cognitively engaging content learning" (Leung and Morton 2016, p. 237) and place more emphasis on individual choice and agency. Group work, therefore, based around assignments that encourage learners to determine their own content, could foster discussions, planning, taking initiative, and taking action to reach objectives. This emphasises a competence-based approach and elicits centrifugal interaction through the integration of content into assignments. In this chapter, I focus on one particular assignment from the syllabus developed using an information and communications technology programme integrated into the learning management system (LMS) Moodle (Dougiamas 2011). The syllabus was informed by the guidelines outlined by the METI and informed by the CEFR benchmarks (see, e.g. deBoer 2017; O'Dwyer and deBoer 2015). In the assignment, learners were asked to research local environmental issues in groups over 15 weeks. They presented a poster in the target language at the midterm, which was attended by peers and invited faculty, who would discuss the poster and provide feedback. Using that experience and feedback, the learners made changes and further developed their research for the remainder of the term, at the end of which, they had a group PowerPoint presentation. Each of the environmental issues was provided by the teacher.

10.1.2 Learning Environments

Two learning environments provided the learners with the means to interact: one was the online forum in the LMS, and the other was the face-to-face classroom once a week for 90 minutes.

The asynchronous online forums in the LMS were available for the learners to collaborate when they were not co-located. Each group had their own forum that could be used to send messages and/or upload files. The teacher had access to all group forums. It is the language that occurred in the online forums that is particularly of interest as it contained a complete record of the dialogue and shared files between the learners. The online forum interaction provided details of the process to help the teacher understand how the learners were able to create the final product, much more than what might be learned by the teacher only observing the groups by walking around the classroom.

Saying that, though, the classroom time was used for learners to work face-to-face on their group projects and for them to explore what other groups were doing by talking with them face-to-face. The teacher was also available for discussion if needed, but it is the observation of the online forums that provided the teacher with

assessment opportunities during this face-to-face time. I discuss in this chapter the conceptual understanding of the teacher utilising both learning environments to understand and guide the learning process.

10.1.3 Online Forum Interaction

The asynchronous forums can be used to communicate in two ways. Learners can write in the message area of the forum and send that to the others in their group and they can also send files attached to their message. The attached file can be multimodal, i.e. can contain images as well as text, and can be shared and edited by learners through many iterations; an IO (Scardamalia et al. 1994) as the focus of learner collaboration and 'the transformation of that object by means of those actions' (Wells 2000, p. 67). The IO is co-constructed through a process, and it is defined as an object that "can be reviewed, rethought and revised … and engaged with dialogically" (Wells 1999, p. 115). Conceptually, the CLIL vehicular language used by the learners in the message area and the language embedded in the IO can be different from three interrelated perspectives (Coyle et al. 2010); the language of learning, language for learning, and language through learning (p. 36). The language *of* learning is the language that is needed 'to access new knowledge and understanding when dealing with content' (Coyle et al. 2010, p. 61), including new vocabulary and "the language of describing, defining, explaining, or hypothesising" (p. 61). Language *for* learning is the "language needed by learners to operate in a learning environment where the medium is not their first language" (Coyle et al. 2010 p. 62). This includes language to build arguments, answering and asking questions, and language for project work (Coyle et al. 2010). The language *through* learning, is language that emerges as a result of the development of new knowledge, skills, and understanding. This language is unplanned and is language that teachers "learn how to capitalise on, recycle and extend [new language] so that it becomes embedded in the learners' repertoire" (Coyle et al. 2010, p. 63). This includes using feedback, recycling discussion skills, presenting evidence, and developing dictionary skills. How learners enact the content or knowledge in this competence-oriented assignment can be identified through the function of the language, i.e. cognitive discourse functions (CDFs) (Dalton-Puffer 2016).

The poster itself is important as a multimodal resource for constructing meaning in the CLIL classroom (for discussion of other such resources, see Evnitskaya and Morton 2011; Kupetz 2011; Nikula et al. 2013). The use of technology to co-locate learners beyond the spatial and temporal boundaries of the classroom is essential for creating a context for multimodality constructed meanings (Cope and Kalantzis 2017). Learners have access to a variety of multimodal sources (e.g. images, videos, and text) and bring these into the interaction through the use of the online forum or integrated into the IO. The role of multimodality should be, therefore, carefully considered in learners' construction of meaning. In this chapter, mediated action (Wertsch 1994), will be used to understand how the learners' forum contributions

were mediated by symbolic or physical tools, in particular when individuals determine how to use the information shared in the online forum (or IO) to mediate their actions to reach their objective.

10.1.4 Assessment Promoting Learning and Learner-Learner Interaction

What occurred in the online forum was based on the learners' understanding of how they reach their objective. Similar to Chap. 9 of this volume, there is no specific set of assessment strategies that can be offered in this online interaction where the assessment is entirely unplanned. Here, I will study the centrifugal interaction, not only from the teacher's perspective, but also from the learners'; after all, it is primarily the learners who are relying on their collective interaction in the online forum to reach their objective. There are a number of salient aspects that the online forums afford for both the learners and the teacher for assessment purposes. Each time a learner posts a message, content, and/or uploads an interim version of the poster, other learners in the group can view the messages and attached files. Learners used language to direct the process forward, as they suggested content, edited their work, and suggested edits.

Informed by the sociocultural theory (SCT) (see Chap. 9, this volume), learner interaction can be viewed as a mediated process, through *intentionality and reciprocity* and *mediation of meaning* (e.g. Feuerstein et al. 2010). The notion of *intentionality* in mediation is the attempt to guide the performance of the learner, or "deliberate efforts to mediate the world, an object in it, or an activity for another student" (Poehner and Lantolf 2005, p. 241). *Reciprocity* is the learner's response to that intentionality in such ways as "responding to task, negotiating mediation, use of mediator as a resource, creating opportunities to develop, seeking mediator approval, and rejecting mediation" (Poehner 2008, p. 42). In *mediation of meaning,* learners select specific information that is relevant and meaningful to the group. This information has no meaning to the members of the group unless it bears meaning to the mediator and provides a background "against which categorization becomes possible" (Lidz 1991, p. 76). During this learner-learner interaction, learners "complete tasks that would otherwise be beyond their level of ability" (Lantolf and Poehner 2014, p. 163).

Learners do assess the interim IO based on their understanding of the current iteration of the IO relative to their understanding of the content to determine how it can be improved further. To reach their objective, i.e. the completion of the IO, learners provide feedback to each other in the online forum through comments, or through direct edits to the IO which helps shape further iterations and fuels further dialogue. Through repeated efforts, this allows the learners to drive the process forward in such a way that the poster eventually forms a finished product that is agreed upon by all members of the group (see Bereiter 1994 for information on progressive dialogue).

The teacher can also observe the content and the language of each member of the group in the forum. The poster presentation is a planned assessment, but observations of the interaction in the forums although unplanned, can be used to collect information about students' learning through the process. There are a number of sources of information that can be used from the forum: (1) the poster iterations, which includes content and the language of that content, (2) the online forum dialogue from each learner, (3) the responses to that dialogue from other learners, and (4) how the content changes or was changed as a result of that dialogue. The teacher can examine learner performance and also the performance of the group working collaboratively. The teacher, from these observations, can make professional judgements (Davison 2008; see Fig. 1.1 in Chap. 1, this volume) about the content and the interactions between the learners. These judgements can then be used to provide feedback or advice as well as to inform the subsequent assessment cycles.

10.2 Research Questions, Data, and Method

10.2.1 Research Questions

In this chapter, I will address the following questions:

1. How do learners use language and content to mediate their interaction?
2. What insights into learner abilities emerge from their interaction?

To answer these questions, I examined learner-learner interaction in the online forums to determine how that shaped the development of the content. By examining the language that was used by the learners to manage the process, i.e. the language *for* learning, and by examining the language that emerged in the content (language *of* and *through* learning), I could determine how the learners used these to mediate their interaction. The interaction indicates their ability to make meaning and their understanding of the content through their co-construction of a coherent poster. This knowledge is then used to inform how assessment *for* learning practices can be implemented into this kind of assignment, as the teacher reflecting back on unplanned classroom-based assessment opportunities that were afforded to him.

10.2.2 Participants and Data

The university in Japan where the study took place has four faculties, Engineering, Agriculture, Humanities, and Education. Language courses are divided based on student scores of the TOEFL (Test of English as a Foreign Language) ITP (Institutional Testing Program) (Educational Testing Service 2018). The 36 students enrolled in this general English language course were from the Engineering and

Agriculture faculties and were placed in this class as a result of their TOEFL ITP scores (average 400 or equivalent of CEFR high A2). This was considered to be an advanced language class for this university. The learners were all first-year learners, just graduated from high school.

The assignment for the learners was to create a poster presentation in groups of 3–4 centred around an environmental issue. The learners worked on this mostly outside of the class face-to-face time, but some class time was dedicated to allow learners to discuss their assignment with the teacher and other groups. During times when the learners were not co-located, they had access to an online forum for collaborative purposes. At the end of the course, the data from the online forums were collected and anonymised. The data in this chapter come from one group who researched P.E.T. bottles (single use plastic bottles); S1 and S3 were male, and S2 and S4 were female. The forum discussion thread from the P.E.T. bottle group was analysed for contingency. The series of posts were chronological in order, but they were not necessarily contingent upon each other (see e.g. Longacre 1996). For example, in one post, learners may have been discussing data they have collected, but in the next post be discussing the layout of the poster. So, while the posts unfolded chronologically, and collectively each is a step towards the learners reaching their objective, not all of the learners' discussions occurred in perfect linearity.

Each post, considered a mediated action (Wertsch 1994), was examined for its relation to the previous posts, and those posts that were contingent on previous posts were labelled based on what the learners were doing. Identically labelled posts were strung together to provide 'threads' of dialogue. In other words, the connections among the posts were studied with the goal of exploring how the learner posts were mediated with the particular focus on the academic language they used and the content they brought in their posts. Furthermore, language *of*, language *through*, and language *for* learning was identified to indicate the language learners used to manage their process versus the language that emerged as a result of their interaction. Finally, I coded the learner posts using Dalton-Puffer's classification of CDFs, i.e. classify, define, describe, evaluate, explain, explore, and report (2016 p. 33) as well as the functions within each of these categories, informed by the revised Bloom's taxonomy (Anderson et al. 2001; see Chap. 1, this volume). In other words, I focused on examining the *function* of the language the learners used and how it shaped their interaction. This conceptual basis informed my assessment of learners' abilities. One point that needs to be made here is that in Dalton-Puffer's (2016) examples, the dialogues largely included the teacher. In this chapter, I study the functions of the learners' language, the teacher not intervening into learner-learner interaction.

In this chapter, I focus on three threads of dialogue (Excerpts 1, 2, and 3), each unique in how it contributed to the learners' overall objective and was useful for the teacher to understand the process of the learners' co-construction. The text of dialogue from the learners' online forum message area has been formatted for this chapter to identify the type of text: (a) if it is content-related, it is identified through

italicised text, (b) text shown in a **bold underlined** font is academic language, and (c) all other text that is used to convey information has been left unformatted. When the learners have uploaded a file with the message, I indicated that with (attached file: file name). If any text has been added to the uploaded file, I indicated that with **{Text added to file name:** text**}**. The data examined in this chapter were collected as part of a doctoral research project (deBoer n.d.).

10.3 Results

In this section, I illustrate how learners mediated each other in the online forum through the use of language in the forum message area and through the content embedded in the IO. I also illustrate how this interaction provided assessment opportunities for the teacher observing the forum to understand and promote the learners' abilities. I investigate three excerpts that show the progression of the learners' mediating each other to reach their objective of completing the poster presentation, guiding each other to a mutually agreed direction (Bereiter 1994). They did so through the use of language that is used to manage the interaction, i.e. the language *for* learning (Coyle et al. 2010). In this competence-based approach (Quadrant 4 of Leung & Morton's integration matrix; see Chap. 1, this volume), the roles of the learners oscillated between the mediator and the mediated as they made suggestions, edited, and introduced content to push the IO to completion. In essence, it was not just one learner's performance that determined the content of the poster, but a joint construction mediated by multimodal resources (images and text) that shaped their understanding. The three extracts represent 25 posts and 13 files shared out of a total of 123 posts and 57 files shared in this group over the four-month term.

The following Excerpt 1 illustrates the initial online interaction in the forum. At the outset, S1 provides the other learners in the group with a research instrument and instructions on whom to interview. The learners were to find out about the recycling policies and use of single-use plastic P.E.T. bottles on the university campus, so this research instrument is used to determine the amount of use. S1 also gives a deadline. Interview data is collected and shared by the members of the group which prompts S1 to research additional data, possibly from the Internet. One of the learners also gathers information from the campus store.

Excerpt 1: Collecting, Sharing, and Analysing Data

1-1. S1: Please interview the students of the subject same as ourself about next questions by next Tuesday. *1.How many pet bottles do you use in a week? 2.Do you have your own my bottle?*

1-2. S1: I interviewed the students of the same department with me *1.How many pet bottles do you use in a week? 0→4 people, 1→15 people, 2→12 people, 3→9 people, 4→6 people, 5→11 people, 6→3 people, more than→10 people. 2.Do you have your own my bottle? Yes→13 people. No→54 people*

1-3. S4: I have 23 people's answers. *How many petbottle do you use in a week? 0→1 people 1→5 people 2→4 people 3→9 people 4→no people 5→4 people ·Do you have your own my bottle? Yes→8 people. No→15 people*

1-4. S3: I have asked 12people questionnaire. *question1 1:2people, 2:3people, 3:1person, 4:1person, 5:2people, 7:1person, 0:2people. question2 have my bottle: 4people don't have my bottle : 8people. Some people have my bottles, but don't use it.*

1-5. S1: I researched about ***annual consumption***. *The annual consumption of the 500-ml pet bottle per one Japanese amounts to 166 ones in 2006.* ***Converting into*** *per week,* ***it is equivalent to*** *3.5 ones. Probably, it is more than 166 now, since the* ***amount of consumption*** *is* ***increasing*** *every year.*

1-6. S1: About question 1. ***I calculated*** the University student's amount of pet bottle average used. *The* ***result is*** *about 3.75 per one person. The amount is almost the same as the general average 3.5.*

1-7. S1: About question 2 ***The ratios of*** *the number of people with My Bottle and the number of people without it were 1:3. I found that many people don't have their own My Bottle, or not use it.*

1-8. S4: I interviewed University cooperative. I did it without your decision, sorry. I asked: *"How many petbottles do you sell in month?" The* ***result is*** *2013. March: 3,425 petbottles. April: 8,950 petbottles. May: 9,911 petbottles (by 28th May)*

S1 is the mediator in post 1–1; his intentionality is directed at the other learners to guide their performance. In other words, to find out about the use of P.E.T. bottles on campus, S1 describes a procedure to the other learners and includes a research instrument. Post 1–2 becomes a mediational means directing the others to share their data in the same fashion as S1 has reported. In posts 1–3 and 1–4, both S4 and S3 report that they have uploaded their interview data, mediated by both S1's initial instructions in post 1–1 and his uploaded data in 1–2. Yet, S3 reports additional data, namely *'Some people have my bottles, but don't use it'*, as there may have been some discrepancy between the data and what S3 discovered.

In post 1–5, S1 studies data from an external source which he uses to mediate his understanding of the group's data. Through the use of academic language, '*annual consumption*', '*converting into*', and '*equivalent to*', S1 presents his calculation of the annual consumption converted to weekly data, mediated by question 1 of the research instrument '*in a week*?', and this calculation will be later used to make a comparison. The information from the external source indicates a trend over a longer period of time allowing S1 to estimate that '*probably, it is more than 166 now*' as '*consumption is increasing*'. Mediated by the interview data units being per week, S1 essentially presents annual consumption as weekly consumption, the interview shaping his understanding of the group's data. His next post, 1–6, is to present both sets of data (interview data and data from the external source), but the function of his language is to first *explain* how he arrived at the weekly consumption for the university students (3.75/person) and then *compare* that information with the external data (3.5/person), concluding that the data is almost the same. In post 1–7, S1, using the data from question 2 of the research instrument is able to calculate the ratio of people with their own 'my bottle' and those without. S3's data from post 1–4 becomes the mediational means for S1 to draw conclusions. S4's action in post 1–8 is mediated by the task but she approaches the task in her own unique way. She reports the sales data collected from the on-campus store. In this Excerpt 1, the learners used academic language to discuss content. Through this discussion, they gained a fuller conceptual understanding of P.E.T. bottle consumption.

The learners use this data to initiate the making of the poster, and it becomes evident that they have been influenced by the data, their language indicating their negative stance toward the use of P.E.T. bottles (Excerpt 2).

Excerpt 2: Creating the Poster

2-1. S3: Let's make poster together!!!!!!

2-2. S1: I made poster. Please check! As the deadline approaches, let's finish our poster as soon as possible!! (***attached file: Pet bottles poster 1.docx***)
{Text in Pet bottles poster 1.docx: Pet bottle is made of ***polyester*** and ***used oil*** to make it. Because it is easy to carry the pet bottle around, ***the consumptions*** of the pet bottle are ***increasing rapidly*** year by year. 80% of the mineral water and tea are sold in pet bottles now. Pet bottles ***have become essential*** for us.}

2-3. S4: Everyone good job! S2's idea is good, I think. I made scripts in the poster.(text omitted)(***attached file: Pet bottles poster3.docx***)

{Text in Pet bottles poster 3.docx: The result of interview. Our group interviewed [Name omitted] University students. 1. How many pet bottles do you use in a week? 2. Do you have "my bottle"? This graphs show the result.}

2-4. S1: Hello. I think that we should put the following contents into our poster. 1.***Explanation about*** pet bottles. 2.About the ***influence of the environment*** on pet bottles, ***and recycling***. 3.About the ***result of the interview***. 4.About ***our opinions*** and the ***solution to the problem***. What do you think about this? Please give me your opinions!

2-5. S4: I made poster about **influence of the environment** on pet bottles and recycling. But I want to make simple and easy to see. Please check and give me advise. (***attached file: Pet bottles poster 6.docx***)

{Text added to Pet bottles poster 6.docx: Pet bottle makes ***CO2***!! When pet bottles are made in factory, ***they emit a lot of*** CO2. For example: 500ml pet bottle water makes ***CO2*** 500 times as large as the same amount tap water. 2. ***Problem of recycling*** pet bottles Pet bottle ***which is made from oil*** costs 7.4 yen. Recycling pet bottle costs 27.4 yen}

The activity changes in post 2–1, when S3 invites the others to begin making the poster now that data has been collected. In the following post (2–2), S1 uploads a poster, responding to S3's invitation. The language he uses in his poster has been recycled from his post in 1–5. He is effectively using the academic language to summarise and present information about P.E.T. bottles which indicates his understanding of the issue. In post 1–5, S1 uses the language to mediate his understanding of the interview data, but in his poster, the functional use of the language differs. To begin, to define P.E.T. bottles, chemical language (polyester) is used to indicate the manufacturing material and (oil) used in the process. He identifies a cause-effect relationship between the ease of carrying around a P.E.T. bottle with the increased consumption, concluding that due to the types of beverages sold in P.E.T. bottles, that they have become "*essential for us*."

S4, in post 2–3, evaluates the previous posts (S2's post is omitted), and has also added content to the poster, mediated by the language from the research instrument in post 1–1 (although the graphs were not added). The activity changes again in post 2–4, because here S1, indicating a fuller conceptual understanding of the issue, mediates the others by suggesting the poster be structured using four distinct areas. The content about P.E.T. bottles has already been added in post 2–2; and S4, in post 2–3, has added some information about the interviews. This list becomes a

mediational means for the others in the group to use as a resource when adding content to the IO (the poster).

In post 2–5, S4, responding to the task, recycles the language verbatim from post 2–4 to indicate what she has added to the new iteration of the IO. The language that has emerged in previous posts by S1 (see posts 1–5, 2–2, and 2–5) has helped develop S4's conceptual understanding of the issue and this is reflected through her explanation of the manufacturing of P.E.T. bottles involving the use of oil which results in the emission of CO2. She uses that explanation to compare CO2 emissions and then using a cost analysis, compares recycling costs versus manufacturing costs, concluding that results in recycling problems. The learners', mediated by the content in the previous excerpt, here use academic language to define and explain their conceptual understanding of the larger issues surrounding P.E.T. bottle use.

As the learners continue to add content to the IO, they instruct each other through the forum message area what they have done and suggest changes. The learners continue to build the IO and in the following excerpt (3), they use the IO to mediate each other. This excerpt follows directly after 2–5.

Excerpt 3: Putting the Poster Together

3-1. S1: Good job, S4!I think that it is easy to see. I put the graph of the interview. Let' think about the result! (***attached file: File: Pet bottles poster 7.docx***)

3-2. S2: I think this poster is simple, but it doesn't have impact... Maybe, there are many sentences, so it is difficult to see a bit. How do you think?

3-3. S3: I suggest that we should write this *" We need to have my bottles to* ***reduce emission of CO2****"* Please give another ideas!!!

3-4. S1: I made the part of the end. Please check it ! Let's make the remaining part! (***attached file: Pet bottles poster 8.docx***)**{Text in Pet bottles poster 8.docx:** It is important that we don't use pet bottles as much as possible. You should use a canteen or my bottle!}

3-5. S4: I made graph result. please give me some advises! It doesn't easy to see , I think... (***attached file: Pet bottles poster 10.docx***)**{Text in Pet bottles poster 10.docx:** These graphs show that·50% students use 3 or more pet bottles in a week.·Many students don't have my bottle. Bad effect on environment}

3-6. S2: I changed the last of poster, *"It is important that we don't ~ possible." to "We should not ~ possible." Also, I changed "You should use a canteen or my bottle!" to "Let's use my bottles!!"* If you don't like this, please correct! (***attached file: Pet bottles poster 14.docx***)

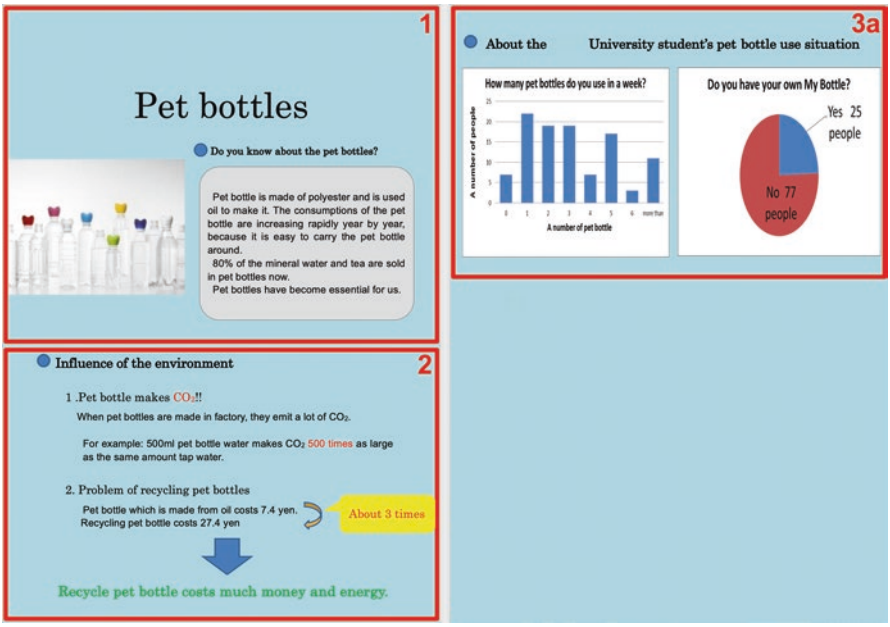


Fig. 10.1 The poster file Pet bottles poster 7.docx (see Post 3–1)

Continuing from 2–5, S1 in post 3–1 evaluates S4's post and uploads a newer version of the IO with graphs added (mediated by S4 in post 2–3, i.e. *"This graphs show the result"*) (See Fig. 10.1). The numerical data collected has been changed to a different modality, i.e. graphs. S1 invites the rest of the group to '*think about the results*', now that he has added the graphs. The focus of the following discussion is to illustrate how the learners co-construct the remaining part of the poster. For this chapter, I have divided the poster into three sections, each represented by a square and a number (1, 2, 3a), which have been populated with content, corresponding with the suggestions made by S1 in Post 2–4: ***1***. *Explanation about pet bottles.* ***2***. *About the influence of the environment on pet bottles, and recycling.* ***3***. *About the result of the interview.*

In post 3–2, S2 critically assesses the IO. Her comments are directed at the lack of impact of the poster, in particular, the language and the layout. In the following post (3–3), S3 suggests that *"we need to have* ***my bottles*** *to* ***reduce emission of CO2****"* be added. This is partially mediated by S1's post 2–4, suggesting adding opinions and a solution to the problem, but notably, S3 explores the possibility that in order to reduce the emissions of CO2, '*my bottles*' are needed. While the language has appeared prior, e.g. S4 discussed CO2 emissions in post 2–5, and the concept of 'my bottle' has been discussed since post 1–1, this is the first time that they have been brought together to allow S3 to predict what CO2 levels "would be

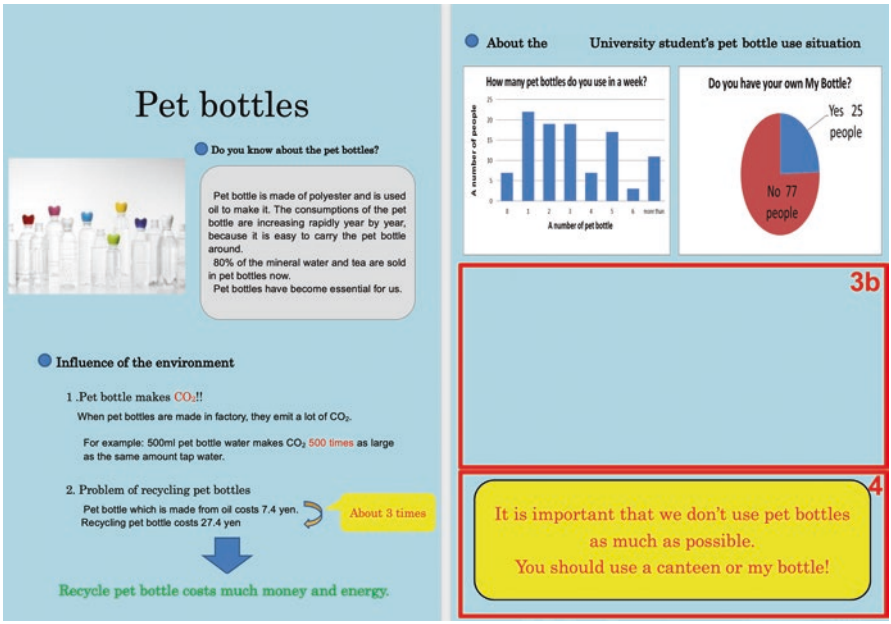


Fig. 10.2 The poster file Pet bottles poster 8.docx (see Post 3–4)

like if certain conditions are met" (Dalton-Puffer 2016, p. 47), i.e. the condition of more people using 'my bottle'. S3 does not add his suggested statement to the IO, but this becomes a mediational means for S1 in the following post (3–4) (see Fig. 10.2, Sect. 4). In post 2–4, S1 suggested *"About **our opinions** and the **solution to the problem**"*, and here in post 3–4, mediated by S3's suggested statement, adds a revised version, the first sentence presents an opinion and in the second sentence offers a solution.

It is S4 in Post 3–5 who responds to S1's invitation in post 3–1 *"Let' think about the result!"* by summarising the graphs through two statements, each of them corresponding to one of the graphs (Fig. 10.3, Pet bottles poster 10.docx). Her interpretation of the graphs shows her conceptual understanding of the problem, exemplified in her statement "*bad effect on environment*". She was able to explain and describe the graphs and draw conclusions from their meaning. She has also used the graphs as mediational means to calculate the percentage of students that use three or more pet bottles a week. The data has now had representation in three different modalities; first the initial raw numerical data, then the graphs, and now the text to describe the graphs.

In the final post (3–6), S2 makes revisions based on her previous assessment of the IO in post 3–2 (see Sect. 4 of the poster in Fig. 10.4). First, her post builds on her emerging understanding of the issue, but it also shows how she can use language

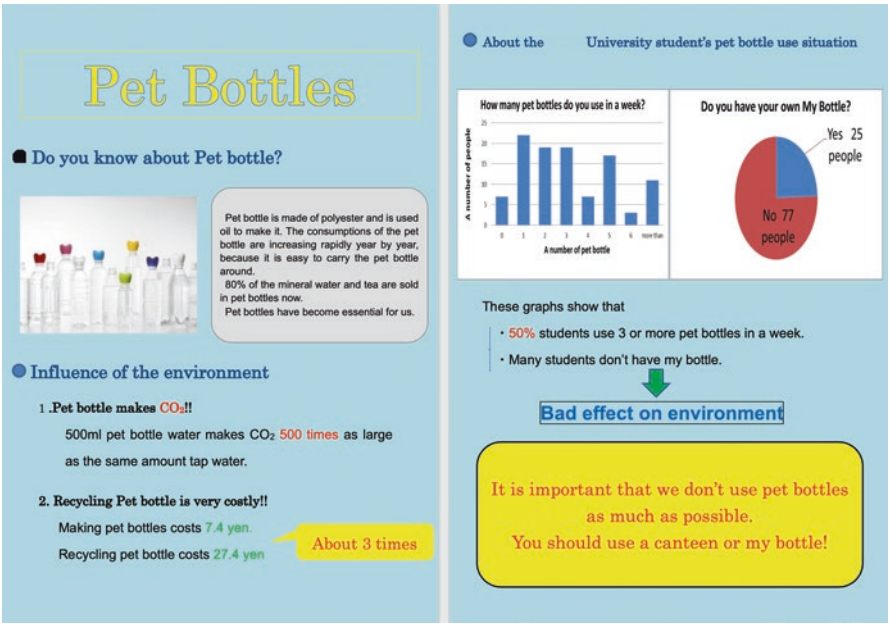


Fig. 10.3 The poster file Pet bottles poster 10.docx (see Post 3–5)

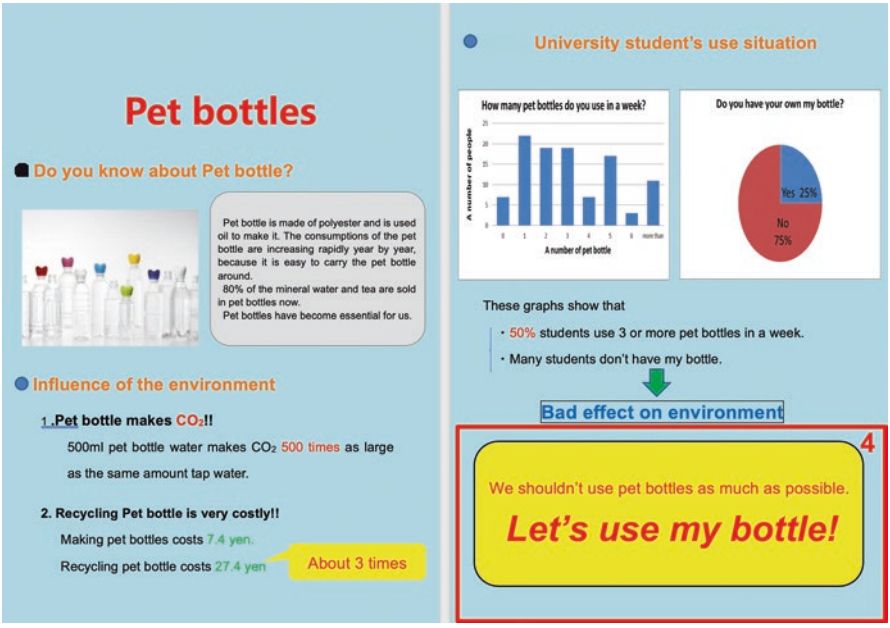


Fig. 10.4 The poster file Pet bottles poster 14.docx (see Post 3–6)

to make the message stronger; i.e. from "*it is important that*" to "*we shouldn't*", indicating a stronger stance on the issue, at the same time making it a shorter sentence. She also revises the second sentence into a friendly invitation to encourage students to use '*my bottle*' and enlarging the font to make it stand out, giving it impact. S1's post in 3–4 where he added a revised version of S3's statement has become the mediational means for S2 to indicate her emerged conceptual understanding of the issue. In this excerpt, the learners have identified a cause-and-effect relationship between P.E.T. bottle use and environmental issues and recommended a solution to counter the issue.

What has been presented here is the learners' use of language and content to mediate their interaction and the development of the poster. The intricacies of this will be discussed in the following section with regards to the function of the language. I will also discuss how the teacher can use this interaction as unplanned assessment, informing the subsequent classroom-based assessment cycles.

10.4 Discussion

In this present chapter, I aimed at showing my understanding of how learners mediate each other and how these interactions can be used by the teacher to promote learning through classroom-based assessment practices (see Fig. 1.1 in Chap. 1, this volume). This understanding emerged as I enriched my theoretical/conceptual knowledge, which enabled me to now approach these data differently from when I collected these data, especially with regards to assessing learners' joint performance by the teacher. As a teacher turned researcher, I will discuss the findings from the forum entries and speculate what I could have learned from their interaction, what unplanned assessment opportunities were available, and how feedback could have impacted their learning.

Throughout the online forum discussion there are a number of salient observations that can be made about the learner-learner interaction and how it contributed to reaching their objective. These findings can be summarised as the following, though it should be noted that due to the small scale of the research, caution should be exercised with regard to their generalisability:

- The learners moving through the assignment show how they can use academic language to develop their understanding of the content. Centrifugal tendencies in the interaction were the result of learners using language to focus on the development of the content in their IO and the teacher giving them freedom in how they approach the assignment.
- Learner agency gave the learners the ability to work on their own objectives, working through areas that they found problematic. This is consistent with weakly framed pedagogy (Leung and Morton 2016).

- The language used was multidisciplinary, i.e. mathematics and chemistry language were evident in their interaction, as in weakly classified pedagogies (Leung and Morton 2016).
- The learners indicated their understanding of the language, i.e. the function of the language (see Dalton-Puffer 2016 on CDFs) to show their communicative intentions.

The objective for me was to create a syllabus to develop abilities in learners that would fulfil the mandates presented by the METI (2010). Creating activities within the syllabus with a focus on choice, creativity, and contingency resulted in the learners using whatever means at their disposal to reach their own objective. This is consistent with Quadrant 4 in the Matrix (see Chap. 1, this volume); and there are a number of key observations that I discuss which are evident in the interaction between the learners and are important to what they contribute to the understanding of learner-learner mediation in a CLIL pedagogy of this nature. Centrifugal interaction allows for unique insights into learner abilities to emerge. These insights would have not emerged should the task have been more structured and focused.

What begins to emerge from the onset of the learner interaction in the online forum is the language *of* learning (Coyle et al. 2010), or the language of the genre, namely about P.E.T. bottles and consumption. The proficiency level of language of the learners (in this case CEFR level A2) does not appear to hinder their ability to communicate or collaborate. As seen throughout their interaction, the learners have successfully used the language for specific functions, i.e. language *for* learning to "work successfully in groups, [and] carry out their research" (Coyle et al. 2010, p. 62). The un-formatted text (See Sect. 10.2.2) in each of the posts of the excerpts (1, 2, and 3) indicates the language that is used to manage the interaction. That being said, the learners are using this language with communicative intentions, which can be classified by function type (CDFs, see Dalton-Puffer 2016). The majority of that language has been used to report or inform the others what they have done, i.e. '*I interviewed*', '*I researched*', '*I changed*', but in addition, there are other functions of the language that have emerged from the interaction. In post 3–1 for example, S1 uses the language to evaluate S4's poster '*I think that it is easy to see*', and in post 2–4, S1 suggests that the poster be divided into four separate categories of information. S2 in post 3–2 is using the language to evaluate, critiquing it and then explaining her reasons why.

Due to the centrifugal nature of the interaction, the academic language that emerges is unplanned and develops though learner interaction. An excellent example is S1 in post 1–5, where he uses the academic language he associates with P.E.T. bottles, the key concept being annual consumption. He informs the others that this is data from 2006, indicating that there is a cause/effect relationship, i.e. hypothesising that since consumption is increasing, more P.E.T. bottles are probably being used. It is in post 2–2 that language *through* learning emerges in S1's initial upload of the IO. Here S1 essentially uses the same language that mediated his understanding of the issue to now describe the P.E.T bottle situation. He is recycling his discussion skills at a higher level (Coyle et al. 2010) by exploring reasons and possibly

attempting to justify the increase in consumption of P.E.T. bottles. His communicative intention (Dalton-Puffer 2016) is a mediational means to help the others conceptualise the issue.

Similarly, in post 2–5, language *through* learning has emerged in S4's explanation of the influence on the environment. Her communicative intention was to identify the connection between the manufacturing of P.E.T. bottles with CO2 emissions, indicating a cause/effect relationship. Her example of CO2 levels in manufacturing compared with drinking regular tap water is an argument against their use. Using a cost analysis comparison, she concludes that the problem with recycling is that it is too expensive, compared to that of manufacturing. The functions of her language clearly indicate to the teacher of her understanding of the issue and how it relates to the information S1 provided.

As the process continued, language *though* learning emerged in the content the learners share which could not have been planned for, gave a clear indication of the group's conceptual understanding of the issue. Indeed, there is evidence that all types of the language of the triptych (Coyle et al. 2010, p. 36) were apparent in the interaction between the learners. Based on the competence approach to the syllabus, the learners needed to use language to mediate their understanding of the issue. The language became the mediational means for the other learners to understand the issue and be able to co-construct the IO with that understanding in mind.

As the process unfolded, the interaction that occurred in the online forum afforded the teacher with unplanned assessment opportunities (Davison 2008; see Chap. 1, this volume) and also an insight into learner abilities. The forum interaction provided examples to the teacher about the learner abilities in both the academic and everyday language. Although Dalton-Puffer (2016) argued that teachers should make content, including the scientific language, available to the learners, I suggest that in a syllabus with less visible language pedagogy and a lower disciplinary orientation to language, learners need to be able to explore the language available to them and develop it based on their needs. It is the role of the teacher then to be aware of the emerging language and to capture, recycle, and develop it strategically through classroom interaction and dialogic activity (Coyle et al. 2010; Wells 1999). As the teacher, reflecting on this data, I next suggest how assessment of process involving learner interaction can be done with reference to classroom-based assessment cycles (see Chap. 1, this volume).

As stated, there were times when the learners were given time in the classroom in a face-to-face setting to discuss their assignment and research with the teacher. To begin, the teacher could collect information about the learners' through observing the online forums (Excerpt 1, for example). Then, face-to-face, the teacher could engage with the learners to discuss the data and data comparisons, using the academic language introduced by S1, with the goal to recycle the language from the forum. During this dialogic interaction, the teacher could use the CDFs the learners had used, with the intention of reinforcing their language use and conceptual understanding (Coyle et al. 2010). This would require "students to call upon their existing knowledge, concepts" (Met 1998, p. 38) to solidify the connections between the concepts and the language. The face-to-face dialogue could build on the forum

interaction and provide information about the learners' conceptual understanding and create more opportunities for unplanned assessment. The teacher could, for example, ask the learners to explain the steps they had taken so far, using this as an educational opportunity to expand the interaction in the classroom and have learners discover what other learners had done. To summarise, the teacher's role would be to collect information about the learners through observation (Davison 2008) and engage with the learners when unplanned assessment opportunities arise, e.g. having learners expand on their explanations, or even ask other groups to identify what they had learned, in order to assess their understanding of other groups' processes.

In order to gain insight into the learner abilities to develop learners' conceptual thinking, the teacher, could again engage in dialogic interaction about S1's decision about the content for the poster (post 2–5) above what the teacher observed in their forum interaction. The teacher could assess each learner's understanding of content, and where and how they intended to find that content. This would allow for providing feedback (Davison 2008) with the intention to *promote* learning. During the assessment of the actual poster presentation, insights from both the process and the final product could inform the teacher's decisions in the overall assessment process.

To a certain extent, a large part of the teacher's role would be to collect information through observation of the online forum interaction to discuss with the learners during the poster presentation. This holds true for understanding the process through which the learners went through to arrive at the final poster file. It would also indicate to the teacher of the extent to which the learners each contributed to the co-construction of the poster. S3 for example, did not contribute very much to the overall poster, but the teacher could also judge from his contribution that he had been actively involved in observing the interaction through his addition to the dialogue. The teacher could ask S3 to explain his contribution to post 3–3 through dialogic interaction, determining his understanding of the issue, and how that relates to a solution, or even other possible solutions.

Each assignment in the syllabus I designed focused on choice, creativity, and contingency. A more structured assignment could have also been used, based on a specific scientific concept and assessed using academic proficiency benchmarks (see Chap. 2, this volume). This would, too, develop the teacher's understanding of the reasons for the learners' strengths and weaknesses and allow for feedback intended to promote learning. Understanding how learners arrived at their final product, such as what content they included or did not include and why, would also help solidify both their understanding and the teacher's understanding of the learners' abilities. However, observing the centrifugal learner interaction allowed for unique insights to emerge.

10.5 Conclusion

The initial objective was to create a syllabus that would engage the learners in ways that would foster the abilities outlined in the METI report (2010). What has emerged from this process is an understanding of how learners in a CLIL classroom supported by an online forum can mediate each other through the process of co-constructing knowledge. In previous studies (see e.g. Ohta 2001), it has been shown that learners of a second language can mediate each other to understand the structure of the language, or in L1 content classroom studies (see e.g. Lemke 1990; Wells 2001) the role of language in the classroom plays a large role in helping develop learner understanding of the content. In the CLIL classroom in this study, the language and content both played a role in the interaction, the learners used L2 to co-construct content knowledge and vice versa. What it reveals is that, even with their low level of proficiency, the learners used the language for specific functions, which enabled them to make meaning and co-construct a poster that was cross-curricular in nature. The can-do lists of the CEFR scale (Council of Europe 2018) can inform the teacher of benchmarks achieved by the learners (see Chap. 2, this volume), but *how* the language is used, i.e. its function, can bring insights into the learners' abilities to use the language above and beyond their level of proficiency.

Saying that, in using an assignment as illustrated in this chapter, there is a danger of the outcome of learner interaction not being successful. The results of this course are not generalisable, i.e. what was successful in this case may not be successful in another classroom with different learners. Assessment-wise, it would still yield insights into learners' abilities, but it might not be the pedagogical outcome the teacher is expecting. Another limitation to the study is that it lacks transcripts of the classroom interactions among learners or between the learners and the teacher. Learners had opportunities to discuss their research with each other which may have helped generate ideas and advance their thinking. Feedback loops (see Kalantzis and Cope 2012) were built into the schedule to assist learners in reaching deadlines and to provide opportunities to share and practice explaining their research, opening the classroom environment to allow for mediation to occur between groups. Mehisto and Ting's (2017, p. 224) definition of assessment, which is to "help students become knowledgeable partners in the learning process" through "rich opportunities to assess and reflect on their own work and the work of others" is applicable here. Indeed, such interactions provide a multitude of assessment *for* learning opportunities for the educator.

What has been demonstrated in this chapter is that the CDF constructs are where the "conceptual orientations of content-subjects and language education intersect" (Dalton-Puffer 2016, p. 51). In other words, in these interactions, I have identified that the CDFs are not about the language and the content separately, they are about both. The insights into learners' abilities seen through the interaction of content and language show that the development of the learners' understanding of the content came from their development of the language to discuss the content and vice versa.

As a final note, this chapter brings forth of the importance of the collaboration between the teaching and research communities. Researchers and teachers need to collaborate so that more discussions occur that assist teachers in developing their assessment practices, which will feed back to the research community (see Chap. 11, this volume). Researchers can contribute with their theoretical understanding while teachers can contribute with their teaching experience. More research is needed to explore creativity and contingency in CLIL (see Leontjev and deBoer 2020) and its contributions to assessment of the process in CLIL classrooms.

Further Reading

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics, 1*(2), 216–253. https://doi.org/10.1515/eujal-2013-0011.

Ferguson, R., Whitelock, D., & Littleton, K. (2010). Improvable objects and attached dialogue: New literacy practices employed by learners to build knowledge together in asynchronous settings. *Digital Culture & Education, 2*(1), 103–123.

Kumaravadivelu, B. (2003). *Beyond methods. Microstrategies for language teaching*. New Haven: Yale University.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Pearson.

Bakhtin, M. M. (1981). *The dialogic imagination. Four essays*. Austin: University of Texas Press.

Bereiter, C. (1994). Implications of postmodernism for science, or, science as progressive discourse. *Educational Psychologist, 29*(1), 3–12.

Bereiter, C., & Scardamalia, M. (1996). Rethinking learning. In D. R. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 485–513). Cambridge, MA: Blackwell.

Bernstein, B. (2000). *Pedagogy, symbolic control and identity: Theory, research, critique*. Lanham: Rowman and Littlefield Publishers, Inc.

Cope, B., & Kalantzis, M. (2017). *E-learning ecologies: Principles for new learning and assessment*. New York: Routledge.

Council of Europe. (2018). *Common European framework of references for languages: Learning, teaching, assessment*. Companion Volume with New Descriptors. Provisional Edition. Language Policy Programme. Education Policy Division. Education Department. Council of Europe. www.coe.int/lang-cefr. Accessed 3 May 2020.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Dalton-Puffer, C. (2016). Cognitive discourse functions: Specifying an integrative interdisciplinary construct. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualizing integration in CLIL in multilingual education* (pp. 29–54). Bristol: Multilingual Matters.

Davison, C. (2008). Assessment for learning: Building inquiry-oriented assessment communities. *Paper presented at 42nd Annual TESOL Convention and Exhibit*, New York.

deBoer, M. (2017). Implementing the CEFR as a benchmark for "action-oriented" syllabus design. In F. O'Dwyer, M. Hunke, A. Imig, N. Nagai, N. Naganuma, & M. G. Schmidt (Eds.), *Critical, constructive assessment of CEFR-informed language teaching in Japan and beyond* (Vol. 6, pp. 155–174). Cambridge: Cambridge University Press.

deBoer, M. (n.d.) Unpublished doctoral dissertation. *Analysis of online asynchronous learner-learner dialogue in a CLIL classroom*. University of Birmingham.

Dougiamas, M. (2011). *Moodle*. Perth: Curtin University of Technology.

Educational Testing Service. (2018). *ETS: TOEFL ITP*. http://www.ets.org/toefl_itp. Accessed 13 Sept 2018.

Engeström, Y., & Sannino, A. (2012). Whatever happened to process theories of learning? *Learning, Culture and Social Interaction, 1*(1), 45–56. https://doi.org/10.1016/j.lcsi.2012.03.002.

Evnitskaya, N., & Morton, T. (2011). Knowledge construction, meaning-making and interaction in CLIL science classroom communities of practice. *Language and Education, 25*(2), 109–127.

Feuerstein, R., Feuerstein, R. S., & Falik, L. H. (2010). *Beyond smarter: Mediated learning and the brain's capacity for change*. New York: Teachers College Press.

Green, A. (2016). Testing four skills in Japan. *British Council New Directions in Language Assessment*. *JASELE Journal Special Edition*, 135–144.

Kalantzis, M., & Cope, B. (2012). *New learning: Elements of a science education* (2nd ed.). Cambridge University Press.

Kupetz, M. (2011). Multimodal resources in students' explanations in CLIL interaction. *Research on Youth and Language, 5*(1), 121–142. http://www.novitasroyal.org/Vol_5_1/kupetz.pdf. Accessed 24 April 2020.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York: Routledge.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Westport: Ablex Publishing.

Leontjev, D., & deBoer, M. (2020). Multimodal mediational means in assessment of processes: An argument for a hard-CLIL approach. *International Journal of Bilingualism and Bilingual Education.* https://doi.org/10.1080/13670050.2020.1754329.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol: Multilingual Matters.

Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment. The Guilford school practitioner series*. New York: The Guilford Press.

Longacre, R. E. (1996). *The grammar of discourse* (2nd ed.). New York: Plenum Publishing Corporation.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge, UK: Cambridge University Press.

Met, M. (1998). Curriculum decision-making in content-based language teaching. In J. Cenoz & F. Genesee (Eds.), *Beyond bilingualism: Multilingualism and multilingual education* (pp. 35–63). Clevedon: Multilingual Matters.

METI. (2010). *Release of report by global human resource development committee of the industry-academia partnership for human resource development*. Tokyo: Ministry of Economy, Trade and Industry.

Nikula, T., Dalton-Puffer, C., & Llinares, A. (2013). CLIL classroom discourse: Research from Europe. *Journal of Immersion and Content-Based Language Education, 1*(1), 70–100. https://doi.org/10.1075/jicb.1.1.04nik.

Ohta, A. S. (2001). *Second language acquisition processes in the classroom: Learning Japanese*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Dwyer, F., & DeBoer, M. (2015). Approaches to assessment in CLIL classrooms: Two case studies. *CercleS, 5*(2), 1–25.

Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development*. Norwell: Springer.

Poehner, M. E., & Lantolf, J. P. (2005). Dynamic assessment in the language classroom. *Language Teaching Research, 9*(3), 233–265.

Richards, J. C. (2012). *New interchange 1* (4th ed.). New York: Cambridge University Press.

Scardamalia, M., Bereiter, C., & Lamon, M. (1994). The CSILE project: Trying to bring the classroom into world 3. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 201–228). Cambridge, MA: MIT Press.

Skinnari, K., & Bovellan, E. (2016). CLIL teachers' beliefs about integration and about the professional roles: Perspectives from a European context. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 145–167). Bristol: Multilingual Matters.

van Lier, L. (2008). Agency in the classroom. In J. P. Lantolf & M. E. Poehner (Eds.), *Sociocultural theory and the teaching of second languages* (pp. 163–186). London: Equinox Publishing Ltd.

Wells, G. (1999). *Dialogic inquiry: Towards a sociocultural practice and theory of education*. Cambridge: Cambridge University Press.

Wells, G. (2000). Dialogic inquiry in education: Building on the legacy of Vygotsky. In C. D. Lee & P. Smagorinsky (Eds.), *Vygotskian perspectives on literacy research: Constructing meaning through collaborative inquiry* (pp. 51–85). New York: Cambridge University Press.

Wells, G. (2001). The case for dialogic inquiry. In G. Wells (Ed.), *Action, talk and text: Learning and teaching through inquiry* (pp. 171–194). New York: Teachers College Press.

Wertsch, J. V. (1994). The primacy of mediated action in sociocultural studies. *Mind, Culture, and Activity, 1*(4), 202–208. https://doi.org/10.1080/10749039409524672.

Chapter 11
Conclusion: Dialectics in CLIL Classrooms

Mark deBoer and Dmitri Leontjev

Based on the insights that emerged in the chapters to this volume, in this chapter, we revisit the following: (a) relationship between teaching, learning, and assessment in the classroom-based assessment cycle (Davison 2008) and (b) integration in assessment in CLIL (Leung and Morton 2016). Two guiding questions will mediate our discussion: *What is assessment promoting learning in CLIL?* and *How can assessment promoting learning in CLIL help to conceptualise assessment promoting learning in general?*

We will then sketch directions that future research could address in order to further conceptualise assessment promoting learning in CLIL classrooms.

11.1 Reconstructing the Models

The conceptual discussions of assessment that the contributors to the present volume engaged in, building on our joint understanding of assessment in CLIL with reference to Davison's (2008) assessment cycle and Leung and Morton's (2016) integration matrix, served as our starting point. Every chapter approaches assessment in CLIL from a different angle. However, what unites them is that they discuss connections both among teaching, learning, and assessment and between content and language. This serves an important basis for forming a coherent understanding

M. deBoer (✉)
English for Academic Purposes, Akita International University, Akita, Japan
e-mail: markdb@aiu.ac.jp

D. Leontjev
Department of Language and Communication Studies, University of Jyväskylä, Jyväskylä, Finland
e-mail: dmitri.leontjev@jyu.fi

M. deBoer, D. Leontjev (eds.), *Assessment and Learning in Content and Language Integrated Learning (CLIL) Classrooms*,
https://doi.org/10.1007/978-3-030-54128-6_11

of assessment promoting learning in CLIL classrooms that brings two central models in the volume together.

11.1.1 Teaching, Learning, and Assessment in Classroom-Based Assessment Cycle

Wiliam and Leahy (2015) discuss classroom-based assessment as an interface between teaching and learning in the classroom. We can further understand this relationship if we critically engage with the central part of Davison's model of classroom-based assessment cycle (Fig. 1.1 in Chap. 1, this volume). That is, Davison (2008) placed teaching, learning, and assessment at the centre of the figure, which implies that all three are equal and contingent on one another and all three change within and across assessment cycles. We note that this idea of interaction of teaching, learning, and assessment is not new, and has been proposed and discussed previously both in the field of CLIL (Mehisto and Ting 2017) and elsewhere (Turner and Purpura 2016).

We, however, argue for the usefulness of viewing the relationships among teaching, learning, and assessment in the classroom as *dialectical* (see Lantolf and Poehner 2014). Such understanding, we propose, allows for seeing teaching, learning, and assessment in the classroom as a coherent whole, without losing the importance of the role of each of these three.

Cause and effect is an example of a non-dialectical relationship: if you flip a switch, lights go on. Dialectics, on the other hand, is a way of seeing separate, or even conflicting, phenomena or processes as forming a unity. For example, a dialectical relationship between a pencil and an eraser during a writing process (see Lantolf and Poehner 2014) can be understood from how they are *both* used as a way of writing something on a piece of paper. To be clear, writing and erasing *are* different. However, they have a quality that allows for considering them as a unity qualitatively different from the sum of its parts. The pencil is a tool that provides the writer with the means to write his/her ideas down. The eraser serves as a negation tool. The pencil and the eraser do not direct the writing process separately, as using one cannot be considered without using the other. The writer knows about what both the pencil and the eraser afford.

We can take the above example and with the same thinking, conceptualise the relationships among teaching, learning, and assessment in the classroom. The following Fig. 11.1 helps to conceptualise these dialectical relationships.

Figure 11.1 helps to visualise teaching, learning, and assessment as three dialectical unities (the oval shapes marked by different shades of grey: (a) teaching-learning, (b) teaching-assessment, and (c) learning-assessment. As one examines one of the unities, for example teaching-learning, one can think about how the remaining third element impacts and is simultaneously impacted by this unity.

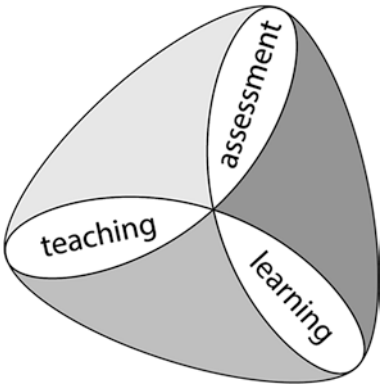


Fig. 11.1 Dialectical relationships between teaching, learning, and assessment

Teachers' instruction is never fully followed by learners (see Chap. 1, this volume) but is taken up and used differently by the individual learners. It is through this realisation that the dialectical unity of teaching and learning emerges. We propose that assessment whose purpose is to guide the development of learners' content *and* language knowledge mediates the dialectical unity of teaching and learning in CLIL.

The teaching and assessment relationship is not cause-and-effect either. Assessment informs teaching, and as teaching changes, so does assessment. A teacher's interaction with learners, for example, can be analysed both as assessment and as teaching, though in reality, one cannot be considered without the other (see Poehner and Infante 2015, for a detailed discussion of this relationship). As learners react differently to teachers' turns in interaction, so do the ways that teachers assess and instruct learners. Learning, thus, impacts on the teaching-assessment unity, which, in turn, shapes and directs learning.

The final dialectical unity is learning-assessment. Through assessment (teacher-, peer-, or self-assessment), learners gain understanding of their learning. In turn, learners' interpretation of the assessment, the way they perform during the assessment, and how they assess themselves and their peers, shapes their learning. Their learning, in turn, also guides their self- and peer- assessment, their understanding of teacher's assessment in the classroom, and their performance during the assessment. Teaching impacts on changes in the unity of learning-assessment simultaneously being impacted by this unity.

The key point to understand from this is that each dialectical unity teaching-learning, teaching-assessment, and learning-assessment cannot be considered as separate. Chapters 8, 9, and 10 can serve as an illustration of these relationships at the level of classroom interaction. In Chaps. 8 and 9, the teacher's on-going assessment needed teaching. Likewise, teaching required assessment of the learners' reactions to the guidance/instruction. One could not happen without the other. The interaction in Chap. 10 of this volume was among learners. However, the same understanding can be applied there. In order to progress, learners were required to constantly assess themselves and their peers and guide their own and other's

performance as a result. The interpretation of this process, too, changed as the learner interaction unfolded.

The same dialectical understanding can also be extended across assessment cycles. One may argue that such tools as scoring rubrics or the CEFR descriptors (see Chaps. 2, 3, 4, and 5, this volume) are fixed and, therefore, are difficult to understand as a part of the dialectical assessment/teaching/learning process. We, however, argue that they can be integrated into this process, as learners' and teachers' interpretation and utilisation of these tools changes, too, as the process unfolds.

To come back to our initial argument, we suggest there is no cause-and-effect relationship between assessment, teaching, and learning (be it on the level of single activities or across assessment cycles). For practical purposes, one can focus on any one element in Fig. 11.1. However, the dialectical relationships among teaching, learning, and assessment in the classroom should not be lost sight of.

11.1.2 Revisiting Integration of Content and Language in CLIL

Even though this sounds evident, we would like to underscore once again that CLIL stands for content-and-language *integrated* learning. Leung and Morton's (2016) matrix is a useful way of introducing different approaches to how content and language could be assessed in CLIL.

Perceiving content and language in CLIL as a dialectical unity helps to see the development of learner content and language knowledge *not* as two separate processes but as developing together. Thinking of content and language as the two sides of the same coin (the same construct) allows to see how content mediates the use of language while language mediates the understanding and knowledge of the content in learners' performance. It also helps to understand how content can *be used* to mediate learners' acquisition of language and language *be used* to mediate their content knowledge in the instruction following assessment which aims at promoting the learning of this integrated construct.

Thinking dialectically helps to perceive more clearly how the relationship between content and language is realised in different chapters to this volume. We next give brief examples from chapters in the volume to help the reader further understand this relationship.

In Chap. 9, once the learner verbalised that to save energy in the home, insulation is needed because it 'keeps out the cold', the teacher directed the learner to using scientific language, i.e. insulation is something that keeps in the heat, which guided the learner to understanding of how energy can be saved in a house. By eliciting a scientific term to explain the same phenomenon, the teacher simultaneously promoted the learner's linguistic knowledge ('heat' as a scientific term) and conceptual understanding. The outcome was more than simply one or the other—these were parts constituting the same development, one not possible without the other.

In Chap. 8 (pp. 267–268), on the surface, the teacher has done the reverse. Once a learner produced an academically and linguistically correct response 'bones can fracture', the teacher provided the learner with a non-academic synonym 'break'. While the intention was to give learners a strategy to mediate their language use, a connection between the words 'fracture' and 'break' was made more salient, thus promoting the learners' conceptual understanding of the word 'fracture'. In Lin's (2016, p. 12) words, CALP does not come naturally and requires instruction; therefore, teachers need to help learners move comfortably between BICS and CALP. The two examples show how this can be done.

The scales discussed in Chaps. 2 and 3 are yet another illustration of how the dialectical relationship between content and language can be realised in CLIL. Using the scales allows for systematically assessing where learner problems lie in the use of language and conceptual understanding. More importantly, they allow for establishing how content knowledge always plays a role in learners' linguistic performance and their linguistic knowledge impacts the development of their content knowledge. This process is facilitated if the scale includes an explicit component of mediation, as discussed in Chap. 2. It is namely the dialectical thinking that helps to understand how such scales can be used systematically for building on learners' strengths in content to address their weaknesses in language and vice versa.

The integration matrix informs the understanding of how specific assessment processes unfold. The dialectical understanding of the relationship between content and language in CLIL deepens this understanding. It compels CLIL researchers and educators to explore how content and language develop together. A metaphor we propose for this construct is that of a sphere (Poehner, personal correspondence). One can turn this sphere and focus on one side of it (content) and then turn it again and focus on the other side (language). However, without thinking about the other (language or content), one cannot comprehend the whole—it is still one and the same sphere, regardless of the part of this sphere one chooses to focus on for practical or empirical purposes.

The power of using Leung and Morton's (2016) matrix lies in that it informs what can be learned from various assessment activities and in which way this information can promote learning. However, in order to conceptualise classroom-based assessment as a continuous process, the unities that we discussed so far, teaching-learning-assessment and content-language, should be considered together.

11.2 Assessment in CLIL as a Coherent Whole

In this section, we bring the two central conceptualisations in the volume—the classroom-based assessment cycle (Davison 2008; Davison and Leung 2009) and the integration matrix (Leung and Morton 2016)—together.

Different assessment activities can be perceived as either focusing on the language, on the content, or oscillating between the two. The understanding of content and language as entering into a dialectical relationship changes the way that the

inferences are made from learners' assessment performance, how the information is delivered to learners, and which adjustments to teaching are made. Thinking of content and language as impacting on one another is different from thinking in terms of the focus of the assessment on either or both of them. The teacher will always think about how content mediates the language in learners' performance and vice versa regardless of the focus of the assessment activity. The concept of classroom-based assessment cycle, in turn, allows for adjustments in teaching, learning, and assessment to be informed by the previous assessment cycles, shaping the planning of assessment activities and further shaping the inferences that are made about learner performance. It should not be forgotten that these adjustments are still made with reference to the goals of the course and the curriculum (Mehisto and Ting 2017). However, the understanding of how the path towards these goals goes becomes more systematic as a fuller picture of learner performance emerges due to using both the integration matrix and the classroom-based assessment cycle.

To better illustrate how the two models reimagined from the perspective of dialectics inform classroom-based assessment in CLIL, we next show how the assessment activities and approaches discussed in different chapters of this volume can be used together keeping the two models (classroom-based assessment cycle and integration matrix) in mind.

Referring back to Sect. 1.8 of Chap. 1, a rubric (see Chaps. 2 and 3, this volume) can inform the teacher what the learner can and cannot do with regard to a particular benchmark. This information mediates the teaching-learning process in that the teacher in a performance-oriented, more visible language pedagogy, then directs the learner's performance depending on whether the identified gaps refer to the language or the content or both. In the subsequent assessment cycle, assessment as a part of dialogic interaction can be used (Chaps. 8 and 9). This assessment, building on the previous cycle, mediates the teaching-learning process differently. Now it is a part of centrifugal interaction, allowing the teacher to see how much external assistance the learner needs in order to develop. The teacher can also probe how the learner's strengths in conceptual knowledge identified in the previous cycle can be used to mediate the learner's linguistic knowledge and vice versa.

We argue that whichever of the two central frameworks of assessment promoting learning CLIL educators subscribe to, assessment *for* learning or the more detailed learning-oriented assessment (or any other framework), changes in teaching, learning, and assessment should be systematically traced *together*. The assessment *for* learning conceptualisation elicits that assessments should be designed with the purpose to promote learning. Hence assessment should give information on how learning should be promoted. This implies that assessment should itself change as more information about learners is gained and as learning happens. The LOA framework (see Chaps. 4 and 5, this volume) helps teachers understand assessment as happening at all of the different dimensions it entails as well as to consider how these dimensions interact in the classroom-based assessment cycles.

The classroom assessment cycle and the integration matrix models make assessment in CLIL classrooms systematic. Using a framework/model as the one

discussed in Chap. 7 can add to this systematicity, helping teachers trace how different assessment activities promote learning.

11.3 What Can Assessment in CLIL Offer to Other Educational Contexts?

Throughout the process of creating this volume, the contributors, us included, asked the question of the role of CLIL in assessment promoting learning. The latter was one of the themes of the symposium in Tokyo in which several contributors to the present volume participated (Leontjev and deBoer 2018). The symposium's roundtable discussion provided ideas as to what CLIL can offer to A*f*L. Hence, the following is a product of a collective thinking. Above all, we owe the following discussion to our contributors and to the participants in the symposium in Tokyo.

As discussed during the symposium, CLIL is both a way to teach language alongside content and a way to understand what happens in the classroom. Rich insights into learner abilities are possible because CLIL teachers are both content- and language-aware. The outcome of our discussion during the symposium was that we expanded the oftentimes used statement 'every teacher is a language teacher' (FNBA 2014; Gottlieb and Ernst-Slavit 2014; Zwiers 2008; Walqui and van Lier 2010) to *every content teacher is a language teacher and every language teacher is a content teacher*. This statement brings language awareness to content lessons and content awareness to language lessons, which should lead to teaching, learning, and assessing of language and content as an integrated construct.

The underlying theoretical principles of CLIL can inform other educational contexts. These include (a) the sociocultural understanding of development as mediated and knowledge as co-constructed, (b) a necessary increase in both linguistic and cognitive demands placed on learners as their development happens (Bloom's Taxonomy) (Anderson et al. 2001), and (c) contextualisation of language use (Long's (1996) Interaction Hypothesis). (See also Chaps. 2, 5, and 6, this volume; Wewer 2014 for discussions.) These principles of CLIL are hopefully shared by CLIL teachers around the world. It is due to these principles that pedagogical processes in the classroom that relate to the development of content and language knowledge together become visible in CLIL classrooms.

One useful way that content teachers can think of the development of learners' conceptual knowledge is with reference to the development of disciplinary language.

Vygotsky (1978, p. 148) stated that in the developmental process, scientific concepts approach concrete phenomena whereas everyday concepts move towards scientific generalisations. As learners acquire ways of talking about concepts (using their first language or any additional languages), their conceptual knowledge develops, too. We argue that using academic language has a central role in this development, as learners acquire ways they talk and write to various communities of practice.

Recognising uncertainty that content teachers may have with regard to language and language teachers, with regard to content, we suggest collaboration between content teachers and language teachers (see also Zappa-Hollman 2018). One goal of such collaboration can be developing assessment criteria and scales (see Chaps. 2, 3, and 6, this volume) for content and language lessons. We suggest that such collaboration can be especially fruitful in higher education. Oftentimes, writing courses in higher education convey a lack of collaboration between academic language and content instructors, the outcome being that the language instructors are not fully aware of what it means to write to specific academic communities. The expectation is that learners are to transfer what they learned in academic writing courses to writing in their respective subjects. Jointly developing a scale having language and content criteria can become a starting point for collaboration between language and content instructors.

A way of being aware how content mediates language and vice versa and use this information to systematically direct learning is at the micro level of classroom interaction (Chaps. 8, 9, and 10, this volume). Content teachers, through being conscious towards the learners' use of language as they discuss academic concepts, can start consciously developing learners' disciplinary language alongside conceptual knowledge. Language teachers can start appreciating learners expressing their conceptual understanding on certain topics, and consciously mediating their conceptual understanding rather than using these topics to introduce grammatical and linguistics categories.

Chapter 4 brings in language awareness in a different way—on the level of program development. The gradual move from EAP (English for Academic Purposes; emphasis on language) to EMI (English-Medium Instruction; emphasis on content) works well at the macro level of the overall progression and the growing demands, as learners are socialised into the respective academic communities.

Our main argument here is, to repeat, that richer insights into learners' abilities that CLIL contexts allow are possible to obtain in other educational contexts. We emphasise that CLIL lessons are not the same as content lessons or language lessons. Differently from language lessons, in CLIL, the language the learners are expected to learn is disciplinary. In content lessons, the emphasis is rarely on the language, disciplinary or otherwise, even if it is expected that learners should socialise into the academic community, learning to talk and write scientifically. That said, the understanding of content and language as two sides of the same coin, acquired simultaneously can inform teaching, learning, and assessment in content and language lessons alike. The outcome should be that content ceases to be but a context for introducing linguistic and grammatical categories in language lessons (as argued in Chap. 7), and disciplinary language and writing conventions are consciously and systematically paid attention to in content lessons. Language teachers, as a result, should become content-aware and content teachers, language-aware and start eliciting both in their lessons, paying attention to how one mediates the other and consciously using both content and language to direct learning.

11.4 Ways Forward

We have not given the reader a one-size-fits-all answer to the question of what and how to assess in CLIL but attempted to conceptualise classroom-based assessment in a way that gives teachers a range of possibilities to implement it. The power of this approach is in that it, in Coyle, Hood, and Marsh's (2010, p. 69) words, allows teachers for "sharing their own understanding of what it is to be taught and learned, transforming ideas into 'teachable' and 'learnable' activities, connecting these with decisions about the optimal organisation of the learning environment, followed by evaluation, reflection and new understandings for classroom teaching and learning."

With regard to research, further work in both conceptualising assessment promoting learning in CLIL and developing assessment tools and approaches is needed. We suggest three main directions that the future research on assessment in CLIL can aim: (a) curriculum and pedagogy planning, (b) participant perspectives, perceptions, and beliefs, and (c) classroom practices (see Nikula et al. 2016b).

The conceptualisation of integrating content and language and assessment in curricula is essential for understanding how teaching, learning, and assessment are organised at different levels of education. This should create a stronger basis for truly integrated curricula, which, as Nikula et al. (2016b) rightfully note, are a rare find. Educational policy research with the focus on assessment in CLIL is, therefore, much needed. The move from educational policy to classroom practices requires also looking into stakeholders' (above all teachers and learners) perspectives, perceptions, and beliefs with regard to integration of content and language and its assessment. Understanding teachers' beliefs and perceptions of content and language in CLIL and the relationships among teaching, learning, and assessment is crucial for understanding their teaching, learning, and assessment practices as well as for changing these same beliefs and perceptions and for developing these practices. CLIL teachers' perceptions and beliefs can range from considering language learning as 'a side effect' and not assessing it systematically (Quadrant 2), to seeing content more as context for teaching and assessing language (Quadrant 3), to placing equal importance to both content and language (Quadrant 1), to letting the focus of assessing and teaching emerge in interaction (Quadrant 4). These beliefs and perceptions can also change as teachers gain new perspectives and understandings of teaching, learning, and assessment in CLIL classrooms, shaping, in turn, teachers' practices. Finally, classroom-based research in teaching, learning, and assessment in CLIL and beyond can provide insights which are invaluable for conceptualising and developing practices in assessment in CLIL. This implies that further research should be more fruitful to continue with an interdisciplinary orientation, bringing together researchers in applied language studies, researchers in assessment, educational policy researchers, and educational researchers in various CLIL contexts and in various content disciplines.

We suggest that for developments in CLIL research and practice to be the most impactful, research and practice should enter into a dialectical relationship—praxis (Lantolf and Poehner 2014; Lantolf and Poehner with Swain 2018). The dialectical

understanding of research-practice relationship changes collaborations between teachers and researchers. Researchers enter the contexts with a view of developing teacher practices rather than only observing them, using their theoretical and conceptual understandings. Teachers, in turn, through their practices, validate and develop researchers' theoretical and conceptual understandings, building on their expertise as educational professionals. To the best of our knowledge, there have been but a few recent examples of researchers collaborating in a similar way with CLIL teachers (Banegas 2013; Lo 2019).

11.5 Teacher Collaboration

We would like to end this volume by explicitly addressing the part of our prospective readership who are educators. There are many excellent examples of practical teaching and assessment activities and ideas that help educators bring CLIL into their classrooms (see Ball et al. 2015; Lin 2016; Mehisto and Ting 2017). Lin's (2016) Chap. 5 is particularly useful with regard to assessment activities that CLIL teachers can adapt for their classes. We propose that CLIL teachers could use these activities alongside those discussed in the chapters of this volume with the view of bringing assessment, teaching, and learning together in their classroom practices.

Lastly, we would like to expand on our argument in Sect. 11.3 for collaboration among teachers. We envision three general ways collaboration can happen. First of all, at the school level, CLIL teachers can collaborate with other teachers, e.g. a CLIL teacher having more of a content teacher identity working together with a language teacher. On the national level, CLIL teachers can share their teaching and assessment practices through CLIL teacher associations, such as *J-CLIL* in Japan and *Suvikyky r.y.* in Finland. CLIL teacher identity discussed in Nikula et al. (2016a) underscores the importance of yet another level of collaboration—internationally. CLIL teachers in Japan often identify themselves as language teachers, whereas in European countries, the identity of CLIL teachers is oftentimes that of content teachers, as has also been illustrated in the present volume. International CLIL teacher collaboration, therefore, implies sharing markedly different perspectives on what CLIL and assessment in CLIL are in classrooms around the world. These understandings, perspectives, and practices can then be brought back to the research community for the development to continue.

References

Anderson, L. W., Krathwohl, D. R., & Airasian, P. W. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Pearson.

Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice*. Oxford: Oxford University Press.

Banegas, D. L. (2013 Doctoral Dissertation). *Teachers developing language-driven CLIL through collaborative action research in Argentina*. Coventry: University of Warwick.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge: Cambridge University Press.

Davison, C. (2008, March). *Assessment for learning: Building inquiry-oriented assessment communities*. Paper presented at 42nd Annual TESOL Convention and Exhibit, New York.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x.

Finnish National Board of Education (FNBE). (2014). *Perusopetuksen opetussuunnitelman perusteet 2014* [National Core Curriculum of Basic Education 2014]. Retrieved March 23, 2020 from http://www.oph.fi/download/163777_perusopetuksen_opetussuunnitelman_perusteet_2014.pdf.

Gottlieb, M., & Ernst-Slavit, G. (2014). *Academic language in diverse classrooms: Promoting content and language learning: Definitions and contexts*. Thousand Oaks: Corwin.

Lantolf, J. P., & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. New York, NY: Routledge.

Leontjev, D., & deBoer, M. (2018, October 20). *Symposium: Assessment for learning in CLIL*. Conceptualisations and practical applications. Tokai University, Japan.

Leung, C., & Morton, T. (2016). Conclusion: Language competence, learning and pedagogy in CLIL – Deepening and broadening integration. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 235–248). Bristol/Buffalo: Multilingual Matters.

Lin, A. M. Y. (2016). *Language across the curriculum & CLIL in English as an additional language (EAL) contexts: Theory and practice*. Singapore: Springer.

Lo, Y. Y. (2019). Development of the beliefs and language awareness of content subject teachers in CLIL: Does professional development help? *International Journal of Bilingual Education and Bilingualism, 22*(7), 818–832. https://doi.org/10.1080/13670050.2017.1318821.

Mehisto, P., & Ting, T. Y. L. (2017). *CLIL essentials for secondary school teachers*. Cambridge, UK: Cambridge University Press.

Nikula, T., Dafouz, E., Moore, P., & Smit, U. (Eds.). (2016a). *Conceptualising integration in CLIL and multilingual education*. Bristol/Buffalo: Multilingual Matters.

Nikula, T., Dalton-Puffer, C., Llinares, A., & Lorenzo, F. (2016b). More than content and language: The complexity of integration in CLIL and bilingual education. In T. Nikula, E. Dafouz, P. Moore, & U. Smit (Eds.), *Conceptualising integration in CLIL and multilingual education* (pp. 1–25). Bristol/Buffalo: Multilingual Matters.

Poehner, M. E., & Infante, P. (2015). Mediated development: Inter-psychological activity for L2 education. *Language and Sociocultural Theory, 2*(2), 161–183. https://doi.org/10.1558/lst.v2i2.26982.

Turner, C. E., & Purpura, J. E. (2016). Learning-oriented assessment in second a foreign language classrooms. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 255–273). Berlin: Walter de Gruyter Inc.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walqui, A., & van Lier, L. (2010). *Scaffolding the academic success of adolescent English language learners*. San Francisco: WestEd.

Wewer, T. (2014). *Assessment of young learners' English proficiency in bilingual content instruction CLIL* (Doctoral dissertation). University of Turku. http://urn.fi/URN:ISBN:978-951-29-5754-5. Accessed 12 May 2018.

Wiliam, D., & Leahy, S. (2015). *Embedding formative assessment: Practical techniques for K-12 classrooms*. United States of America: Learning Sciences International.

Zappa-Hollman, S. (2018). Collaborations between language and content university instructors: Factors and indicators of positive partnerships. *International Journal of Bilingual Education and Bilingualism, 21*(5), 591–606. https://doi.org/10.1080/13670050.2018.1491946.

Zwiers, J. (2008). *Building academic language: Essential practices for content classrooms, grades 5–12*. San Francisco: Jossey-Bass.