ORIGINAL RESEARCH

CrossMark

Challenges for IT-Enabled Formative Assessment of Complex 21st Century Skills

Mary E. Webb[1] · Doreen Prasse[2] · Mike Phillips[3] · Djordje M. Kadijevich[4] ·
Charoula Angeli[5] · Allard Strijker[6] · Ana Amélia Carvalho[7] · Bent B. Andresen[8] ·
Eva Dobozy[9] · Hans Laugesen[10]

Abstract

In this article, we identify and examine opportunities for formative assessment provided by information technologies (IT) and the challenges which these opportunities present. We address some of these challenges by examining key aspects of assessment processes that can be facilitated by IT: datafication of learning; feedback and scaffolding; peer assessment and peer feedback. We then consider how these processes may be applied in relation to the assessment of horizontal, general complex 21st century skills (21st CS), which are still proving challenging to incorporate into curricula as well as to assess. 21st CS such as creativity, complex problem solving, communication, collaboration and self-regulated learning contain complex constructs incorporating motivational and affective components. Our analysis has enabled us to make recommendations for policy, practice and further research. While there is currently much interest in and some progress towards the development of learning/assessment analytics for assessing 21st CS, the complexity of assessing such skills, together with the need to include affective aspects means that using IT-enabled techniques will need to be combined with more traditional methods of teacher assessment as well as peer assessment for some time to come. Therefore learners, teachers and school leaders must learn how to manage the greater variety of sorts and sources of feedback including resolving tensions of inconsistent feedback from different sources.

✉ Mary E. Webb
mary.webb@kcl.ac.uk

[1] King's College London, London, UK

[2] Schwyz University of Teacher Education, Goldau, Switzerland

[3] Monash University, Clayton, Australia

[4] Institute for Educational Research, Belgrade, Serbia

[5] University of Cyprus, Nicosia, Cyprus

[6] SLO - National institute for Curriculum Development in the Netherlands, Enschede, Netherlands

[7] University of Coimbra, Coimbra, Portugal

[8] Aarhus University, Aarhus, Denmark

[9] Curtin University, Bentley, Australia

[10] National Union of Upper Secondary Teachers, Copenhagen, Denmark

🍞 Springer

1 Introduction and Background

The future of assessment faces major challenges including making the best use of information technologies (IT) to facilitate formative assessment that is important for improving learners' development, motivation and engagement in learning. The underpinning work at EDUsummIT 2017, on which this article is based, focused on the one hand on identifying the range and nature of opportunities for formative assessment provided by IT, and on the other hand on the associated challenges and evidence of how these challenges are addressed by research and current practice, and what known and as yet unresolved challenges remain.

While a variety of definitions are evident in the literature, we adopted a definition by Black and Wiliam (2009) who characterised formative assessment as the generation and interpretation of evidence about learner performance by teachers, learners or their peers to make decisions about the next steps in instruction. This definition is widely used and captures the purpose of formative assessment to support learning, the integration of assessment into instructional activities and the main players in the process. Formative assessment is often referred to as 'Assessment *for* learning' (AfL), because it views assessment as an integral part of the learning and teaching cycle, designed to support learning, allowing decisions about future performance to be better founded than decisions made in the absence of formative evidence (Black and Wiliam 2009). For our purposes, we can also incorporate computers or mobile devices along with teachers, learners and their peers into the processes for generating and interpreting evidence.

Evidence from broad-scale meta-analysis has demonstrated that formative assessment improves learning with strong effect sizes (Hattie 2009) and has led to a renewed impetus for assessment to support learning in a variety of cultural contexts (for example, see Carless and Lam 2014). Formative assessment sits in contrast to summative 'assessments *of* learning' which are used to assess student learning and make judgements, that are typically based on standards or benchmarks, at the conclusion to a learning sequence.

In addition to assessment *for* and assessment *of* learning, assessment *as* learning is a phrase that has crept into common use in education and reflects a renewed focus on the nature of the integration of assessment and learning. It highlights the importance of the dialogue between learners and teachers and between peers engaged in formative assessments. We argue that such integration can be supported and promoted by IT (Webb et al. 2013).

In many countries, in recent years, a renewed focus on assessments to support learning has been pushing against the burgeoning of testing for accountability, which in some countries, renders effective formative assessment practices almost impossible (Black 2015). Moreover, a systematic review by Harlen and Deakin Crick (2002) revealed that a strong focus on summative assessment for accountability can reduce motivation and result in the disengagement of many learners from the learning and teaching process. At the same time, use of IT-enabled assessments has been increasing rapidly, as they offer promise of cheaper ways of delivering and marking assessments as well as access to vast amounts of assessment data from which a wide range of judgements might be made about learners, teachers, schools and education systems (Gibson and Webb 2015). Current opportunities for the application of IT-enabled formative assessment, including the harnessing of data that are

being collected automatically, are underexplored and less well understood than those for summative assessments. Previously, the possibilities and challenges for IT-enabled assessments to support simultaneously both formative and summative purposes were analysed (Webb and Gibson 2015; Webb et al. 2013). Therefore, while these challenges remain, in this article we focus on the opportunities and challenges of IT supporting formative assessment, rather than summative, because effective formative assessment is recognised as being important for learning (Black and Wiliam 1998) and has tended to be under-represented in discussions of computer-based assessment. More specifically, we identify a range of challenges for using formative assessment enabled by IT (Webb et al. 2017). In this article, we address some of these challenges by examining key assessment processes that can be facilitated by IT and then considering how they may be applied in relation to the assessment of horizontal, complex and transferable 21st century skills (21st CS). As we will discuss, 21st-century skills are considered to be central to citizenship in the 21st-century (Voogt et al. 2013) but refer to complex constructs encompassing multiple components, including motivational and affective elements. Therefore, the assessment of 21st century skills is both important and challenging.

We focus first on highlighting the importance, for assessment, of affective aspects of learning and assessment because these are too often overlooked and yet are crucial for the success of the forms of assessment that we discuss throughout this article as well as being integrated into 21st CS constructs. Second, we address key aspects of assessment processes that promise to be greatly facilitated by the use of IT, such as datafication of learning, feedback and scaffolding, peer assessment and peer feedback. Third, we discuss and characterise the main challenges for using IT-enabled formative assessment and draw on recent research to examine ways of addressing them. Fourth, we focus more specifically on the challenges of assessing 21st CS, which we use as a context to examine how some of the approaches that we have identified may begin to address these challenges. Finally, we describe briefly some remaining challenges that we identified but considered that addressing them in depth was beyond the scope of this article. Thus, we aim to provide an overview of approaches to formative assessment and how they can benefit from IT as well as a specific focus on IT-enabled assessment of 21st century skills, including motivational and affective aspects.

2 Motivational and Affective Aspects Influencing Learners' Engagement with Formative Assessment

While recognition that engagement and motivation are critical for learning goes back at least as far as Dewey (1913), the importance of affective factors in accurate assessment and feedback processes have not generally been given the attention they deserve (Sadler 2010). Instead, assessment has focused predominantly on cognitive factors. Vygotsky identified the separation of cognitive and affective factors as a major weakness of traditional psychology since "it makes the thought processes appear as autonomous flow of 'thoughts thinking themselves' segregated from the fullness of life, the personal needs and interests, the inclinations and impulses of the thinker" (Vygotsky 1986, p. 10). The importance of non-cognitive factors in learning and attainment is now well recognised (see for example Khine and Areepattamannil 2016) but taking account of such factors in assessments remains a challenge. Therefore, a key recommendation for all stakeholders is to develop awareness of the importance of emotional and motivational factors in both learning and assessment.

More specifically it is important to identify, represent/visualise learners' emotional and motivational states and to use the data to inform the learning and teaching process.

A review of approaches to measuring affective learning has shown that while a range of different methods have been developed, measuring affective learning has proved to be difficult (Buissink-Smith et al. 2011). This is because affective attributes are wide-ranging and often involve complex interactions with each other and with cognitive aspects of learning. One subject where assessment of affective attributes is of obvious importance and has been developed is physical education where rubrics and checklists of both specific and holistic affective teacher assessment are available (see for example, Glennon et al. 2015). Another area where assessment of affective factors has been developed is in relation to professional behaviour of, for example, health professionals. Here learner involvement in a reflective process of assessment has been found to be a valuable part of the formative assessment process (Rogers et al. 2017). A relatively simple use of IT to facilitate this process is seen in the use of reflective blogs (see for example Olofsson et al. 2011; Wilson et al. 2016).

Current challenges for the use of IT are to develop tools addressing affective attributes that can: (1) provide information to facilitate instructional decisions; (2) support teachers in developing emotional aspects of the content they are teaching, and (3) help learners to increase their awareness. A useful first step in this regard is Harley et al.'s taxonomy of approaches and features (Harley et al. 2017). Their taxonomy is designed to support the development of complete learning systems such as intelligent tutoring systems and it highlights one of the key benefits of ongoing formative assessments during the learning process, that assessments can be modified to take account of learners' emotional responses and current state. There remain major challenges for the design and development, across all content domains, of rubrics that take account of affective aspects. Furthermore, as we will discuss later, crucial affective aspects are critical for, and integrated into constructs of 21st CS.

3 Assessment Processes Enabled by IT

A range of different types of IT can support a wide variety of processes involved in assessment. We have identified key aspects that show particular promise in relation to making use of IT as: datafication of learning processes; feedback and scaffolding; peer assessment and peer feedback. Although none of these processes are particularly new, they all can be supported by recent developments in IT and are also important for effective formative assessment.

3.1 Datafication of Learning Processes

In this article, we focus on the value of datafication for formative assessment, i.e., how to collect data, interpret/analyse and use that meaningful information to support teachers and learners in the process of learning. This includes data that are immediately processed and presented as part of the interactive learning processes as well as data analysed in the background and available for future analysis, e.g., "stealth assessment" (Shute 2011) or "quiet assessment" where learners and teachers are able to "turn up the volume" whenever they wish, in order to review progress (Webb et al. 2013).

In the earlier stages of research into datafication of education, "learning analytics" (LA) was of limited use because there was little focus on assessment purposes (Ellis 2013) so

Ellis proposed the need for "assessment analytics". More recently the theory and practical elements of analytics have been further developed towards their use for assessment purposes but developments for formative assessment purposes are still in their relatively early stages (Ifenthaler et al. 2018). Some learning contexts lend themselves to the use of assessment analytics: for example Fidalgo-Blanco et al. (2015) developed a LA system to examine the performance of individuals in teamwork where they were interacting online through discussion boards. Their study focused on the value for teachers of obtaining timely information about interactions and progress and therefore being able to improve their teaching decisions. Another study enabled learners themselves to access data about their learning and hence develop their self-regulation (Tempelaar et al. 2013). This study focused particularly on emotional factors, whose importance we discussed earlier. Thus, the current literature points the way towards making use of LA for formative purposes to support learners and teachers and the two examples mentioned here are particularly relevant for 21st CS as discussed later in this article. However, there is an ongoing need for research, in multidisciplinary groups, across different subject areas and modes of learning on how to collect, analyse and represent data, in such a way that it is useful to learners and teachers.

Learning data are usually visualised and analysed by using dashboards that can present the data in a variety of different ways (see Verbert et al. 2014 for a review). Although the use of dashboards can support all stages of learning, and the analytics processes, e.g., awareness, (self-) reflection, sense making, and impact, have considerable potential to improve learning, it is not yet clear to what extent the use of a dashboard would result in behavioural changes or new understanding, because research on this impact is still limited (Verbert et al. 2013). A recent empirical study by Kim and colleagues illustrates the complexity of dashboard design (Kim et al. 2016). The research found that students who used dashboards in their online learning had higher final scores than those who did not, however, the dashboard usage frequency did not influence their learning achievement, because more capable students tended only to access the dashboard on one occasion. Moreover, the Kim et al. (2016) study identified a range of factors that need to be further researched in relation to dashboard design including motivational elements for different types of learners, gender effects and how to match presentations to learners' current needs.

3.2 Forms of Feedback and Scaffolds for Teachers and Learners to Make Sense of Data

In the meta-analysis by Hattie (2009), feedback was found to have one of the highest effects on student learning of all learning interventions. However, the value of feedback depends on the type of feedback and how it is delivered. Despite a renewed focus on providing detailed feedback, many learners failed to make use of the feedback because they had insufficient background knowledge to make sense of it (Sadler 2010). Sadler's analysis suggested that it was not only necessary to pay attention to the technical aspects of the feedback in terms of the knowledge content and relevance but also to take account of learners' emotional responses, as discussed earlier in this article. In order to take account of these affective factors we suggest the need for ongoing dialogue involving teachers, learners and system designers in the process of creating systems that can be adaptive to contextual sensitivities. Artificial intelligence techniques are supporting the development of adaptive systems. For example Chen (2014), in an experimental study of 170 eighth grade students,

found that an adaptive scaffolding system, that addressed both cognitive and motivational aspects of learning, promoted the learning of velocity and acceleration.

In line with Sadler's earlier work in higher education, in a review of recent developments in computer-based formative assessment for learners in primary and secondary education, Shute and Rahimi (2017) concluded that a key challenge was to design feedback that learners actually use. Furthermore, in order to encourage learners to use the feedback, evidence from many recent studies have confirmed the need to deliver feedback in manageable units and to use "elaborated feedback", which includes explanations, rather than simple verification of whether or not the answer was right (Shute and Rahimi 2017).

Developments in IT have led to additional challenges through the availability of more different sorts and sources of feedback, including automatic feedback systems (see Whitelock and Bektik 2018 for a review). Thus, feedback can come from humans or be processed from data. Therefore, learners, teachers and school leaders have to learn how to manage this greater variety of sorts and sources including resolving tensions of inconsistent feedback from different sources. In order to facilitate this additional aspect of assessment literacy, we believe that it is important, in designing feedback systems, to give teachers and learners access to the data collection and processing model in addition to the final data state, using appropriate visualisation techniques as discussed earlier, to better understand the formative elements of the tasks.

3.3 Peer Assessment and Peer Feedback

Peer assessment and peer feedback are playing an increasingly important role in education so a key issue concerns the extent to which peer assessment and feedback can replace or complement teacher assessment (Erstad and Voogt 2018). In the context of higher education, Liu and Carless (2006) argued for a focus on peer feedback, as a formative process that enables learners to take an active role in the management of their own learning, rather than peer assessment using grades. In school education, Black and Wiliam (2009) identified a key role for peer assessment and feedback as a precursor to self-assessment and development of self- regulated learning. In relation to using IT, Van der Kleij and Adie (2018), based on a review of recent research, found that IT can support peer feedback processes in various ways, including: the use of social networking platforms for feedback discussions about homework; collaborative writing using online word processing and video. However, they concluded that it is how teachers support learners in developing their capability to assess the work of their peers, rather than the particular IT-use, that determines the effectiveness of the feedback. Learners need to have a good understanding of the assessment criteria and the assignment task. These prerequisites vary, in their challenge, according to the subject area and learning focus. Thus, for example, in creative writing, while it is relatively easy to enable young learners to identify key features in writing, such as alliteration or powerful adjectives, enabling them to assess the overall quality of the piece of writing is much harder to explain and model. Likewise, in learning science, a checklist approach for assessment criteria can easily be developed for enabling students to comment on the clarity and completeness of an explanation of a science experiment, but it is much harder, for example, to enable students to assess the quality of argument in an exploration of an environmental dilemma. These examples illustrate both the challenge of enabling peer feedback and its value because, in order to utilise peer feedback effectively, teachers need clear learning objectives, a scheme of work showing progression in development of understanding and ways of making assessment criteria accessible to the students. These

pedagogical principles not only support peer assessment but their use would also help students to understand what is required for their own learning.

In addition to developing learners' content knowledge for peer assessment, there are emotional, social and cultural issues for consideration. For example, learners may not accept peer feedback as accurate or they may feel uncomfortable in assessing their peers or be unwilling to take responsibility (Carvalho 2010; Topping 1998). Thus, key challenges for enabling effective peer feedback include: establishing a safe environment in which learners feel comfortable and confident in their assessment capabilities; promoting, managing, timing and designing peer assessment and managing learners' expectations.

Some research is beginning to point towards ways in which IT can address some of these issues. For example, in order to help a learner select an appropriate helper, an online peer assessment tool may provide him/her some information about social context (e.g., willingness to help) and knowledge context (e.g., achievement level) for each helper candidate. In the event of an incorrect answer, a learner can see the list of candidates for help, choose one of them based on that information, and send him/her a message with a request for help (e.g., about the correct answer and the reasoning applied). By implementing this approach, Lin and Lai (2013) found that compared to the traditional formative assessment, this approach resulted in better learning achievements, probably because of a high response rate of requests for help. Note that learners with higher centrality (i.e., social network position) were more likely to ask for help from peers, and then they themselves would gradually taking over the role of target helpers for these peers.

4 Assessing Horizontal, Complex 21st Century Skills

21st CS is often used as an umbrella term to describe a wide range of competencies, "habits of mind", and attributes considered central to citizenship in the 21st century (Voogt et al. 2013). Theoretical constructs commonly employed and studied under this perspective are, for instance, creativity and critical thinking, complex problem solving, communication and collaboration, self-regulation, computer and information literacy (e.g. see Geisinger 2016; Griffin and Care 2014 for conceptual clarification). These 21st CS are considered to be of growing importance in the context of current and future societal challenges, in particular with regard to the changing nature of learning and work driven by the evolution of the job market under the influence of automatization, digitalization and globalization. The discussion around 21st CS also emphasises competencies which enable responsible action when faced with complex and often unpredictable problems of societal relevance. Increasingly, this shift of focus towards complex competencies relating to authentic, complex problems is also being called for in current psychological research on problem solving, where the past emphasis on primarily cognitive and rational forms of learning is being criticised (compare Dörner and Funke 2017). Here, we discuss first the challenges of clarifying the constructs of 21st century learning in order to consider how they might be measured. Then, we examine how IT may enable such assessment.

4.1 Challenges of Clarifying Constructs for Assessment of 21st-Century Skills

21st CS are complex constructs comprising multiple elements, attitudes, behaviours or modes of action and thought that are transferable across situations and contexts. Many of these constructs lack a sharp definition and/or display varying degrees of theoretical

overlap in their definitions or the meaning of their sub-concepts (Shute et al. 2016). To give an example, the construct Collaborative Problem Solving (CoIPS), described in Care et al. (2016), includes sub-concepts (e.g., critical thinking) which also appear in other constructs such as creativity (Lucas 2016). Certain skills defined in the construct "Computer and Information Literacy", for example "evaluating information" (Ainley et al. 2016) form a part of the CoIPS construct, and so forth. This overlap on the level of theoretical constructs becomes even more pronounced on the level of concept operationalisation in the shape of certain behavioural, cognitive and emotional patterns (Shute et al. 2016).Many of the 21st CS constructs, such as collaborative and complex problem solving and computer and information literacy, have recently been studied more closely in comprehensive research projects such as those associated with the PISA surveys and the "Assessment and Teaching of 21st Century Skills" (ATC21) Project (for example see Griffin and Care 2014). However, incorporation into curricula and integration of formative and summative assessment practices of 21st CS in schools often lags behind (Erstad and Voogt 2018). On the one hand, typical barriers might be attributed to certain social and educational policies, such as the traditional organisation of the curriculum by subjects or accountability structures which prioritise typical indicators of academic success, such as mathematics, science, or language literacy. On the other hand, the complexity of 21st CS constructs presents another significant challenge to their assessment, which can only insufficiently be addressed by the classic repertoire of methods, e.g., multiple-choice questions or self-report measures (Ercikan and Oliveri 2016; Shute and Rahimi 2017). Furthermore, 21st CS contain an assortment of diverse, but interconnected skills and competencies, which are latent and thus not directly measurable constructs. Therefore, we argue that they must first be linked to specific complex and context-dependent, and therefore possibly dynamic, behavioural patterns via a theoretical model. If, for example, the aim is to assess the quality of collaboration in a group, a number of questions arise: What would constitute a good measure? The quality of the end-product, the creativity of the solution or the satisfaction of the team members with the social interactions in that group? Normative considerations enter the equation here as well. Furthermore, how do different patterns of learning activities relate to a (latent) trait, e.g., creativity? And how stable are these patterns with regard to different types of problems, or social/cultural contexts of the learning situation? The translation of theoretical (and normative) considerations into an adequate measurement model and the derivation of meaningful interpretations of learners' performances which then enable possible adjustments of learning processes is not only important for summative measurement. When making use of the new possibilities for tracking and analysis of learning activities in digital environments it is crucial to explicitly state and theoretically justify ascriptions of meaning and possibilities for interpretation when analysing this data in the context of formative assessment.

4.2 New Opportunities Provided by IT for Assessing 21st Century Skills

Considering the challenges for formative assessment of 21st CS that go hand in hand with the endeavour to capture, visualise and feedback these complex cognitive, emotional and behavioural patterns, IT-based developments create high hopes for new opportunities (Shute and Rahimi 2017; Webb and Gibson 2015). An example would be the assessment of multidimensional learner characteristics, such as cognitive, metacognitive and affective, using authentic digital tasks, such as games and simulations (Shute and Rahimi 2017). Working in digital learning environments also brings with it a set of expanded possibilities

with respect to documentation and analysis of large and highly complex sets of data on learning processes, including log-file and multichannel data, in varying learning scenarios (Ifenthaler et al. 2017). For example, the retrieval of the time dimension, the context, and the sequence of occurrence of different behaviours, which could also involve the use of certain strategies, the posting of certain comments or the retrieval of specific learning content at given times in the problem-solving process, allow for the digital analysis of these "traces of learning" through sequence analysis or social network analysis. Furthermore, behavioural patterns of interest can be combined with data derived through more "traditional" methods, such as test-scores for digital literacy, self-report measures for motivation, self-efficacy, personality or information obtained from data in open language-based formats, e.g., reflective thoughts in chats, blogs or essays, which can be put through digitally assisted analysis e.g., natural language processing.

Some of the current research on digitally assisted assessment explicitly focuses on the "theory-driven measurement" of 21st CS. Examples are recently designed tests for collaborative problem solving (Herde and Greiff 2016), complex problem solving (Greiff et al. 2016) or ICT-literacy (Ainley et al. 2016; Siddiq et al. 2017). In tests for collaborative problem solving, as developed in the international project, ATC21S (Griffin and Care 2014), as well as in the PISA assessments (Herde et al. 2016), learners interact with peers (ATC21S) or an intelligent virtual agent (PISA) to solve problems of varying complexity. These assessments use (more or less) controlled digital learning scenarios for capturing and analysing a variety of behavioural process data in order to create indicators which form scale values, competence levels or prototypes. A game-based example is the learning environment "Use Your Brainz", where four areas of problem-solving competence can be assessed: analysing, planning, using tools and resources, monitoring and evaluating (Shute et al. 2016). The development of these tests provides a good illustration of the complexity of the design process, starting with theory-based modelling of analytic categories, the development of a learning environment in which the heterogeneous data sources can be captured, and the design of supportive tools for automated analysis and feedback. Feedback, in these test environments, is usually designed for teachers, researchers or other stakeholders in educational administration, who can identify areas of development for learners or classrooms. The challenge remains to identify the types of information and the feedback format that will provide effective learning impulses directly to learners, as discussed earlier.

In addition to the body of research focusing on theory-driven measurement, other studies take what might be characterised as a more "data-driven" approach. Here, the new possibilities for continuous "quiet" capture and analysis of rich process data in digital learning environments, such as learning management systems, blogs, wikis etc., can be used to explore and identify behavioural patterns in relation to 21st CS. For example, specific performance outcomes may be measured, or certain learning patterns or "error patterns" may be correlated with a large number of other user data, to allow predictions regarding effective next steps towards obtaining specific skills, such as critical thinking. Greif et al. (2016), for instance, analysed log-files of performance data from a computer-based assessment of complex problem solving using the "MicroDYN approach". They found certain behavioural patterns were associated with better performance. Similarly, the identification of particular decision patterns occurring during a digital game can be typical of pupils with differing creative thinking skills. In addition, with regard to automated assessments of collaborative processes, the knowledge contributions and interaction patterns of different learners can be analysed in real time and compared with ideal/typical interaction patterns in order to derive recommendations for the use of effective cooperation strategies for

learners or for effective group compositions for teamwork (Berland et al. 2015; Fidalgo-Blanco et al. 2015). Going beyond the data analysis process to provide a tool to enable learners to engage in peer support, Lin and Lai (2013) used Social Network Analysis, as discussed earlier.

In both the theory- and the data-driven approach, the focus is often on the identification of meaningful information from which recommendations for the next steps of the learning process can be derived. Although these steps are not always fully automated, the results of the data analysis guide and structure the decisions of learners and teachers to a large extent. If, instead, one focuses on the processing of data by the learners themselves, real-time feedback can be seen as the trigger for self-regulating, cognitive and metacognitive learning processes and thus contributes to the development of competences in this area. Generally speaking, this applies to most 21st CS, which in some form all include reflexive, metacognitive processes, whether it is about adopting differing perspectives in collaborative problem solving, weighing up diverse lines of argumentation and reflecting one's personal attitudes in critical thinking or the use of particular problem-solving heuristics in creative thinking. Research here focuses on the development of tools for the visualization and presentation of data for learners and for pedagogical scaffolding of learning processes in order to initiate effective cognitive and metacognitive processes. Computer-based assessments for learning which include such tools, e.g., articulating the rationale for making a specific response, stating confidence, adding and recommending answer notes may support the development of self-regulated learning skills (see for example Chen 2014; Mahroeian and Chin 2013; Marzouk et al. 2016). In the context of self-regulated learning, the question of what kind of feedback will actually engage individual learners and motivate them to become self-regulated learners becomes critical (Tempelaar et al. 2013). In summary, research challenges for self-regulated learning in the context of assessment of 21st-century learning include:

- Development of tools for creating automated knowledge/concept visualization for individuals or groups, e.g., concept maps from written text or other formats, which can be compared to expert maps/reference models (see for example, Ifenthaler 2014) to indicate issues for the learner to consider, e.g., cohesion gaps in critical writing (Lachner et al. 2017).
- Applications for Social Network Analysis for visualisation and analysis of learner status, e.g., knowledge/expertise/willingness to help, and learner interaction, which might potentially promote important preconditions for effective teamwork, such as transactive memory and perspective taking. Indicators such as social distance and centrality in a network can be used to visualise collaborative efforts in groups, which might help to develop self-regulated learning strategies such as managing resources, and seeking support (e.g., Lin and Lai 2013).
- Analysis of learners' free-text responses via natural language processing techniques to automatically detect rhetorical patterns or indicators which can be interpreted in terms of reflective or creative thinking and support formative assessment of these skills (see for example Bektik 2017; Rodrigues and Oliveira 2014).
- Research on appropriate quantity, complexity and timing of metacognitive prompts in different user/age groups to tackle problems associated with cognitive load or motivation. Studies have for instance shown that metacognitive scaffolding (e.g., reflection prompt, goal setting) can also have negative impacts on intrinsic motivation and self-concept (Förster and Souvignier 2014; Maier et al. 2016).

- Research on pedagogical virtual agents tutoring learners on how to organise and guide their own learning contexts (Johnson and Lester 2016).
- Automated analysis and integrated visualization of rich e-portfolio data for the development of reflective and critical thinking capabilities.

Learning analytics and educational data mining generate high hopes for a renewed focus on formative assessment of 21st CS (Gibson and Webb 2015; Spector et al. 2016). In addition, a continuous unobtrusive background measurement of performance (Shute 2011; Webb et al. 2013) enables minimal disruption of learning processes and immediate feedback, which is very important for the automation and routinisation of self-regulatory learning strategies. Furthermore, progress with automated, real-time natural language processing opens new possibilities in the area of reflective and critical thinking. However, meaningful analysis of the data collected is often very difficult and requires strong theoretical grounding and modelling as well as verification of validity, gained for instance in complex evidence-based design processes. Due to the complexity of the 21st CS constructs, validity of detected behavioural patterns should be investigated in a comprehensive manner, i.e., not only via correlations with certain outcome measures, but also by identification of causal chains that lead to such outcomes. Case studies using think-aloud protocols might be a promising approach here (Siddiq and Scherer 2017). With regard to validity issues of complex and collaborative problem solving, formative assessment of 21st CS should aim to address authentic and complex learning opportunities and, when possible, not limit itself to "simpler" problems for ease of measurement (Dörner and Funke 2017). In this context, game environments and virtual worlds have a great potential for development, but require a concerted interdisciplinary effort by a variety of stakeholder groups.

5 Conclusion and Recommendations

In this article, we examined some of the key challenges for formative assessment and ways of addressing them especially in relation to 21st CS. More specifically, we highlighted the importance of affective aspects of learning and assessment, which we argued, are particularly important for 21st CS such as creativity, complex problem solving, communication, collaboration and self-regulated learning. We focused particularly on opportunities and issues associated with learning/assessment analytics; feedback and scaffolding; peer assessment and peer feedback. Regarding the challenges of assessing horizontal, general complex 21st CS, we identified some developments in ways of assessing these skills and competences especially concerning datafication of learning processes and the use of analytics. While there is currently much interest and research in developing learning/assessment analytics for assessing 21st CS, it is highly likely that the complexity of assessing such skills, together with the need to include affective aspects will mean that using IT-enabled techniques will need to be combined with more traditional methods of teacher assessment as well as peer assessment for some time to come. Therefore, learners, teachers and school leaders have to learn how to manage the greater variety of sorts and sources of feedback including resolving tensions of inconsistent feedback from different sources. In order to facilitate this additional aspect of assessment literacy, we believe that it is important, in designing feedback systems, to give teachers and learners access to the data collection and processing model in addition to the final data state, using appropriate visualisation techniques as discussed earlier, to better understand the formative elements of the tasks.

Particularly with regard to 21st CS, a significant challenge for the design of formative assessment is to find a good balance between automated assessment, i.e., highest possible adaptivity to the individual characteristics of the learner, and the active and "constructivist" role of learner and teacher. When making use of feedback information, a highly restricted space for interpretation and decision-making could be counterproductive with regard to the learning benefits concerning 21st CS. In this context, it is relevant to also consider potential unintended consequences of feedback, for instance with respect to the learners' experience of autonomy and competence (Sadler 2010). Here the value of peer feedback, as discussed earlier, needs to be considered as an alternative or in conjunction with automated assessment and assessment analytics.

Learners and teachers should be supported regarding interpretation of the data, the associated learning decisions and the possible consequences. It is important to design pedagogical and technological scaffolds for learners and teachers that are directly integrated into the tools applied. Furthermore, the significance of formative assessment of 21st CS should be reflected in educational standards and curricula, as this influences the investment of resources in this area.

In addition, to make formative assessment of 21st CS effective, teachers and learners require a high degree of assessment literacy as discussed earlier (Erstad and Voogt 2018). To complicate matters, their interpretations and decisions are also influenced by their beliefs regarding the value of formative assessment and the achievement of 21st CS (e.g., Valtonen et al. 2017). Research looking at the interplay between knowledge, beliefs and the implementation of assessment practices might identify hindering factors, which could for instance be addressed in the context of teacher training.

In order to provide a comprehensive overview of formative assessment practices in relation to developments in IT-enabled assessment, we must mention privacy and ethical issues, which although not addressed in this article are of critical importance if the use of digital data in assessments is to serve learners well. Learners and teachers leave "digital traces", but are, at the same time, often not aware of the possible consequences of their digital activities. The same is true for different groups of stakeholders, who enable the collection and use of different types of data at different levels of the educational system (learner, teacher and classroom, school, etc.). Therefore, important questions must be addressed regarding who has access to the data, and for which purpose. For instance, companies might control access to certain data, and governing institutions of an educational system may use information for steering purposes. As Breiter and Hepp (2018) point out, the digital data generated and the digital traces left behind are not 'neutral phenomena' which reflect the natural behavior of learners, but rely on the technical and analytical procedures of the researcher, administrations, and companies that produce, shape and use the data. Therefore, schools need to be careful when arranging contracts with providers of digital learning materials to ensure that the data belong to the school and are used only for dialogue between teachers and their students.

The complex role of IT in the provision of formative feedback is highly situated and can be shaped by numerous micro-, miso-and macro-contextual factors. As such, this research stream requires ongoing research to ensure that high quality, usable information is provided to teachers and learners. In summary, our recommendations for major research challenges to be addressed include multidisciplinary investigation into assessment analytics in order to determine how to make data available in useful ways for learners and teachers and research into how to support the development of self-regulated learning of 21st CS. For teachers and learners, we recommend increasing awareness of the importance of emotional aspects of learning and assessment, in particular in relation to complex 21st CS. Furthermore, we

suggest that it is important for teachers and learners to expect to be able to participate fully in not only the assessment processes but the design of new assessments and therefore developing assessment literacy is crucial. For policymakers, awareness of the importance for learning of formative assessment and how it can be supported by IT, may enable a move away from the strong focus on summative assessment to supporting the development of formative assessment. Furthermore, addressing the need for curricula to represent fully 21st CS and for these to be adequately assessed should be a priority.

References

Ainley, J., Fraillon, J., Schulz, W., & Gebhardt, E. (2016). Conceptualizing and measuring computer and information literacy in cross-national contexts. *Applied Measurement in Education, 29*(4), 291–309. https://doi.org/10.1080/08957347.2016.1209205.

Bektik, D. (2017). Learning analytics for academic writing through automatic identification of meta-discourse. The Open University.

Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 425–447. https://doi.org/10.1007/s11412-015-9217-z.

Black, P. (2015). Formative assessment—an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice, 22*(1), 161–177. https://doi.org/10.1080/0969594x.2014.999643.

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

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

Breiter, A., & Hepp, A. (2018). the complexity of datafication: putting digital traces in context. In *Communicative figurations* (pp. 387–405). Berlin: Springer.

Buissink-Smith, N., Mann, S., & Shephard, K. (2011). How do we measure affective learning in higher education? *Journal of Education for Sustainable Development, 5*(1), 101–114. https://doi.org/10.1177/097340821000500113.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education, 29*(4), 250–264. https://doi.org/10.1080/08957347.2016.1209204.

Carless, D., & Lam, R. (2014). Developing assessment for productive learning in confucian-influenced settings. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 167–179). Dordrecht: Springer.

Carvalho, A. (2010). Revisão por pares no ensino universitário: desenvolvimento da capacidade de criticar construtivamente. *Transformar a pedagogia universitária–Narrativas da prática*, 175–198.

Chen, C.-H. (2014). An adaptive scaffolding e-learning system for middle school students' physics learning. *2014, 30*(3). https://doi.org/10.14742/ajet.430

Dewey, J. (1913). *Interest and effort in education*. Houghton Mifflin.

Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology, 8,* 1153. https://doi.org/10.3389/fpsyg.2017.01153.

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662–664. https://doi.org/10.1111/bjet.12028.

Ercikan, K., & Oliveri, M. E. (2016). In search of validity evidence in support of the interpretation and use of assessments of complex constructs: discussion of research on assessing 21st century skills. *Applied Measurement in Education, 29*(4), 310–318. https://doi.org/10.1080/08957347.2016.1209210.

Erstad, O., & Voogt, J. (2018). The twenty-first century curriculum: issues and challenges. In J. Voogt, G. Knezek, K. Wing, & R. Christensen (Eds.), *International handbook of IT in primary and secondary education* (2nd ed.). Berlin: Springer.

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015). Using learning analytics to improve teamwork assessment. *Computers in Human Behavior, 47,* 149–156. https://doi.org/10.1016/j.chb.2014.11.050.

Förster, N., & Souvignier, E. (2014). Learning progress assessment and goal setting: Effects on reading achievement, reading motivation and reading self-concept. *Learning and Instruction, 32,* 91–100. https://doi.org/10.1016/j.learninstruc.2014.02.002.

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them? *Applied Measurement in Education, 29*(4), 245–249. https://doi.org/10.1080/08957347.2016.1209207.

Gibson, D. C., & Webb, M. E. (2015). Data science in educational assessment. *Education and Information Technologies, 20*(4), 697–713. https://doi.org/10.1007/s10639-015-9411-7.

Glennon, W., Hart, A., & Foley, J. T. (2015). Developing effective affective assessment practices. *Journal of Physical Education, Recreation & Dance, 86*(6), 40–44.

Greiff, S., Niepel, C., Scherer, R., & Martin, R. (2016). Understanding students' performance in a computer-based assessment of complex problem solving: An analysis of behavioral data from computer-generated log files. *Computers in Human Behavior, 61,* 36–46. https://doi.org/10.1016/j.chb.2016.02.095.

Griffin, P., & Care, E. (2014). *Assessment and teaching of 21st century skills: Methods and approach.* Berlin: Springer.

Harlen, W., & Deakin Crick, R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning.* Retrieved from London: http://eppi.ioe.ac.uk/cms/Default.aspx?tabid=108

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education, 27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Abingdon: Routledge.

Herde, C. N., Wüstenberg, S., & Greiff, S. (2016). Assessment of complex problem solving: What we know and what we don't know. *Applied Measurement in Education, 29*(4), 265–277. https://doi.org/10.1080/08957347.2016.1209208.

Ifenthaler, D. (2014). Toward automated computer-based visualization and assessment of team-based performance. *Journal of Educational Psychology, 106*(3), 651.

Ifenthaler, D., Gibson, D. & Dobozy, E. (2017). The synergistic and dynamic relationship between learning design and learning analytics. In H. Partridge, K. Davis, & J. Thomas. (Eds.), *Me, Us, IT! Proceedings ASCILITE2017: 34th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education* (pp. 112–116).

Ifenthaler, D., Greiff, S., & Gibson, D. (2018). Making use of data for assessments: Harnessing analytics and data science. In J. Voogt, G. Knezek, & K. Wing (Eds.), *International handbook of IT in primary and secondary education (2nd ed.)* (pp. 191–198). Berlin: Springer.

Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education, 26*(1), 25–36. https://doi.org/10.1007/s40593-015-0065-9.

Khine, M. S., & Areepattamannil, S. (2016). *Non-cognitive skills and factors in educational attainment.* Berlin: Springer.

Kim, J., Jo, I.-H., & Park, Y. (2016). Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review, 17*(1), 13–24. https://doi.org/10.1007/s12564-015-9403-8.

Lachner, A., Burkhart, C., & Nückles, M. (2017). Formative computer-based feedback in the university classroom: Specific concept maps scaffold students' writing. *Computers in Human Behavior, 72,* 459–469. https://doi.org/10.1016/j.chb.2017.03.008.

Lin, J.-W., & Lai, Y.-C. (2013). Online formative assessments with social network awareness. *Computers & Education, 66,* 40–53. https://doi.org/10.1016/j.compedu.2013.02.008.

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290. https://doi.org/10.1080/13562510600680582.

Lucas, B. (2016). A five-dimensional model of creativity and its assessment in schools. *Applied Measurement in Education, 29*(4), 278–290. https://doi.org/10.1080/08957347.2016.1209206.

Mahroeian, H., & Chin, W. M. (2013, 15–18 July 2013). *An analysis of web-based formative assessment systems used in E-learning environment.* Paper presented at the 2013 IEEE 13th International Conference on Advanced Learning Technologies.

Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education, 95,* 85–98. https://doi.org/10.1016/j.compedu.2015.12.002.

Marzouk, Z., Rakovic, M., & Winne, P. H. (2016). *Generating learning analytics to improve learners' meta-cognitive skills Using nStudy trace data and the ICAP framework.* Paper presented at the LAL@ LAK.

Olofsson, A. D., Lindberg, O. J., & Hauge, E. T. (2011). Blogs and the design of reflective peer-to-peer technology-enhanced learning and formative assessment. *Campus-Wide Information Systems, 28*(3), 183–194. https://doi.org/10.1108/10650741111145715.

Rodrigues, F., & Oliveira, P. (2014). A system for formative assessment and monitoring of students' progress. *Computers & Education, 76,* 30–41. https://doi.org/10.1016/j.compedu.2014.03.001.

Rogers, G. D., Mey, A., & Chan, P. C. (2017). Development of a phenomenologically derived method to assess affective learning in student journals following impactive educational experiences. *Medical Teacher.* https://doi.org/10.1080/0142159X.2017.1372566.

Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education, 35*(5), 535–550. https://doi.org/10.1080/02602930903541015.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age Publishers.

Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted learning, 33*(1), 1–19. https://doi.org/10.1111/jcal.12172.

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior, 63,* 106–117. https://doi.org/10.1016/j.chb.2016.05.047.

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in digital networks—ICT literacy: A novel assessment of students' 21st century skills. *Computers & Education, 109,* 11–37. https://doi.org/10.1016/j.compedu.2017.01.014.

Siddiq, F., & Scherer, R. (2017). Revealing the processes of students' interaction with a novel collaborative problem solving task: An in-depth analysis of think-aloud protocols. *Computers in Human Behavior, 76,* 509–525. https://doi.org/10.1016/j.chb.2017.08.007.

Spector, J. M., Ifenthaler, D., Sampson, D., Yang, L. J., Mukama, E., Warusavitarana, A., et al. (2016). Technology enhanced formative assessment for 21st century learning. *Journal of Educational Technology & Society, 19*(3), 58.

Tempelaar, D. T., Heck, A., Cuypers, H., van der Kooij, H., & van de Vrie, E. (2013). *Formative assessment and learning analytics.* Paper presented at the Proceedings of the Third International Conference on Learning Analytics and Knowledge, Leuven, Belgium.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276.

Valtonen, T., Sointu, E., Kukkonen, J., Kontkanen, S., Lambert, M. C., & Mäkitalo-Siegl, K. (2017). TPACK updated to measure pre-service teachers' twenty-first century skills. *Australasian Journal of Educational Technology, 33*(3), 15–31.

van der Kleij, F., & Adie, L. (2018). Formative assessment and feedback using IT. In Springer (Ed.), *International handbook of IT in primary and secondary education (2nd ed.).*

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., et al. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing, 18*(6), 1499–1514. https://doi.org/10.1007/s00779-013-0751-2.

Voogt, J., Erstad, O., Dede, C., & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted learning, 29*(5), 403–413. https://doi.org/10.1111/jcal.12029.

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

Webb, M. E., Andresen, B. B., Angeli, C., Carvalho, A. A., Dobozy, E., Laugesen, H.,… Strijker, A. (2017). Thematic working group 5: Formative assessment supported by technology. In K. W. Lai, J. Voogt, & G. Knezek (Eds.), *EDUsummIT 2017 summary reports.*

Webb, M. E., & Gibson, D. C. (2015). Technology enhanced assessment in complex collaborative settings. *Education and Information Technologies, 20*(4), 675–695. https://doi.org/10.1007/s10639-015-9413-5.

Webb, M. E., Gibson, D. C., & Forkosh-Baruch, A. (2013). Challenges for information technology supporting educational assessment. *Journal of Computer Assisted learning, 29*(5), 451–462. https://doi.org/10.1111/jcal.12033.

Whitelock, D., & Bektik, D. (2018). Progress and challenges for automated scoring and feedback systems for large-scale assessments. In Springer (Ed.), *International handbook of IT in primary and secondary education (2nd ed.)*.

Wilson, A., Howitt, S., & Higgins, D. (2016). Assessing the unassessable: making learning visible in undergraduates' experiences of scientific research. *Assessment & Evaluation in Higher Education, 41*(6), 901–916. https://doi.org/10.1080/02602938.2015.1050582.