Chapter 11
Transforming Assessment Research: Recommendations for Future Research

Jan Alexis Nielsen, Jens Dolin, and Sofie Tidemand

Introduction

As a research project, ASSIST-ME produced a large number of results both within and across the eight participating partner countries using a variety of research methods. Based on the preceding chapters, this chapter will organise, prioritise and summarise the principal outcomes. It seems reasonable to assume that many of the findings presented in the preceding chapters can inform further research with the fields of classroom assessment or science education (or both). For example, it is quite clear that the concept of inquiry teaching, while being central in the field of science education for two decades, still is difficult to define in clear and uniform terms (Rönnebeck et al. 2016). In concert, the chapters at the very least provide a state-of-the-art terminology about inquiry-related learning outcomes and how they are assessed that can act as a strong scaffold for future research on inquiry Science, Technology and Mathematics (STM) teaching.

We will in this chapter identify and outline current gaps in research into assessment practice and tie the results of the ASSIST-ME project onto this outline. In this way, the chapter will present concrete research vistas that are still needed in international assessment research. The chapter concludes with a key theme that appears across many of the chapters in this volume, namely, issues concerning the *operationalisation* of complex learning goals into teaching and assessment activities.

J.A. Nielsen (✉) • J. Dolin • S. Tidemand
Department of Science Education, University of Copenhagen, Copenhagen, Denmark
e-mail: janielsen@ind.ku.dk

Cross-Cutting Trajectories

We start by extrapolating three trajectories across the chapters that seem to be particularly promising for future research. These pertain to (i) using competences as a theoretical foundation in assessment, (ii) placing summative and formative assessment on a continuum and (iii) identifying the need for teachers to be supported when introducing new assessment formats. These trajectories will subsequently be discussed in the ensuing sections of this chapter.

First, Ropohl et al.'s (Chap. 1) exposition of *competences* as a way of parsing learning objectives in science education and Rönnebeck et al.'s (Chap. 2) delineation of how competences related to STM inquiry could be *assessed*, provide a very direct vision of how the field of science education could understand and make operational concepts such as *Bildung*, scientific literacy and inquiry teaching (for a discussion of the connection between scientific literacy and *Bildung*, see Sjöström and Eilks 2017). Ropohl et al.'s (Chap. 1) analysis of the concept of competence indicates that the concept is multifaceted and often used ambiguously, but in so doing, the chapter provides much needed clarity and indicates ways forward by approaching competences from what they call a 'holistic' perspective. Further, Ropohl et al. provide a general vantage point for an understanding of competences within science, technology and mathematics education. In particular, Ropohl et al.'s push to conceptualise complex learning objectives in competence terms forms a backdrop for Rönnebeck et al.'s (Chap. 2) analysis of assessment of learning in inquiry teaching. Rönnebeck et al.'s analysis builds on a detailed systematic literature review reported in Bernholt et al. (2013), and the chapter provides a much needed translation of inquiry-related learning outcomes into competence terms, not just for research to reach a shared understanding of inquiry but also indirectly to support teaching and assessment practice concerning inquiry teaching.

Second, while researchers within the field of classroom assessment have long distinguished between formative and summative assessment, Dolin, Black, Harlen and Thiberghien (Chap. 3) provide a framework for understanding the dynamics of the interplay between formative and summative assessment. In particular, the chapter contains concrete ideas for ways of linking formative and summative forms of assessment and for how formative and summative forms of assessment can be seen as belonging to a spectrum rather than as binary forms of assessment. Now, these ideas have bearing on the empirical studies of both *structured assessment dialogues* (Dolin et al., Chap. 3) – a regimented assessment procedure that allows for multifarious layers of assessment activities and perspectives – and the scaffolded approach to teacher *written feedback* explored by Holmeier et al. (Chap. 7).

Third, it seems that the teachers who were involved in the various empirical studies presented in this volume in general needed support to use the various assessment formats as well as to plan and implement inquiry STM teaching (see Chaps. 4, 5, 6, 7 and 9). In particular, at least three chapters conclude that concerted professional development efforts are needed to support teachers to provide formative feedback in inquiry teaching (see Chaps. 4, 7 and 9). The need for support notwithstanding, it

seems equally clear that the process of repeatedly trying out the different methods for assessing inquiry-related competences did provide a strong basis for the teachers to establish an understanding of inquiry (or rather inquiry-related competences) that in turn stabilised their formative assessment practice.

Recommendations for Future Research Foci and Methods

We next present what we find to be key lacunae in classroom assessment research that future research should address. As such, we reflect on key research foci that we feel should be pursued in the future. Further, we identify an underlying theme that reoccurs in various ways across many of the research findings in the chapters of this volume. We present this theme, argue why this theme is important and provide a terminological and analytical framework that may open a new research vista for classroom assessment research.

The chapters in this volume all address issues that are boundary objects (Star and Griesemer 1989). Indeed, the issues that are explored can be approached from the perspective of classroom assessment research which we can take to be a strand of general education research *and* from the perspective of science education research. When looking across the issues in the preceding chapters, it is difficult to define exactly when it is advisable to draw on background research from the general educational field and when to draw on research about assessment in subject-specific contexts – such as the STM subjects. Clearly, drawing on classroom assessment research in general would have afforded a much more comprehensive background into typical issues surrounding teachers' assessment practices regardless of which subject the teachers are teaching. But it is still an open question whether assessment practices differ across disciplines and if so how they differ and what such differences signify (Ruiz-Primo and Li 2012). Also, focussing on the assessment of complex inquiry competences may to some extent preclude the application of much of the existing research into subject-specific assessment practices that will often focus on assessment of concrete subject content. For example, for science education, there exists some research on the development and validation of learning progressions (Wilson 2009), but such work typically focusses on concrete subject-specific conceptual content rather than generic competences such as those that are the aims of inquiry teaching. Generic competences are here taken to be competences that can be at play in or the aim of multiple different disciplines, e.g. communication competences, innovation competences, problem-solving competences, collaboration competences and argumentation competences (see, e.g. Belova et al. in press).

Future research endeavours must take this into account: research into assessment in relation to inquiry teaching still requires a fair amount of extrapolation – either by assimilating findings from general education assessment research or by transposing findings and principles about assessment of more specific disciplinary skills.

Cowie (2012) recently argued that '[t]he current focus on more directly aligning the system of assessment (national to classroom), curriculum, and pedagogy comes

with the prospect of this suggesting the need for research that tracks development at all levels of the system and across all stakeholders' (p. 484). The primary idea behind Cowie's call for comprehensive research of this type is the fact that the *assessment culture* of an educational system manifests itself in various ways across stakeholder types – for example, teachers' narratives about the assessment culture in which they are actors will probably be very different from the narratives of parents (compare, e.g. Brookhart 2012) or government officials overseeing centralised assessment (see Moss 2008, for an argument to a similar effect). To be sure, the ASSIST-ME project that formed the basis of the preceding chapters took an important step in this direction by adopting an involvement of key stakeholder groups: researchers, teachers, leaders and policymakers (see Dolin et al., Chap. 10). In the project, this stereoscopic involvement ultimately resulted from a *change agenda* – with the aim of actively capitalising on the research results in order to impact teaching and assessment practice (see Dolin 2012). Arguably, more stakeholder types – most notably the learners – could and should play a larger role in research that comprehensively tracks the development of assessment culture in increasingly aligned educational systems.

An important issue in assessment research that had a somewhat implicit role in this volume is that of *validity*. The coding of the data in the project did entail some aspects of assessment validity – in particular in the case of peer feedback (Chap. 6) and to some extent in the case of written feedback (where the coders had to focus on the level of justification of the teachers' feedback to his/her students; Chap. 7) – and of course validity is a key aspect considered in the theoretical outcome of the project focussing on how summative and formative assessment can be linked constructively (Chap. 3). Future large-scale cross-national research projects similar to the ASSIST-ME project may focus more directly on validity aspects. Indeed, the role of validity in educational assessment research is difficult to circumvent (for a historical overview and an exposition of the importance of validity in educational assessment, see Newton and Shaw 2014). At this point it is relevant to emphasise that the ASSIST-ME project worked with complex learning objectives (inquiry competences in STM) that were not necessarily familiar to the participating teachers and that this may indicate validity concerns (even though reliability concerns may trump validity concerns in matters about formative assessment; see Chap. 3).

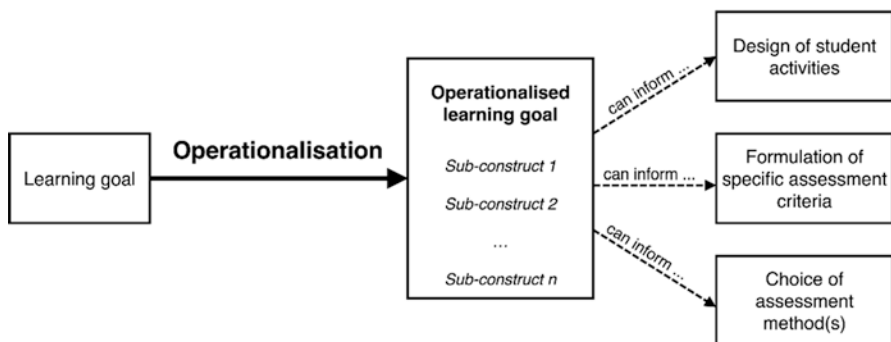## Operationalisation of Learning Goals: A New Research Vista?

In this section, we want to point to an underlying thematic process that we find permeates the research findings in most of the preceding chapters. We provide a first attempt to define that process and in so doing point to new research vistas for research on teachers' assessment practices. Let us start by making some observations from three chapters that describe empirical studies.

Evidence from the use of *structured assessment dialogues* (see Dolin et al., Chap. 6) indicates that the translation of relatively complex learning goals into more

concrete and operational constructs that can function on the level of assessment criteria proved to be important for the quality of feedback. Further, teachers' practice of engaging in on-the-fly formative interaction benefits from a close exposition of the assessment criteria or the construction of rubrics (see Harrison et al., Chap. 4). Similarly, the work put into translating more complex competence goals into criteria for the feedback templates used for written feedback seems to have been beneficial (see Holmeier et al., Chap. 7). To be sure, such templates that delineate the potential progression trajectories of students' competence development can aid the teacher in providing valuable feedback. But beyond this, the very fact that competence development is given a typified description seemed to help some of the participating teachers in making assessment transparent to their students.

What connects these findings is that the participating teachers went through a process of translating learning goals and that, in the context of these studies, this process to some extent was necessary for establishing high-quality assessment practices. Translating learning goals belongs to a process of *operationalising* learning goals. As such, the process has the aim of *making operational* learning goals, and the rationale behind the process is that the operationalised learning goals may provide *better* guidance for the teacher than the initial learning goals – by guidance we mean guidance on how to structure his or her teaching and on how to assess students' level of attainment of the learning goals. We have depicted a graphical model of the process of operationalising learning objectives in Fig. 11.1.

Take, for example, the following predefined learning goal for biology in a Danish upper secondary school: '[students should be able to] assess far-reaching biological issues and their significance on a local and global level' (Danish Ministry of Education 2013). On its face value, this particular learning goal arguably provides little guidance on how to structure teaching and assessment activities. In order for that learning goal to be instructive for a teacher (e.g. a teacher that is confronted with this learning goal for the first time), there must be a process through which the reader can *negotiate the meaning* of the goal (this negotiation of meaning could be

Fig. 11.1 A graphic illustration of the process of operationalising a learning goal. Notice that the product – the operationalised learning goal – does not necessarily involve sub-constructs; in many cases the product is an interpretation of the initial learning goal that enables the teacher to make decisions vis-á-vis teaching and assessment activities

scaffolded by attaching some sort of commentary to the curriculum). This kind of negotiation of meaning is exactly what the concept of operationalisation signifies. As we will describe below, the operationalisation of a learning goal can be more or less specific – ranging from a general interpretation to a detailed parcelling out of the initial learning goal into sub-constructs that can indicate assessment criteria on various taxonomical levels (see, e.g. Biggs and Collis 2014; Krathwohl 2002). So, at the very least, the operationalised learning goal, as a product of the process of operationalisation, is the teachers' interpretation of the initial learning goal.

The process that we are describing here should resonate quite well with both practitioners and researchers. Indeed, there is nothing new in that process. But it seems to us that it is important to specify the process in more detail than has been done in the existing literature. In fact, the process has often remained implicit in theoretical expositions. Clearly, the process of operationalising learning goals must be an important part of what Kattmann, Duit, Gropengiesser and Komorek (1996) called *Educational Reconstruction* – the process through disciplinary content is reconstructed into a curriculum or into teaching activities. Similarly, the process of operationalising learning goals would be a part of what Chevallard (1991) called the *internal didactic transposition*, i.e. the process with which a teacher transposes the aims of a curriculum into actual teaching (Winsløw 2011). But in both conceptualisations of the process from the disciplinary knowledge over the curriculum to the classroom activities, what we have called the process of operationalisation is at best only implied. While there seems to be no established body of work related to the particular process that we want to refer to as operationalising learning goals, the terminology we have chosen should be familiar to the field. For example, in the German curriculum, competence goals are fleshed out using 'operators' (German: *operatoren*) that are action verbs describing student activities that should be expected when the student is developing a particular competence.[1] Notice that our usage of operationalisation as a term indicates an active part of the teacher. While a curriculum can perform a part of the task of making learning goals operational, the teacher will need to perform at least a minimal operationalisation him- or herself (such as described by the notion of the internal didactic transposition).

We contend that any teaching practice will involve at least minimal processes of operationalisations of the sort we are describing here. But it seems reasonable to assume that the more complex the initial learning goals are, or the more unfamiliar a learning goal is to the teacher, the more there is a need for support for operationalisations. Arguably, the extent to which teachers go through interpretative processes that can be categorised as operationalisation of learning goals will vary between different educational cultures (see, e.g. Desurmont et al. 2008). A reasonable hypothesis could be that teachers in systems that belong to what can loosely be called the north and continental western European tradition have a strong tradition for going through such processes, for these are systems where the teacher traditionally has a relatively autonomous role of designing his or her teaching using a curriculum as a guide. But the explicit familiarity with processes of operationalisation

---

[1] See, e.g. https://lehrerfortbildung-bw.de/u_gewi/gwg/gym/bp2004/fb1/modul1/geo/operator/

can also differ between educational levels within one country. For example, findings from the ASSIST-ME project indicate that Danish lower secondary school teachers were much more familiar with processes that resemble operationalisation of learning goals than teachers from upper secondary school – probably due to differences between the curricula for lower and upper secondary school in Denmark (Nielsen and Dolin 2016).

It seems to us that there is a real need for thematising how to operationalise learning goals that teachers perceive as new and/or unclear (such as 'innovation competence'; see Nielsen 2015) or for other reasons perceive as unclear (such as is often the case with learning goals that relate to technology issues in science teaching; see Bungum 2006). Indeed, the previous chapters in this volume indicate that teachers for whom inquiry teaching introduced a new set of learning goals need substantial support in identifying viable strategies to plan their teaching and operate during their teaching vis-à-vis identifying and acting on opportunities to provide formative assessment (see, e.g. Harrison et al., Chap. 4).

We hypothesise that the process of operationalisation is of paramount importance – whether or not a given learning goal is complex, unclear or novel. As has been argued by Dysthe et al. (2008), if assessment 'criteria are explicitly formulated as reifications of continuous negotiations and participation, they become part of a meaningful learning process[; …] [e]xplicit criteria cannot be understood in isolation from the negotiation process' (p. 127). Indeed, the findings in the preceding chapters corroborate this statement. Crudely put, a criterion in itself is not yet operational for teaching and assessment.
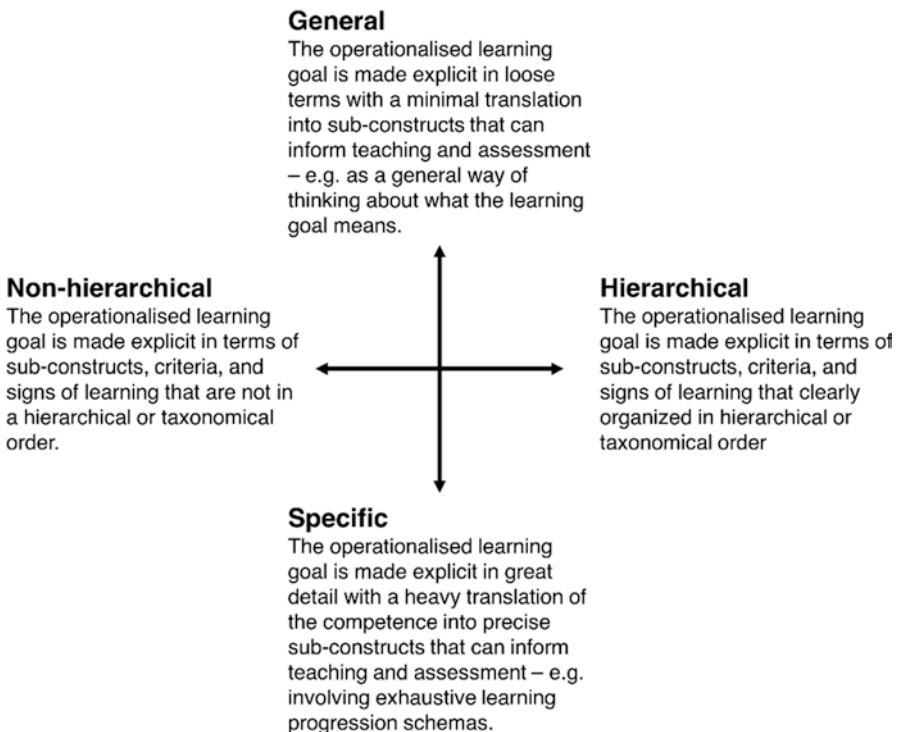
The term 'operationalisation' is frequently used in the field of validity research. For example, when investigating the validity of a construct, the key question is whether that construct was appropriately operationalised for functional measurement (see, e.g. Drost 2011). But we do *not* want invoking specific psychometric connotations with our usage of 'operationalisation' here. Our way of using 'operationalisation' also relates to questions regarding validity of assessment (*am I, as a teacher, really assessing the construct that I intend to assess?*), but we aim to signify a process which is closer to actual teaching practice and which pertains not just to assessment. Further, the term 'operationalisation' has been used in curriculum research. For example, Wiek et al. (2015) use 'operationalisation' to signify a process of making explicit a given general competence through a set of 'specific learning objectives for different educational levels' so as to inform curriculum design (p. 242). Again, we want to use 'operationalisation' here as signifying a process that is closer to actual teaching practice, rather than something that occurs during the construction of a curriculum.

There are clear indications in the findings of the ASSIST-ME project that the participating teachers found the process of operationalising competences overly time-consuming (see, Dolin et al., Chap. 3; Harrison et al., Chap. 4; Dolin et al., Chap. 5; Evans et al., Chap. 9). Moreover, many teachers in the project found it fundamentally difficult to operationalise general inquiry competences through learning progressions – even when teachers are being assisted by researchers. As reported by Dolin et al. (Chap. 5), the teachers who implemented structured

assessment dialogues tended to formulate rubrics that essentially not had the structure of a progression but rather consisted of unstructured or non-taxonomically ordered signs of student learning. Such operationalisations could be called *non-hierarchical* operationalisations of competences.

In the narratives of the participating teachers (for an exposition, see Dolin 2016; Nielsen and Dolin 2016), there are indications that some of the teachers felt that detailed learning progressions could lead to some form of instrumentalist assessment paradigm (see also Torrance 2007) involving teachers and students in following rudimentary learning checklist. As such some of the teachers were opposed to using what could be called *specific* operationalisations of competences. This issue harks back to a discussion by Rönnebeck et al. (Chap. 1) about the extent to which a generic competence can or should be deconstructed into a myriad of smaller constructs or whether a more holistic approach is preferable.

These findings indicate to us the benefit of thinking about teachers' operationalisation of competences as an activity that leads to a product or outcome that can be analysed along two continuums or dimensions (see Fig. 11.2). First, the outcome of a concrete act of operationalising a competence can be more or less *specific*. At one end of the continuum, the sub-constructs and/or assessment guides that are formulated in order to make the competence operational can be very precise and minute,

General
The operationalised learning goal is made explicit in loose terms with a minimal translation into sub-constructs that can inform teaching and assessment – e.g. as a general way of thinking about what the learning goal means.

Non-hierarchical
The operationalised learning goal is made explicit in terms of sub-constructs, criteria, and signs of learning that are not in a hierarchical or taxonomical order.

Hierarchical
The operationalised learning goal is made explicit in terms of sub-constructs, criteria, and signs of learning that clearly organized in hierarchical or taxonomical order

Specific
The operationalised learning goal is made explicit in great detail with a heavy translation of the competence into precise sub-constructs that can inform teaching and assessment – e.g. involving exhaustive learning progression schemas.

Fig. 11.2 A diagram of two possible dimensions of operationalising competences

e.g. by detailing potential signs of learning vis-à-vis multifarious sub-competences to be used in a single lesson. Alternatively, at the other end of the continuum, the operationalisation can result in a more general explication of the competence, e.g. by stipulating one or a few general signs of learning that can guide the assessment of the development of the competence over a year. Second, the outcome of a concrete act of operationalising a competence can be more or less *hierarchical.* At one end of this continuum, the sub-constructs and/or assessment guides that are formulated in order to make the competence operational can be structured in a hierarchical fashion, e.g. projecting potential learning trajectories in a learning progression format. At the other end of the continuum, the operationalisation can result in an array of sub-constructs and/or assessment guides that are not related or hierarchically ordered.

It is important to note that there will surely be more dimensions that are salient for the analysis of competence operationalisation. For example, teacher intentions seem to be an obvious candidate dimension. Further, we think that our model with the two dimensions should not be used normatively. Different contexts may call for different operationalisation strategies and operationalisation aims. Our primary aim with the two-dimensional model is to propose a terminological and analytical framework for analysing, and talking about, a key activity in education that we feel demands more explication.

The two-dimensional model of competence operationalisation hopefully has the potential to support future research vistas into classroom assessment research. One such area could be research into teacher professional development. As argued by Andrade (2012), there is a need for research that focusses on professional development vis-à-vis developing pedagogy based on learning progressions. The two-dimensional model may be used for both conducting professional development activities and analysing teachers' professional development in, e.g. action-research projects. Based on the findings from the preceding chapters, such professional development ought to be implemented over significant periods of time with ample possibility for teachers to negotiate meanings of learning goals together with educators and other teachers.

The two-dimensional model could also support efforts to meet the need for more knowledge about whether and how teachers design instruction on the basis of the cognitive constructs that are tested for in large-scale testing systems (for an argument for this need, see McMillan 2012). In general, it must be important for the field of classroom assessment research to study the efficacy of more organised operationalisation as compared to less organised operationalisations. Studies of this kind could become an important theme in the further investigation of the potency of pedagogy based on learning progressions that many scholars call for (see, e.g. Andrade 2012).

Schneider and Andrade (2013) argued that the questions of whether 'teachers have sufficient skill to analyse student work' and of how 'teachers use evidence of student learning to adapt instruction on the intended learning target' (p. 159) are among some of the key research questions for the future of classroom assessment research. The two-dimensional model proposed here could offer an interpretive

framework for analysing observations, narratives and other data collected in order to elaborate on research questions such as these. Further, as Randel and Clark (2012) argued, there is a growing need for the development of instruments that can be used to measure teachers' assessment practices. The two-dimensional model may provide us with an outline for formulating items that pertain to the specificity and level of organisation of teachers' operationalisation of competences. In relation to this, the model may assist future research into teacher assessment preparation – a research focus that some argue needs to be systematically pursued (see, e.g. Campbell 2012).

Conclusion

In this chapter, we have pointed to several aspects that seem to cut across the multifarious research findings from the different contexts and studies involved in the ASSIST-ME project. In particular, we identified an underlying theme in the findings that pertains to how teachers interpret complex learning goals and make them more operational in order to be instructive for designing and implementing teaching and assessment activities. By beginning to talk about a process of operationalising learning goals, we hope that the fields of classroom assessment and science education will gain a more explicit nomenclature for approaching some of the perennial issues that emerge from studying teachers' assessment practice.

References

Andrade, H. L. (2012). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 17–34). Thousand Oaks: SAGE Publications.

Belova, N., Dittmar, J., Hansson, L., Hofstein, A., Nielsen, J. A., Sjöström, J., & Eilks, I. (in press). Cross-curricular goals and raising the relevance of science education. In K. Hahl, K. Juuti, J. Lampiselkä, J. Lavonen, & A. Uitto (Eds.), *Cognitive and affective aspects in science education research: Selected papers from the ESERA 2015 conference*. Rotterdam: Springer.

Bernholt, S., Rönnebeck, S., Ropohl, M., Köller, O., & Parchmann, I. (2013). Report on current state of the art in formative and summative assessment in IBE in STM-Part 1. *ASSIST-ME Report Series, 1*.

Biggs, J. B., & Collis, K. F. (2014). *Evaluating the quality of learning: The SOLO taxonomy*, *Structure of the observed learning outcome*. New York: Academic.

Brookhart, S. M. (2012). Grading. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 257–272). Thousand Oaks: SAGE Publications.

Bungum, B. (2006). Transferring and transforming technology education: A study of Norwegian teachers' perceptions of ideas from design & technology. *International Journal of Technology and Design Education, 16*(1), 31–52.

Campbell, C. (2012). Research on teacher competency in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 71–84). Thousand Oaks: SAGE Publications.

Chevallard, Y. (1991). *La transposition didactique – du savoir savant au savoir enseigné*. Grenoble: La Pensée Sauvage.

Cowie, B. (2012). Assessment in the science classroom: Priorities, practices. And prospects. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 473–488). Thousand Oaks: SAGE Publications.

Danish Ministry of Education. (2013). *Bekendtgørelse om uddannelsen til studentereksamen. Nr 776 af 26/06/2013* (Executive Order nr. 776 of 26/06/2013). Copenhagen: Danish Ministry of Education.

Desurmont, A., Forsthuber, B., & Oberheidt, S. (2008). *Levels of autonomy and responsibilities of teachers in Europe*. Eurydice. Available from: EU Bookshop.

Dolin, J. (2012). *ASSIST-ME project proposal*. Copenhagen.

Dolin, J. (2016). Idealer og realiteter i målorienteret undervisning. [Ideals and realities in goal-oriented teaching]. *Cursiv, 19*(1), 67–87.

Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*(1), 105.

Dysthe, O., Engelsen, K. S., Madsen, T., & Wittek, L. (2008). A theory-based discussion of assessment criteria – The balance between explicitness and negotiation. In A. Havnes & L. McDowell (Eds.), *Balancing dilemmas in assessment and learning in contemporary education* (pp. 121–131). New York: Routledge.

Kattmann, U., Duit, R., Gropengiesser, H., & Komorek, M. (1996). Educational reconstruction – Bringing together issues of scientific clarification and students' conceptions. *NARST, 1996*, 22.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*(4), 212–218.

McMillan, J. H. (2012). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 3–16). Thousand Oaks: SAGE Publications.

Moss, P. A. (2008). Sociocultural implications for assessment I: Classroom assessment. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 222–258). Cambridge: Cambridge University Press.

Newton, P., & Shaw, S. (2014). *Validity in educational and psychological assessment*. London: SAGE Publications.

Nielsen, J. A. (2015). Assessment of innovation competency: A thematic analysis of upper secondary school teachers' talk. *The Journal of Educational Research, 108*(4), 318–330. doi:10.1080/00220671.2014.886178.

Nielsen, J. A., & Dolin, J. (2016). Evaluering mellem mestring og præstation. [Assessment between mastery and performance]. *MONA, 2016*(1), 51–62.

Randel, B., & Clark, T. (2012). Measuring classroom assessment practices. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 145–164). Thousand Oaks: SAGE Publications.

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground –A literature review of empirical research on scientific inquiry activities. *Studies in Science Education, 52*(2), 161–197.

Ruiz-Primo, M., & Li, M. (2012). Examining formative feedback in the classroom context: New research perspectives. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 215–232). Thousand Oaks: SAGE publications.

Schneider, M. C., & Andrade, H. (2013). Teachers' and administrators' use of evidence of student learning to take action: Conclusions drawn from a special issue on formative assessment. *Applied Measurement in Education, 26*(3), 159–162. doi:10.1080/08957347.2013.793189.

Sjöström, J., & Eilks, I. (2017). Reconsidering different visions of scientific literacy and science education based on the concept of Bildung. In Y. J. Dori, Z. Mevarech, & D. Bake (Eds.), *Cognition, metacognition, and culture in STEM education*. Dordrecht: Springer.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'Translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387–420. doi:10.1177/030631289019003001.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. 1. *Assessment in Education: Principles, Policy & Practice, 14*(3), 281–294. doi:10.1080/09695940701591867.

Wiek, A., Bernstein, M., Foley, R., Cohen, M., Forrest, N., Kuzdas, C., … Keeler, L. (2015). Operationalising competencies in higher education for sustainable development. In M. Barth, G. Michelsen, M. Rieckmann, & I. Thomas (Eds.), *Handbook of higher education for sustainable development. Routledge* (pp. 241–260). London: Routledge.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716–730.

Winsløw, C. (2011). Anthropological theory of didactic phenomena: Some examples and principles of its use in the study of mathematics education. In M. Bosch (Ed.), *Un panorama de la TAD. An overview of ATD*, *CRM documents* (Vol. 10, pp. 117–138). Barcelona: Centre de Recerca Matemàtica.