

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Computers & Education

journal homepage: <http://www.elsevier.com/locate/compedu>

Multimodal digital classroom assessments

Henning Fjørtoft

Department of Teacher Education, NTNU Norwegian University of Science and Technology, Trondheim, Norway

ARTICLE INFO

Keywords:

Classroom assessment
Validity
Multimodality
Formative experiment
Literacy

ABSTRACT

Despite the widespread adoption of multimodal and digital modes of representation outside school settings, classroom assessment practices rely more on conventional print media and less on digital technologies. Stronger connections between the use of ICT in schools, contextual factors, and theoretical approaches are needed if teachers are to use digital tools effectively in the classroom. This study explores multimodal digital classroom assessments (MDCAs) as a subset of classroom assessments. Combining multimodal perspectives with performance assessment theory, the paper analyzes three examples of MDCAs developed in collaboration with practitioners as part of a formative experiment and discusses their affordances and potential relevance for practice. MDCAs may offer richer repertoires of modalities for students and teachers. However, implementing MDCAs requires continuous attention to validity, literacy demands, and management of the longitudinal nature of certain MDCAs. Therefore, to provide a meaningful picture of student learning, design processes should consider how evidence from MDCAs complements conventional assessment practices.

1. Introduction

1.1. Assessment practices in the digital and multimodal age

For more than two decades, researchers have called for broader conceptions of assessment appropriate for a multimodal digital era (Jewitt, 2003), urging educators to rethink the representational modes and literacy demands of assessment (Johnson & Kress, 2003; New London Group, 1996) and to explore students' participation, invention, and knowledge building in and beyond school settings (Greenhow, Robelia, & Hughes, 2009). The rapid dispersal of new communication technologies has fueled visions of students as active designers of meaning (Cope & Kalantzis, 2009) and posited the teacher as "a purposeful learning designer, rather than (just) a curriculum implementer" (Kalantzis & Cope, 2010, p. 205). In particular, new modes of representation have required a reconsideration of the relation between composition and design, the nature of literacy (George, 2002; Prain & Hand, 2016), and literacy and assessment (Kalantzis, Cope, & Harvey, 2003). More diverse practices, such as projects, collaborative work, and portfolios, have been suggested as viable alternatives to conventional paper-and-pen formats (Kalantzis et al., 2003) and as a stimulus for creativity (Newfield, Andrew, Stein, & Maungedzo, 2003).

However, this endeavor is challenging. Investments in ICT for education have not resulted in increased performance on PISA tests or bridged the achievement gap between different socioeconomic groups (Organisation for Economic Co-operation and Development [OECD], 2015). Merely providing students with multimodal learning resources does not necessarily lead to the use of such resources in assessment practices (Silseth & Gilje, 2017). For example, studies of digital games in second language (L2) instruction rarely consider

E-mail address: henning.fjortoft@ntnu.no.

<https://doi.org/10.1016/j.compedu.2020.103892>

Received 10 September 2019; Received in revised form 1 April 2020; Accepted 4 April 2020

Available online 8 April 2020

0360-1315/© 2020 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

contextual factors and teacher implementation (Acquah & Katz, 2020). Furthermore, there is a dearth of theory-informed research on educational technology, especially where “middle-range theories” could explain empirical findings and frame research topics to inform methodological approaches (Hew, Lan, Tang, Jia, & Lo, 2019). A review by Haßler, Major, and Hennessy (2016) suggests that schools should create opportunities for professional development, including relevant training, technical support, and policies that support teachers in integrating new technology into existing pedagogy. Effective professional development for teachers is considered more important than factors such as the student-device ratio (Haßler et al., 2016). Improving professional development through the development of elaborate instructional designs could lead to more thorough exploitation of devices in schools (Sung, Chang, & Liu, 2016). In short, stronger connections between the use of ICT in schools, contextual factors, and theoretical approaches are needed if teachers are to use digital tools effectively in the classroom.

Research on the pedagogical use of ICT is characterized by a lack of qualitative studies in general, contributions focusing on disciplinary learning in particular, and a lack of studies involving participants in designing solutions (Pérez-Sanagustín et al., 2017). The purpose of this qualitative study is to explore *multimodal digital classroom assessments* (MDCAs) as a subset of classroom assessment practices. Here, MDCAs refer to any teacher-designed assessment practices requiring students to combine two or more representational modes using digital technology. Practitioners seeking to design and implement such practices are likely to encounter a range of validity issues requiring contextually sensitive research. This study draws on two theoretical perspectives—validity in educational assessment (Crooks, 2011; Messick, 1994; Moss, 2013) and multimodality (Jewitt, 2003, 2014; Johnson & Kress, 2003; Kress, 2010; Kress & Van Leeuwen, 2001; Silseth & Gilje, 2017)—to show how teachers can design MDCAs in disciplinary contexts. MDCAs can share certain traits with performance assessments, and therefore, teachers and researchers should use validity theory to ensure that MDCAs are fit for purpose in classroom settings. To illustrate challenges in the design and use of MDCAs, three examples of MDCA practices are presented. Finally, implications of this view are discussed.

1.2. Outline of the theoretical framework

The theoretical framework for this study consists of two components. First, to understand the challenges teachers face when designing and implementing classroom assessments, research on the validity of performance assessments is drawn on. Second, multimodal theory is discussed to show how it can illuminate the role of MDCAs in the classroom. Finally, these perspectives are synthesized to elaborate on the nature of MDCAs and their role in the research design of this study.

1.3. Validity in classroom assessment

Embedding performance assessment tasks as part of classroom learning activities is considered a viable strategy for teachers seeking to integrate learning and assessment (William, 2011). However, most studies in validity theory are conducted on standardized tests (Moss, 2013), and validity research in classroom assessment contexts is characterized as having “wide gaps” (Bonner, 2013, p. 102). For example, little is known about validity in classroom assessments in academic disciplines (Xu & Brown, 2016), student use of feedback (Gamlem & Smith, 2013), the relationship between self-assessment and increased student self-efficacy (Andrade, Wang, Du, & Akawi, 2009), or reliability in teachers’ inferential decision-making (Parkes, 2013). Potential threats to validity in classroom assessments include professionals’ sense-making capacity in addressing situated problems (Moss, 2013) and the lack of skills, confidence, and understanding of validity necessary to improve classroom assessments (Black, Harrison, Hodgen, Marshall, & Serret, 2010).

Classroom assessments must also be practicable for teachers. Conceptualizing this dilemma as a one-handed clock, Stobart (2008) argues that teachers must balance validity, reliability, and manageability, and that meeting all three requirements is generally impossible. Teachers (or test designers) normally are able to satisfy only two in practice. Performance assessments are often used by teachers to navigate this dilemma. Although performance assessments are defined in many and contradictory ways (Palm, 2008), a typical characteristic is close similarity between the type of performance that is actually observed and the type of performance that is of interest (Kane, Crooks, & Cohen, 2005). However, although research on performance assessments has proliferated for more than two decades (see e.g. Aschbacher, 1991; Borko, Mayfield, Marion, Flexer, & Cumbo, 1997; Messick, 1994; Wiggins, 1998), key notions in the approach have been criticized. Messick (1994) identifies a range of issues related to the validation of performance assessments, such as the role of construct validity in their interpretation and use, the need to address construct under-representation and construct-irrelevant variance, the mix of structured exercises with open-ended tasks to manage the trade-off between issues of domain coverage and generalizability and those of time-intensive depth of examination, as well as elimination of unintended consequences through validation. Messick (1994) concludes that performance assessments “must be evaluated by the same validity criteria, both evidential and consequential, as are other assessments” (p. 13).

In discussing the validity of performance assessments, Crooks (2011) noted several problem areas: how well the evidence focuses on performance at the intended time, or on progress over the intended time period; the fairness to all students of the assessment processes; and whether the evidence coalesces into meaningful pictures of performance (p. 72). Furthermore, if performance assessments are designed and implemented by teachers, and if standards are assessed by teachers and not as part of external accountability systems, then teachers know in detail the tasks to be used and are able to prepare students for these tasks specifically. If this is known by the students, the assessment results may be high within that particular standard, while the students still lack the breadth of learning typically mandated in curricula (Crooks, 2011).

1.4. Multimodality

Multimodality is an approach to representation, communication, and interaction exploring the use of a range of semiotic (or meaning-making) resources and modes, such as image, writing, gesture, gaze, speech or posture, in products or events (Jewitt, 2014; Kress & Van Leeuwen, 2001, p. 20). Multimodal means of communicating are particularly relevant when digital devices are used due to their visual design, intertextual links to other texts, and collaborative interactivity (Beach & O'Brien, 2018; Leu, Zawilinski, & Corrigan, 2016). Thus, the advent of multimodal means of communication requires teachers to be mindful of a new "design-oriented communicative landscape" (Burke & Hammett, 2009, p. 5).

However, although the introduction of multimodal and digital approaches offers possibilities for expanding teachers' and students' repertoires, reconceptualizing teachers as designers of multimodal assessment practices remains a challenge. Among teachers, entrenched views of digital texts as primarily out-of-school media may be common, inhibiting literacy teachers' opportunities for responding to technological change (Hundley & Holbrook, 2013). Teachers may also have difficulties distinguishing between individual students' contributions or between knowledge of academic content and (mere) skill in use of digital technology (Silseth & Gilje, 2017). Furthermore, assessments of multimodal texts must relate to the material properties of digital media (e.g., opportunities for tactile interaction with the text), electronic reproduction, and genre (Davis & Yancey, 2014).

There are several approaches to analyzing multimodality, including systemic functional linguistics, social semiotics, and conversation analysis (Jewitt, Bezemer, & O'Halloran, 2016). Social semiotics aims to study how a community makes meaning using various representational resources to make meaning to create, transform, and innovate (Johnson & Kress, 2003). Social semiotics therefore analyzes textual materials and interaction in order to understand the "communicational landscape" (Jewitt et al., 2016, p. 66) of everyday situations. In this study, key concepts from research on multimodality are adopted and used selectively (see Jewitt et al., 2016, p. 5 for a distinction between "doing" and merely "adopting" research on multimodality). In particular, concepts from social semiotics such as mode and affordance (Johnson & Kress, 2003; Kress, 2010) are used to understand how the various aspects of MDCAs could be understood.

1.5. Multimodal digital classroom assessments

Multimodal perspectives on assessment have many similarities to performance assessments in that they focus on the similarity between the performance that is observed and the type of performance that is of interest. For example, asking students to collaborate on designing digital products using multiple modes shares many traits with the type of creative knowledge work associated with learning and working in a technology-rich setting (see e.g., Belo, McKenney, Voogt, & Bradley, 2016; Tondeur, van Braak, Ertmer, & Ottenbreit-Leftwich, 2017). For the purposes of this paper, MDCAs refer to any teacher-designed assessment practices requiring students to combine two or more representational modes using digital technology. This definition excludes monomodal digital performances and products (e.g., using word processor software to produce verbal texts or digital photography, or creating statistical representations from numeric data), as well as multimodal products or performances created without the use of digital technology (e.g., drama performances, printed picture books, or presentations using verbal language and gestures). This definition also excludes digital assessment practices developed by other stakeholders (e.g., digitally administered standardized tests developed by testing providers or assessments provided as part of electronic textbooks or similar externally developed learning resources). In theory, an MDCA practice could consist of any two modes of representation. For example, students could create silent movies using gestures and written language to demonstrate their learning. In such cases, each mode carries part of the informational load (see Kress, 2003, p. 20). In practice, due to the reliance on verbal language in most educational settings, MDCAs are likely to consist of a combination of linguistic means of representation (e.g., speaking or written text) and at least one other mode of representation. Consequently, attention to the literacy demands of MDCAs is likely to be important.

This study is underpinned by the assumption that MDCAs provide rich possibilities for collecting evidence of and drawing inferences about student learning processes and outcomes. Using several modes of representation in conjunction with computers or other digital devices can often add to the number of variables that teachers must manage in the classroom, thus reducing the practical value of the devices or computers and adding threats to validity in classroom assessment. For example, if the teachers' assessment designs provide a task-driven rather than a construct-driven approach to performance, then validity may be compromised (Messick, 1994). However, the "one-handed clock dilemma" posed by Stobart (2008) is a useful test for any classroom assessment, and for MDCAs specifically in that they can be evaluated based on their validity, reliability, and manageability. Therefore, a key question in designing MDCAs is how one can ensure their validity, reliability, and manageability for teachers. In addition, as teachers are likely to do much of the design work when developing MDCAs, another key question is how teachers can be involved in the design and validation of MDCAs in real-world settings. The research question guiding this study is the following: What validity issues do practitioners face when designing and using multimodal digital classroom assessments?

2. Materials and methods

This study is positioned as a formative experiment (Bradley & Reinking, 2011; Reinking & Bradley, 2008) aiming to improve classroom assessment using digital devices such as personal and tablet computers, video cameras, and mobile phones in collaboration with teachers. By emphasizing problems of practice in the intersection of literacy, technology, and classroom assessment, formative experiments have been suggested as viable approaches to studying the complexity of writing, new literacies, and communication (Leu et al., 2016), experimenting with and optimizing practices, and developing contextualized principles for future assessment practices

(Anderson & Shattuck, 2012). The qualitative descriptions provided in the findings provide insight into the subjective experiences and sensemaking processes of the participants (Morse, 2018). Although the approach used in this study does not yield generalizable findings in the quantitative sense, the findings may resonate with research conducted in similar settings, and the concepts developed can be used to further theorize the role of such practices in school settings (Twining, Heller, Nussbaum, & Tsai, 2017). Specifically, by exploring situated examples of MDCAs, the study aims to show how MDCAs can complement conventional assessments and the potential validity issues teachers may face when designing and implementing MDCAs.

The examples are considered “good examples of practice” (Kelchtermans, 2015) in that they represent situated practices showing potential ways of unleashing student and teacher creativity with emphasis on exploiting digital devices and multiple modes of representation in classroom assessment design. Such “good examples” are not meant as success stories or to impose norms for best practice; instead, they provide rich descriptions of a particular situation and the factors or conditions that determine what is going on, thus allowing others to translate the key insights into other professional contexts (Kelchtermans, 2015).

2.1. Context and participants

Teachers’ responsibilities in summative assessments vary considerably across education systems and policy contexts. For example, in Norwegian secondary schools, approximately 80% of students’ final grades are awarded by teachers based on students’ performance in the classroom. This national policy replaced compulsory exams in most subjects. The teachers collect evidence primarily through classroom observations, various teacher-constructed tests, performance assessments, and participation in classroom activities. The assessment system relies heavily on trust in the teachers’ capabilities to design and implement assessment tasks, as well as to exercise judgment in summative assessment; however, there are few guidelines for constructing and implementing this part of the assessment system (Hopfenbeck, Tolo, Florez, & El Masri, 2013; Prøitz, 2013; Prøitz & Borgen, 2010).

The substantial role of teacher choice in assessment in the Norwegian curriculum, along with the lack of shared principles for assessing and validating the designs, means that this context was suitable for a formative experiment approach. This study draws on data collected as part of a formative experiment that included participants from a university department (the author), four teachers in a Norwegian lower secondary school (students age 13–16 years), and a software engineer. The experiment lasted 16 months and was characterized by constant negotiations in meetings and informal discussions with the teachers and through support from the school principal. The teachers volunteered to participate. Teachers, the researcher, and the software engineer held informal meetings approximately every six weeks, with each meeting lasting two to 3 h. The objective was to support the teachers’ assessment design processes and explore validity issues related to the use of digital and multimodal technologies. A backward design principle framed the collaboration (Wiggins & McTighe, 2006; see also Smith, 2016). Additionally, meetings were held with the principal several times during the project period to discuss the overall progress. The teachers used iPads, video cameras, and laptop computers in the design of the MDCA practices, paying specific attention to the literacy demands of these practices.

2.2. Data collection, analysis, and representation

Data generated in this study consisted of 1) field notes from eight meetings lasting between one and 3 h, 2) written logs and email correspondence with teachers, 3) samples of student work, 4) two classroom observations, and 5) two interviews with participating teachers. Ethics approval for data collection was obtained from the Norwegian Centre for Research Data. The interviews were recorded and transcribed by the author. See Table 1 for an overview of the data collection and analysis methods.

The study started as an exploratory study, but it became clear that all participants shared an interest in developing a theoretical vocabulary to describe and critique the assessment practices created. To explore how multimodal digital classroom assessments might be understood, data from two participating teachers were selected for this study: Robert (Spanish L3 and mathematics) and Elle (Norwegian). Both teacher names are pseudonyms. The students (age 13–15) were majority language users living in the rural area surrounding the school. Mathematics and Norwegian L1 were compulsory subjects; Spanish L3 was an elective. The data collected should be considered a convenience sample. Although the project resulted in a range of experimental practices, the teachers and the researcher agreed that these examples were sustainable ways of assessing students using multimodal and digital means.

During the first stage of the analysis, data were coded using an initial coding strategy, breaking down the data into discrete parts for examination and paying particular attention to processes (e.g., designing or implementing assessment tasks) and conceptual ideas that could bring together seemingly disparate processes (e.g., validity; Saldaña, 2015, p. 117). This open-ended analysis yielded six main categories related to the design of MDCAs: content area, problem of practice, instruction and formative assessment, benefit for summative assessment, the role of students in the assessment process, and teacher professional development needs.

Table 1

Data collection and analysis methods.

Data collection methods	Data collected	Analysis methods
Participant observation	Field notes from 8 meetings	Initial coding
Participant observation	2 classroom observations	Initial coding, multimodal analysis
Document analysis	Written logs and emails from teachers	Initial coding
Document analysis	Samples of student work	Initial coding, multimodal analysis
Interviews	2 interviews (1 with Robert, 1 with Elle together)	Initial coding

The data set contained many examples of loosely developed ideas and false starts. For example, the project was hampered by several technical issues during the first months of the partnership (e.g., how to store files, how to share software licenses, or how to use unfamiliar interfaces). However, three MDCA practices were developed into fully fledged versions. These practices were selected for further analysis.

In the second stage of the analysis, data from the three MDCA practices were analyzed using concepts from social semiotics and performance assessment theory to further understand the three practices. Attention was given to data providing insight into 1) multimodal and digital aspects of these practices and 2) validity issues, such as threats to validity and their manageability in a classroom setting. This approach was chosen to analyze teacher thinking about validity and decision-making in designing and using these specific MDCAs. For example, as Elle's written reflections contained descriptions of activities targeted at understanding the learning objectives and familiarizing the students with the MDCA format she had selected, they show how she managed student learning processes with validity in mind. In the following section, three examples describing a problem of practice and the proposed MDCA design, along with the teachers' reflections on the design changes, are presented.

Teachers are unlikely to design and implement new MDCA practices in the classroom if they do not fit with their conceptions of manageable classroom activities, the purposes of assessment, and other contextual issues. Thus, the focus here is not primarily a micro-level analysis of the details in each practice but rather how a specific problem of practice can be understood through the relation between 1) how the affordances of multimodal and digital technologies affect classroom assessment practices, and 2) the validity issues related to the affordances of MDCAs (i.e., teacher-designed assessment practices requiring students to combine two or more representational modes using digital technology). Teacher reflections is a typical way of accessing their experiences and validity arguments. For this reason, I include excerpts from interviews with the teachers.

3. Results

3.1. Example A: creating mathematics videos using speaking, drawing, writing, and gesturing

Robert's students found equations difficult, and there was limited time to differentiate instruction. This problem of practice encouraged the design of an MDCA practice combining video recordings with pen-and-paper problem solving. The students created instructional videos in mixed-ability pairs, modeling and explaining their thinking while solving equations. Mid- and high-achieving students were recruited to create videos in a room adjacent to the classroom. Cameras captured the students' think-aloud processes, written explanations, and hand gestures, enabling students to combine the think-aloud activity with visual representations of handwriting and calculation, and body language (e.g., pointing at numbers on paper or screen; see Figs. 1 and 2). The rest of the class was encouraged to use the videos to gain access to the creator students' mathematical thinking, as well as to prepare at home for upcoming tests. Viewing the videos would allow students to self-assess their own mathematical reasoning by comparing it with the reasoning processes of mid- and high-achieving students.

One classroom observation focused on two students making a video based on a script Robert had provided. The purpose was to create an explanation for how to solve equation $2x - 2 = x + 5$. While filming, the students combined gestures, writing, and verbal explanations to explain their thinking for others. Although the students did not discuss the prospective audience explicitly, their conversation was replete with deictic markers such as "there," "like this," and sequential markers organizing their talk in segments (e.g., "first" and "then") while pointing to the written mathematical symbols, indicating the process of mathematical reasoning required. Deictic gestures are defined as "pointing gestures which are used to point to places in real or abstract space" (Stam & McCafferty, 2008, p. 9). The gestures in the video served two purposes: 1) to enhance the explanations by relating the mathematical notations in the script to the handwritten notes and to the deictic and sequential markers in the verbal instructions, and 2) to solve practical problems during the creation of the video. During the first attempt at filming the explanation, one student exclaimed, "You have to record my hand!"

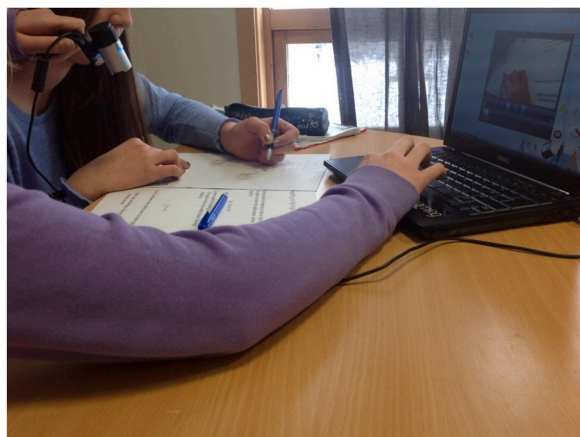


Fig. 1. Creating a mathematics video using handwriting, gestures, and talk.

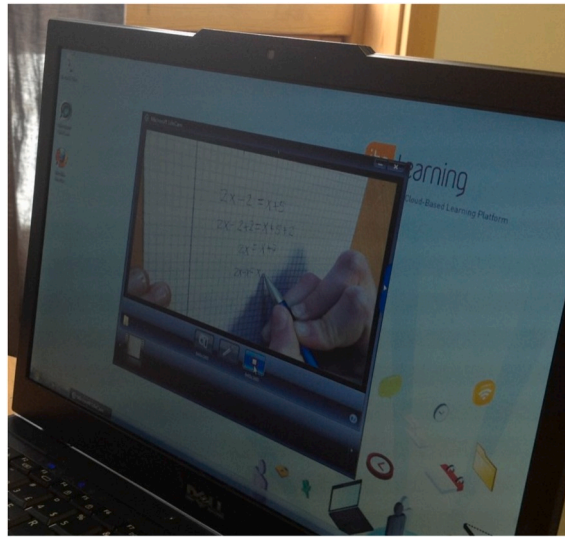


Fig. 2. Illustration of student video.

before realizing that such comments would be a distracting feature in the video. During a later attempt, one of the students used gestures off camera to discretely guide the other student back on task when her attention wandered.

Following Norris (2011), the creating of the video could be seen an example of high modal *density*: Student performance in this MDCA is represented through multiple modes (printed and handwritten notes, gesturing, handling objects, and talking). This MDCA cannot be validly assessed without attention to the *configuration* of these modes (i.e., how these modes in interaction were structured in relation to one another). For example, gestures are used to reinforce the representation of mathematical reasoning in the handwritten and verbal explanations. The deictic gestures provide a higher density of meaning-making in that they afford an interplay between the real space of the video-making scene and the abstract space of the mathematical reasoning.

On one hand, this complexity allows for rich evidence of student learning and positions students as designers of products that elicit higher-level thinking. On the other, evaluating the assessment evidence generated through this MDCA requires sophisticated lenses for analyzing the interplay of the modes, in the process of making the video and in the outcomes it represents. The gestures used for directing the process of making the movie is a second higher-level process taking place simultaneously. This process, however, does not become part of the final product. Teachers who are interested in process data in general, and evidence of student interaction in the creation of the product, therefore must be present while the video is made. In this case, Robert was not able to observe the entire process. It is likely that this MDCA is not manageable for eliciting process evidence in large classes or if teachers are not able to monitor the activity in its entirety.

Verbalizing the mathematical problem-solving strategies allowed the students to assess their own work. “When you have to talk about something for a couple of minutes, you’ll know if you understand or not,” Robert commented in the interview. However, Robert also noted that it can be challenging for students to verbalize their thinking while they are writing. In short, combining modes in real time can be challenging if the tasks are demanding. Therefore, in this case multimodality is a potential threat to validity in that performing in multiple modes represents a source of potential construct-irrelevant variance. Furthermore, disentangling the social roles, student positioning in group work, and meaningful variations in knowledge proved challenging. For example, mid-performing students were likely to spend some time off task while recording videos and approached the task with varying degrees of seriousness. Paradoxically, then, using multiple modes constituted a potential threat to validity and a strategy for minimizing construct-irrelevant variance. Disentangling construct-relevant evidence from construct-irrelevant information is therefore important when planning, creating, and using this MDCA.

3.2. Example B: My Secret Identity: longitudinal self-assessment in L3 instruction

Robert’s instruction of Spanish as a third language (L3) suffered from limited time for oral language activities and limited opportunities to document students speaking and listening. Thus, samples of student writing constituted the primary evidence used for assessment purposes. These written assignments were of little value for monitoring students’ progress in speaking and listening skills, and represented threats to validity in assessment. Moreover, given the incremental process of acquiring vocabulary and communication skills in an L3 class, students found self-assessment challenging.

To confront this problem of practice, Robert designed a longitudinal self-assessment MDCA practice called My Secret Identity. This practice required students to choose a Spanish name and record samples of their oral language development, providing him with documentation of their skill progression across a longitudinal time span. The students also recorded videos in pairs acting out their identities and documenting their ability to participate in conversations combining language use with gesturing and body language. The

digitally stored recordings of the students' language use increased individual student activity and allowed easier access for assessment purposes, for Robert and for the students. The original files with comments and reflections from the teacher and students were used as part of the assessment. Robert reviewed the samples of language use across a longitudinal time span. The students then used these samples to document and reflect on their language development. The students were also asked to review their development across this time span.

Robert evaluated this MDCA practice as having several benefits. First, it allowed multiple voices to be heard in the assessment. Original files with comments and reflections from the teacher and students provided insight into student metacognition and acted as a stimulus for dialogue and feedback. Second, data and interpretations could be aggregated and disaggregated when necessary. Third, the students were actively engaged in and responsible for the documentation of their language development, allowing them to review samples of language use across a longitudinal time span. These samples could be used in self-assessment practices to allow students to reflect on their language development. Finally, the MDCA was manageable for Robert in that the assessment evidence yielded through this practice was useful and easy to collect (see Fig. 3).

This MDCA provided an alternative to using pen-and-paper tests as a means of holding students accountable for their learning. Robert wanted to "hold students accountable by publishing," emphasizing the difference between just learning phrases and moving on through the curriculum, and being able to apply language learning through multimodal means. Students did not appear to see this perspective on accountability as demotivating or stressful. Instead, they seemed to appreciate using the videos as opportunities for recall. Reflecting on the process of using the video as part of language learning, one student wrote: "I think we learn a lot, and if we forget how to say our names, for example, then we can just look at the videos." This response suggests that this MDCA practice could provide a "soft" accountability mechanism that could avoid construct-irrelevant variance stemming from test-induced stress or low effort.

However, despite being manageable, this MDCA practice demanded reliable access to technology and a certain threshold of digital and literacy skills. This could represent a threat to construct validity in that the performances documented through MDCA are affected by construct-irrelevant variance. For example, constructing secret identities requires cultural knowledge in addition to vocabulary and speaking skills. And if data from MDCA practices are used for several purposes, students may be less willing to experiment with practices involving playing with roles and identities.

3.3. Example C: literary fiction interview videos with role-playing

The curriculum for L1 instruction in Norway defines a broad set of learning objectives but does not prescribe specific levels of attainment for these objectives. The curriculum states that students should be exposed to and practice creating a range of genres, including multimodal ones. Literacy teachers must operationalize these objectives as assessment criteria for specific literacy activities. Elle taught a Norwegian L1 class in which the students often gave individual oral presentations discussing famous novels and their authors. Although this practice is commonplace in Norwegian schools, Elle identified it as a problem of practice due to its summative, high-stakes, and time-consuming nature. Furthermore, the oral presentations provided little evidence of reading comprehension, tended to be unengaging, and being live performances, did not make evidence of student learning available for documentation purposes.

Elle designed an MDCA practice encouraging students to create videos simulating a cultural practice in which journalists interview

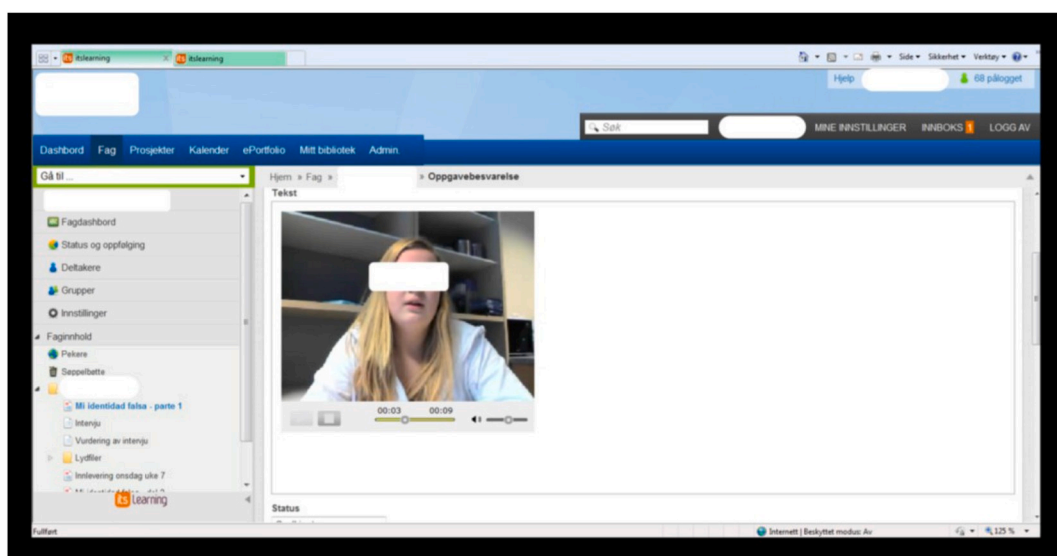


Fig. 3. Screenshot of assessment evidence collected as part of My Secret Identity on the learning management system. The menu contains assessment evidence (e.g., audio recordings of interviews and student self-assessments).

authors on stage. She created a set of assessment criteria describing her expectations for the final product based on authentic examples from mass media and merged them with disciplinary objectives. These criteria were divided into two main categories: content (e.g., biographical information about the author, appropriate use of concepts from literary history, the author's language style and preferred themes, and an analysis of a literary text) and presentation (e.g., students' ability to engage in the role, diction, and creativity in performance). Effectively, these criteria operationalized the curriculum objectives and were intended to help students navigate the expectations, thus ensuring alignment of the task and curriculum goals.

Several students were unfamiliar with the particular multimodal and digital format Elle had selected. Her written reflections illuminate her effort to build a shared understanding of the task (the genre or format of this particular style of interview) with the underlying construct (a complex amalgam of skills such as digital video production, conducting a TV interview, and using concepts from literary theory):

I introduced the lesson with a brief talk about my ideas for this activity before showing a short video clip from a TV show. Then, I distributed the assignment to the students and discussed the task. Several students were excited about the notion of creating videos and were eager to start. I provided three specific tasks for them to do during the lesson: 1) find a partner, 2) agree on an author, and 3) discuss the mode of presentation. The main task was to find as much information about the author as possible. I think it is important to give the students something concrete to relate to, a concrete assignment, even though the project itself is quite open. Some of the students were eager to start, but several others were busy talking about other things.

In the following lessons, Elle noted that more students were engaged: "creativity is bubbling (...)" but I'm unsure if all the groups understand the requirements even though they have been made clear many times," she wrote in her reflections. She also noted that giving students free rein to work often led to variations in effort among the groups. She had to remind them that the presentation was fast approaching, and that it was the responsibility of each group to finish their assignment.

The videos were recorded in the students' homes, echoing the personal atmosphere characteristic of the author interviews (see Fig. 4). The students also had the opportunity to re-record the interviews to revise and improve their performances. The final versions were stored online, making them available for assessment purposes. Implementing this MDCA practice required a combination of general classroom management skills (e.g., orchestrating and focusing a group learning activity, activating and building background knowledge) and specific moves to ensure validity (e.g., managing the challenge of an open task by dividing it into smaller segments, maintaining focus on the learning goals and assessment criteria). Elle chose to use the summative presentation as a way of holding the students accountable. Furthermore, she strove to reduce potential sources of invalidity in assessment by ensuring that the students were familiar with the task and the underlying construct.

Elle selected one video as an exemplar of how students mimicked genre traits from TV shows. For example, using body language, gaze direction, and camera angles, students showed mastery of how TV show hosts face their audiences. Furthermore, students used standardized spoken language commonly heard in TV shows but rarely used in school or casual conversation. These traits from multimodal forms of representation were integrated with curriculum elements covering the assessment criteria (e.g., using disciplinary concepts such as "contemporary author" and "literary traits" to demonstrate their familiarity with literary concepts or mentioning specific works by the author to show in-depth knowledge of her literary production). Elle interpreted this set of evidence holistically as an attempt at meeting the broadly defined content objectives in the curriculum and the assessment criteria for this specific task.

After finishing this assignment, Elle asked her students about their experiences working on this MDCA. Students commented that they appreciated using a different mode than writing or oral presentations as well as the ability to edit the videos before they are assessed:

- "It was more fun than sitting and writing, and we learned about Unni Lindell [the author]."
- "It was fun! You could be creative, and if you made a mistake, you could do it over again."



Fig. 4. Screenshot of assessment evidence collected as part of the interview assignment.

- “[It was] fun and different. And you don’t have to stand in front of the class if you are nervous.”

This suggests that the role-play component along with the opportunity to create a video offered variation in literacy instruction and a formative aspect allowing students to edit videos. In contrast, oral presentations in class do not allow for editing.

However, new challenges were encountered, particularly related to technology and literacy. Although all the students had access to cell phones for basic video recording, opportunities for editing the videos were not accessible for all students. Furthermore, Elle noted that the uploading of student videos could cause privacy issues, even when the learning management system provided by the school authorities is used. Moreover, redesigning one conventional performance assessment practice to include other modes of representation may involve new literacy skills. In this case, the change involved converting a formal speaking role (the oral presentation) into a less formal, interactive style of discourse. The students lacking cultural familiarity with the diversity of language styles and uses across genres may be at a disadvantage, resulting in potential construct-irrelevant variance and inequity in assessment. This threat was not uncovered during the design or implementation of the MDCA in this case. Nevertheless, reflecting on the practice in retrospect, the teachers noted the varying literacy demands inherent in performance assessment tasks.

4. Discussion

4.1. Design and validation of MDCAs

The ubiquity of digital devices in many educational systems allows teachers to integrate learning and assessment through designing new classroom assessment practices. MDCAs can expand teachers’ assessment practices to include a broader set of modes of representation. The examples explored in this study can add to the knowledge base of contextualized studies of classroom assessment using digital and multimodal means. Specifically, by analyzing the affordances of digital and multimodal technologies, and aligning them with curricular goals, the study shows how teachers can tailor assessment practices in the classroom and select the modes of representation most likely to provide relevant evidence of student learning, increase student engagement, and stimulate creativity.

However, the validity of such practices must be evaluated based on evidential and consequential criteria (Messick, 1994), requiring the principled use of validity theory (Brown, 2017). For example, although MDCAs may promote engagement among students, there is no guarantee that such practices will improve the quality of the assessment practices. Therefore, the validation of MDCAs should be seen as an integral part of assessment design decisions.

The research question guiding this study was *What validity issues do practitioners face when designing and using multimodal digital classroom assessments?* Three aspects of MDCAs are important for validation purposes: the multiple purposes of MDCAs, the role of MDCAs in taking a longitudinal approach to assessment, and the literacies required for successful use of MDCAs. These three aspects are important from evidential and consequential perspectives in that they foreground some aspects of student learning while downplaying others, as well as potentially impacting teachers and students in various ways.

First, evidence from a single classroom assessment practice is likely to be used for several purposes by teachers (e.g., disaggregated and used for formative purposes in feedback or aggregated and used for summative purposes in holistic evaluations). Any redesign of assessment practices should take the multiple purposes of the MDCAs into consideration. For example, if the MDCAs are to serve formative and summative ends, the teachers and others involved in their design should be prepared to defend their designs and develop separate validity arguments for each purpose. Developing guidelines for validity arguments could be helpful in this context. Such guidelines should provide support in the development and critique of validity arguments, and should discuss the relation between assessment purposes, validation procedures, and the ways in which assessment evidence is used to support decision-making. Furthermore, such guidelines should consider the affordances of various modes such as gestures, audio, or video.

Second, standardized tests and other types of single-day examinations rarely provide opportunities to represent student learning across longer time spans. If the proposed intent of the assessment is to provide a picture of student growth across time or to represent the breadth and depth of learning, high-stakes and snapshot types of assessment practices could be insufficient. A longitudinal approach using MDCAs could give teachers and students access to rich data representing the development of students’ skills and understanding across longer time spans. The opportunity to collect evidence of learning across many modes (e.g., speaking, writing, listening activities, or interactive and collaborative work) means that a broad picture of student learning can be sampled, effectively broadening construct representation. Storing recordings of language performance along with student self-assessment might provide greater awareness of the many subdomains involved in language learning (e.g., increasing vocabulary and mastery of complex grammar, or speaker positionality and contextual awareness, or participation in dialogues). Therefore, longitudinal MDCAs allow for student self-assessment in that students can access various stages of their development.

Third, the implementation of MDCA practices requires technology-specific literacies for teachers and students. As Messick (1994) suggested, evidence of the intended and unintended consequences of the interpretation and use of student performance data should be evaluated as an integral part of the validation process. Any significant change in classroom routines may cause teachers and students to shift their attention more to the medium of representation and less to the underlying construct to be learned and assessed. A lack of familiarity with digital or multimodal means of communicating could increase the time spent on technical questions and reduce the students’ actual engagement with disciplinary ideas and representations, ultimately decreasing MDCA manageability. In addition, when using new and potentially unfamiliar modes of representation, teachers may risk drifting from construct-to task-driven approaches. To minimize threats to validity, the use of less conventional modes in assessment should be accompanied by sustained awareness of the construct. Teachers and researchers should heed Crooks, Kane, and Cohen (1996) recommendation to consider the extent to which the assessment evidence coalesces into a coherent and meaningful picture of student performance, as well as the

question of whether an MDCA practice is fair to all students in the assessment process.

4.2. Limitations of the study

This study has several limitations. First, the fieldwork was conducted in a low-accountability context which allowed for considerable experimentation with assessment practices. Teachers in other contexts may not have the same opportunities for experimenting with assessment designs. Second, only three curricular areas were covered. Examples of MDCA practices from other curricular areas would enrich the knowledge base. Third, multimodal theory primarily prioritizes the practical design work of individuals over the critique of past actions or historical conventions and regulations (Kress, 2010, p. 6). Assessment practices in schools, however, are highly institutionalized and often draw on long disciplinary traditions. Consequently, their histories, epistemologies, and values should be reflected in the design and validation of MDCAs. Fourth, the focus of this study is the teachers' role as purposeful designers of MDCAs. Hence, teacher thinking about affordances and validity are emphasized over more robust measures of student learning outcomes.

4.3. Suggestions for further research and practice

The validation of MDCAs may reveal situations where reliability and validity may be compromised. The trade-off among validity, reliability, and manageability delineated in Stobart's (2008) one-handed clock are relevant to all the steps in the MDCA design and validation process. For example, understanding the implications of task-versus construct-driven approaches to validity may improve validity through ensuring that teachers collect quality evidence of student learning and base their interpretations on shared expectations of student outcomes. Documenting validation processes might require researchers to employ methodologies that capture new relationships among word, image, and sound (Kendrick, 2015).

Furthermore, assessment policies may pose limits to teachers' creativity and choices, particularly in high-stakes and low-trust contexts. Kress (2003, p. 8) has argued that the affordances of new modes require a reflection on gains and losses as well as a reflection on "human potentials, wishes, desires" beyond social and economic utility. Therefore, design processes should consider how evidence from the MDCAs complement conventional assessment practices to provide a broad and meaningful picture of student learning, and how this picture relates to the ways in which young people collaborate and represent themselves and their learning through digital means.

Credit author statement

The author of this study was responsible for all parts outlined in the credit roles.

Acknowledgements

This work was supported in part by the Department of Teacher Education at NTNU Norwegian University of Science and Technology, and Itslearning AS. The manuscript complies with APA ethical standards in the treatment of samples. The author wishes to thank Øystein Gilje for his helpful comments in the preparation of the manuscript.

References

- Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers & Education*, 143, 103667. <https://doi.org/10.1016/j.compedu.2019.103667>.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>.
- Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research*, 102(4), 287–302. <https://doi.org/10.3200/JOER.102.4.287-302>.
- Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education*, 4(4), 275–288. https://doi.org/10.1207/s15324818ame0404_2.
- Beach, R., & O'Brien, D. (2018). Significant literacy research informing English language arts instruction. In D. Lapp, & D. Fisher (Eds.), *Handbook of research on teaching the English language arts* (4th ed., pp. 1–29). Routledge.
- Belo, N., McKenney, S., Voogt, J., & Bradley, B. (2016). Teacher knowledge for using technology to foster early literacy: A literature review. *Computers in Human Behavior*, 60, 372–383. <https://doi.org/10.1016/j.chb.2016.02.053>.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215–232. <https://doi.org/10.1080/09695941003696016>.
- Bonner, S. M. (2013). Validity in classroom Assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment*. Sage Publications.
- Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, 13(3), 259–278. [https://doi.org/10.1016/S0742-051X\(96\)00024-8](https://doi.org/10.1016/S0742-051X(96)00024-8).
- Bradley, B. A., & Reinking, D. (2011). Revisiting the connection between research and practice using formative and design experiments. In N. K. Duke, & M. H. Mallette (Eds.), *Literacy research methodologies* (2nd ed., pp. 188–212). Guilford Press.
- Brown, G. T. L. (2017). The future of assessment as a human and social endeavor: Addressing the inconvenient truth of error. *Frontiers in Education*, 2, 1–4. <https://doi.org/10.3389/educ.2017.00003>.
- Burke, A., & Hammett, R. F. (2009). Introduction: Rethinking assessment from the perspective of new literacies. In A. Burke, & R. F. Hammett (Eds.), *Assessing new literacies: Perspectives from the classroom* (pp. 1–13) (Peter Lang).

- Cope, B., & Kalantzis, M. (2009). "Multiliteracies": New literacies, new learning. *Pedagogies: International Journal*, 4(3), 164–195. <https://doi.org/10.1080/15544800903076044>.
- Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies In Educational Evaluation*, 37(1), 71–77. <https://doi.org/10.1016/j.stueduc.2011.03.002>.
- Crooks, T. J., Kane, M. T., & Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice*, 3(3), 265–286. <https://doi.org/10.1080/0969594960030302>.
- Davis, M., & Yancey, K. B. (2014). Notes toward the role of materiality in composing, reviewing, and assessing multimodal texts. *Computers and Composition*, 31, 13–28. <https://doi.org/10.1016/j.compcom.2014.01.001>.
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150–169. <https://doi.org/10.1080/0969594X.2012.749212>.
- George, D. (2002). From analysis to design: Visual communication in the teaching of writing. *College Composition & Communication*, 54(1), 11–39.
- Greenhow, C., Robelia, B., & Hughes, J. E. (2009). Learning, teaching, and scholarship in a digital age: Web 2.0 and classroom research: What path should we take now? *Educational Researcher*, 38(4), 246–259. <https://doi.org/10.3102/0013189X09336671>.
- Habler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes: Tablet use in schools: A critical review. *Journal of Computer Assisted Learning*, 32(2), 139–156. <https://doi.org/10.1111/jcal.12123>.
- Hew, K. F., Lan, M., Tang, Y., Jia, C., & Lo, C. K. (2019). Where is the "theory" within the field of educational technology research? *British Journal of Educational Technology*, 50(3), 956–971. <https://doi.org/10.1111/bjet.12770>.
- Hopfenbeck, T. N., Tolo, A., Florez, T., & El Masri, Y. (2013). *Balancing trust and accountability? The assessment for learning programme in Norway—papers—OECD iLibrary*. http://www.oecd-ilibrary.org/education/balancing-trust-and-accountability-the-assessment-for-learning-programme-in-norway_5k3txnpqlsn-en.
- Hundley, M., & Holbrook, T. (2013). Set in stone or set in motion?: Multimodal and digital writing with preservice English teachers. *Journal of Adolescent & Adult Literacy*, 56(6), 500–509. <https://doi.org/10.1002/JAAL.171>.
- Jewitt, C. (2003). Re-thinking Assessment: Multimodality, literacy and computer-mediated learning. *Assessment in Education: Principles, Policy & Practice*, 10(1), 83–102. <https://doi.org/10.1080/09695940301698>.
- Jewitt, C. (2014). (Ed.). *The Routledge handbook of multimodal analysis* (2nd ed., pp. 1–7). Routledge.
- Jewitt, C., Bezemer, J. J., & O'Halloran, K. L. (2016). *Introducing multimodality*. Routledge.
- Johnson, D., & Kress, G. (2003). Globalisation, literacy and society: Redesigning pedagogy and assessment. *Assessment in education: Principles. Policy and Practice*, 10(1), 5–14. <https://doi.org/10.1080/09695940301697>.
- Kalantzis, M., & Cope, B. (2010). The Teacher as Designer: Pedagogy in the new media age. *E-Learning and Digital Media*, 7(3), 200. <https://doi.org/10.2304/elea.2010.7.3.200>.
- Kalantzis, M., Cope, B., & Harvey, A. (2003). Assessing multiliteracies and the new basics. *Assessment in education: Principles. Policy and Practice*, 10(1), 15–26. <https://doi.org/10.1080/09695940301692>.
- Kane, M., Crooks, T., & Cohen, A. (2005). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18(2), 5–17. <https://doi.org/10.1111/j.1745-3992.1999.tb00010.x>.
- Kelchtermans, G. (2015). Learning from 'good examples of practice'. *Teachers and Teaching*, 21(4), 361–365. <https://doi.org/10.1080/13540602.2014.969103>.
- Kendrick, M. (2015). The affordances and challenges of visual methodologies in literacy studies. In J. Rowsell, & K. Pahl (Eds.), *The routledge handbook of literacy studies* (pp. 619–633). Routledge.
- Kress, G. R. (2003). *Literacy in the new media age*. Routledge.
- Kress, G. R. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G. R., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Leu, D. J., Zawilinski, L., & Corrigan, J. A. (2016). Writing research through a new literacies lens. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 41–53). The Guilford Press.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Morse, J. (2018). Reframing rigor in qualitative inquiry. In N. K. Denzin, & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (5th ed., pp. 796–817). Sage.
- Moss, P. A. (2013). Validity in action: Lessons from studies of data use. *Journal of Educational Measurement*, 50(1), 91–98. <https://doi.org/10.1111/jedm.12003>.
- New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66(1), 60–93. <https://doi.org/10.17763/haer.66.1.17370n67v22j160u>.
- Newfield, D., Andrew, D., Stein, P., & Maungedzo, R. (2003). "No number can describe how good it was": Assessment issues in the multimodal classroom. *Assessment in Education: Principles, Policy & Practice*, 10(1), 61–81. <https://doi.org/10.1080/09695940301695>.
- OECD. (2015). *In Students, computers and learning: Making the connection*. OECD.
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research and Evaluation*, 13(4), 1–11.
- Parke, J. (2013). Reliability in classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment*. Sage Publications.
- Pérez-Sanagustín, M., Nussbaum, M., Hilliger, I., Alario-Hoyos, C., Heller, R. S., Twining, P., et al. (2017). Research on ICT in K-12 schools – a review of experimental and survey-based studies in computers & education 2011 to 2015. *Computers & Education*, 104, A1–A15. <https://doi.org/10.1016/j.compedu.2016.09.006>.
- Prairie, V., & Hand, B. (2016). Coming to know more through and from writing. *Educational Researcher*, 45(7), 430–434. <https://doi.org/10.3102/0013189X16672642>.
- Proitz, T. S. (2013). Variations in grading practice – subjects matter. *Education Inquiry*, 4(3), 555–575. <https://doi.org/10.3402/edui.v4i3.22629>.
- Proitz, T. S., & Borgen, J. S. (2010). *Retfærdig standpunktvurdering – det (u)muliges kunst?: Læreres setting av standpunkt karakter i fem fag i grunnskolplæringen (NIFU STEP-rapport 2010:16)*. <https://www.nifu.no/publications/934462/>.
- Reinking, D., & Bradley, B. A. (Eds.). (2008). *On formative and design experiments: Approaches to language and literacy research*. Teachers College Press.
- Saldaña, J. (2015). *The coding manual for qualitative researchers* (3E [3rd ed.]). SAGE.
- Silseth, K., & Gilje, Ø. (2017). Multimodal composition and assessment: A sociocultural perspective. *Assessment in Education: Principles, Policy & Practice*, 1–17. <https://doi.org/10.1080/0969594X.2017.1297292>.
- Smith, K. (2016). Assessment for learning: A pedagogical tool. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *The SAGE handbook of curriculum, pedagogy and assessment* (pp. 740–755). SAGE.
- Stam, G., & McCafferty, S. G. (2008). *Gesture: Second language acquisition and classroom research*. Routledge.
- Stobart, G. (2008). *Testing times. The uses and abuses of assessment*. Routledge.
- Sung, Y.-T., Chang, K.-E., & Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94, 252–275. <https://doi.org/10.1016/j.compedu.2015.11.008>.
- Tondeur, J., van Braak, J., Ertmer, P. A., & Ottenbreit-Leftwich, A. (2017). Understanding the relationship between teachers' pedagogical beliefs and technology use in education: A systematic review of qualitative evidence. *Educational Technology Research & Development*, 65(3), 555–575. <https://doi.org/10.1007/s11423-016-9481-2>.
- Twining, P., Heller, R. S., Nussbaum, M., & Tsai, C.-C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers & Education*, 106, A1–A9. <https://doi.org/10.1016/j.compedu.2016.12.002>.
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance* (1st ed.). Jossey-Bass.
- Wiggins, G. P., & McTighe, J. (2006). *Understanding by design* (Expanded 2). Pearson Prentice Hall International.
- William, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>.