ATS STEM

Report #3 of
ATS STEM Report Series

Digital Formative Assessment of Transversal Skills in STEM

A Review of Underlying Principles and Best Practice

Katherine Reynolds
Michael O'Leary
Mark Brown
Eamon Costello

www.atsstem.eu

#atsstem @stem_ats

Co-funded by the Erasmus+ Programme of the European Union

Please cite as: Reynolds, K., O'Leary, M., Brown, M. & Costello, E. (2020). Digital Formative Assessment of Transversal Skills in STEM: A Review of Underlying Principles and Best Practice. *ATS STEM Report #3*. Dublin: Dublin City University. **http://dx.doi.org/10.5281/zenodo.3673365**

This is Report #3 of #5 in the ATS STEM Report Series. All reports in the series are available from the project website: **http://www.atsstem.eu/ireland/reports/**

- Report #1: STEM Education in Schools: What Can We Learn from the Research?
- Report #2: Government Responses to the Challenge of STEM Education: Case Studies from Europe
- Report #3: Digital Formative Assessment of Transversal Skills in STEM: A Review of Underlying Principles and Best Practice
- Report #4: Virtual Learning Environments and Digital Tools for Implementing Formative Assessment of Transversal Skills in STEM
- Report #5: Towards the ATS STEM Conceptual Framework

ISBN: 978-1-911669-05-0

Typesetting, design and layout by

TABLE OF CONTENTS

FOREWORD

Assessment of Transversal Skills in STEM (ATS STEM) is an innovative policy experimentation project being conducted across eight European Union countries through a partnership of 12 educational institutions **(www.atsstem.eu)**. The project is funded by Erasmus+ (Call reference: EACEA/28/2017 - European policy experimentations in the fields of Education and Training, and Youth led by high-level public authorities). The project aims to enhance formative digital assessment of students' transversal skills in STEM (Science, Technology, Engineering and Mathematics). ATS STEM is co-financed by the ERASMUS+ Programme (Key Action 3 – Policy Experimentation). The project partnership comprises ministries of education, national and regional education agencies; researchers and pilot schools.

The countries and regions in which the digital assessment for STEM skill are being piloted are **Austria, Belgium/ Flanders, Cyprus, Finland, Ireland, Slovenia, Spain/Galicia** and **Sweden** as per below:

- Dublin City University, Ireland
- H2 Learning, Ireland
- Kildare Education Centre, Ireland
- Danube University Krems, Austria
- Go! Het Gemeenschapsonderwijs, Belgium
- Cyprus Pedagogical Institute, Cyprus
- University of Tampere, Finland
- Ministry of Education, Science and Sport, Slovenia
- National Education Institute Slovenia
- University of Santiago De Compostela, Spain
- Consejería De Educación, Universidad Y Fp (Xunta De Galicia), Spain
- Haninge Kommun, Sweden

Dublin City University (DCU) is the project coordinator. A core element of DCU's vision is to be a globally-significant university that is renowned for its discovery and translation of knowledge to advance society. DCU has an interdepartmental team of experts from three different research centres bringing their combined expertise to bear to help lead and deliver the project goals. These centres have expertise in digital learning, STEM education and assessment and are respectively the National Institute for Digital Learning (NIDL), the Centre for the Advancement of STEM Teaching and Learning (CASTeL) and the Centre for Assessment Research, Policy and Practice in Education (CARPE).

NIDL
National Institute for Digital Learning

The National Institute for Digital Learning (NIDL) aims to be a world leader at the forefront of designing, implementing and researching new blended, on-line and digital (BOLD) models of education **(https://www.dcu.ie/nidl/index.shtml)**. The NIDL'S mission is to design, implement and research distinctive and transformative models of BOLD education which help to transform lives and societies by providing strategic leadership, enabling and contributing to world-class scholarship, and promoting academic and operational excellence.

CASTeL
Centre for the Advancement of STEM Teaching and Learning

The Centre for the Advancement of STEM Teaching and Learning (CASTeL) is Ireland's largest research centre in STEM education **(http://castel.ie/)**. CASTeL's mission is to support the development of STEM learners from an early age, and so enhance the scientific, mathematical and technological capacity of society. CASTeL encompasses research expertise from across the Faculty of Science and Health and the DCU Institute of Education, one of Europe's largest educational faculties.

CARPE
Centre for Assessment Research, Policy and Practice in Education

The Centre for Assessment Research, Policy and Practice in Education (CARPE) is supported by a grant from Prometric to Dublin City University **(https://www.dcu.ie/carpe/index.shtml)**. The centre was established to enhance the practice of assessment across all levels of the educational system, from early childhood to fourth level and beyond.

ACKNOWLEDGEMENTS

The feedback from many colleagues on the ATS STEM project on various drafts of this report is gratefully acknowledged. The authors would like to thank Paula Lehane, Vasiliki Pitsia and Conor Scully at the Centre for Assessment Research, Policy and Practice in Education (CARPE) and Prajakta Girme at the National Institute for Digital Learning (NIDL) for their work in helping to edit and proof this document.

EXECUTIVE SUMMARY

This report was written as part of a research project titled, *Assessment of Transversal Skills in STEM (ATS STEM)*. The project is funded by Erasmus+ (Call reference: EACEA/28/2017 - European policy experimentations in the fields of Education and Training, and Youth led by high-level public authorities). The report is based on outputs related to two of the project's work packages as outlined in Appendix A – namely, *Review of digital assessment approaches* (WP1.4) and, *Formative assessment design* (WP2.1).

This report is the third in a set of five based on deliverables related to the ATS STEM project. Reports #1 and #2 are concerned with the research pertaining to STEM education in schools and with national policies for STEM in various European countries, respectively. Report #4 examines the potential of various technology-enhanced tools and architectures that might be used to support assessment for learning in STEM. Drawing on all four of these, the fifth report presents a draft integrated conceptual framework for the assessment of transversal skills in STEM.

Two major themes are addressed here. First, consideration is given to the key ideas and principles underlying formative assessment theory. Second, the current state of the art with respect to how STEM digital formative assessment is conceptualised and leveraged to support learning of transversal skills in STEM is discussed. Here, particular attention is paid to approaches that enable problem/research-based learning, enquiry-based learning, collaborative learning and mobile learning.

Formative Assessment

Formative assessment is considered to be a dynamic or cyclical process involving the elicitation and interpretation of evidence that will be used by learners and their teachers to make decisions about next steps in learning. Formative assessment inherently requires an understanding of instructional goals (i.e., the knowledge or skills to be learned), students' current progress in reaching those goals, and what students must do to close the gap between their current attainment and the end goals. The rationale for the use of formative assessment derives from constructivist and socio-cultural theories of learning which posit that learners build on existing ideas and competences to learn new things and that learning takes place through interacting with others rather than working alone.

In the literature, five key strategies to support the formative assessment process predominate. They involve clarifying and sharing learning intentions and success criteria; eliciting evidence of learning through classroom discussions, questions, and tasks; providing feedback that moves learners forward; peer-assessment and self-assessment. Feedback, in particular, plays a central role in formative assessment. The important thing about feedback is what students do with it. While there is no way to guarantee that students will use feedback in a given situation, there are some forms of feedback that stand a greater chance of being effective than others. Feedback related to process and self-regulation is considered to be most helpful in advancing student learning. Above all, teachers must provide enough time for both the provision of feedback and for students to make sense of and use the feedback. When this happens, the research evidence suggests that formative assessment can be effective for bolstering student achievement in STEM fields, as well as in other areas such as self-regulation and motivation to learn.

In designing formative assessments, a key principle to follow is that if a formative assessment does not support student learning, then it cannot be said to be valid for its intended purpose. Thus, attention to student learning that occurs as a result of formative assessment is an essential part of a validity argument in support of it.

Formative assessment in STEM can present its own special challenges with respect to content domain definition. While the individual STEM subjects are taught separately in most curricular contexts, there is evidence to support the consideration of STEM as a single, unitary domain. This highlights the importance of reaching a shared understanding of the content (knowledge, skills, attitudes), and, perhaps even more importantly, of the learning progressions that underpin STEM subjects and the concept of STEM as an entity in and of itself.

Digital Formative Assessment

Interest in digital formative assessment has grown rapidly in the past few decades and some reasons for that include: the provision of feedback in a timelier manner; the assessment of hard to measure constructs and processes that were previously inaccessible; the inclusion of new item types capable of providing more nuanced information about learning; automation of the feedback process; access for students with disabilities; greater opportunities for student collaboration.

In planning for technology-enhanced formative assessment four types of learning for building critical skills in STEM must be considered.

- Problem and research-based learning (PBL) - a constructivist approach to learning, where students are given meaningful problems to solve.
- Enquiry based learning - similar to PBL but students pose questions and carry out investigations to answer those questions.
- Collaborative learning – where students work in small groups to achieve a learning goal.
- Mobile learning - where students use mobile devices in the context of a learning environment inherently based on the use of technology or when technology provides greater flexibility in terms of where the learning takes place.

Three models are useful for framing digital formative assessment. The Technological Pedagogical Content Knowledge (TPACK) model presents a general way of thinking about integration of technology with teaching and learning. The Formative Assessment in Science and Mathematics Education (FaSMEd) model integrates technological functions with formative assessment. The digital literacy model illustrates different skills and competences that may be required of teachers and students for the effective implementation of a technology-enhanced formative assessment programme.

The exact nature of technological competences or literacies required by teachers and students for any particular initiative will ultimately be determined by the design of the formative assessment program. However, while the specific forms of technology-enhanced formative assessments may vary, research studies highlight the importance of adequate support and training for teachers in the implementation of technology-enhanced formative assessment systems. Explicit consideration of teacher professional development related to formative assessment is also warranted. It is incorrect to assume that simply because a generation of students has grown up with technology that they know how to use it effectively, or that it will enhance their engagement and learning.

Above all, the incorporation of technology into formative assessment is a means to an end, not a goal in its own right. Formative assessment validity evidence is seriously compromised if a digital formative assessment does not improve student learning.

INTRODUCTION

The overarching focus of this report (#3) is on design issues related to the implementation of formative assessment using digital technology. More technical issues, such as descriptions and evaluations of specific software programmes and platforms, are contained in a separate report (#5). For stylistic purposes, the terms *digital assessment* and *technology-enhanced assessment* are used interchangeably throughout this document.

This report begins by defining formative assessment, explicating the theory underlying it and the principles of its practice. Related issues, such as what constitutes effective feedback, are also addressed. The report then discusses validity and reliability in formative assessment contexts. Moving from theory to practice, literature related to the general effects of formative assessment is presented. It should be noted that some of these issues were not included in the specifications for work packages in the original proposal submitted for funding (see Appendix A). However, consideration of them in this report was thought to be an essential first step in the development of any formative assessment programme.

The discussion then shifts to focus specifically on digital formative assessment in STEM fields. Several frameworks and models are presented to guide thinking about what a particular digital formative assessment initiative might look like and how it might be used to enhance learning of core skills and competences in STEM. Studies concerning the implementation of specific digital formative assessments are also reviewed.

Key questions for consideration are posed at the conclusion of this report. These are questions for which there is no one "right" answer; rather, they are intended to guide the decision-making process relating to the use of formative assessment and how digital technology can aid the process. It is important to note at this point that the questions address issues pertinent to digital formative assessment more generally and do not focus on the specific tools that might be used to implement it.

FORMATIVE ASSESSMENT DEFINITIONS AND PRACTICES

Defining Formative Assessment

The concept of formative assessment can be traced to work in the area of programme evaluation, which began by differentiating between evaluative work for formative and summative purposes. While summative evaluation was focused on rendering judgements about particular programmes or interventions, formative evaluation was aimed at facilitating improvement (Bennett, 2011). This distinction roughly translates to the assessment field: summative assessment is typically thought to constitute assessment *of* learning, while formative assessment is assessment *for* learning.

It is important to note that this distinction refers to *the purpose of an assessment's use*, not an assessment tool itself. An individual assessment cannot unequivocally be declared formative or summative, as this depends on the inferences to be drawn. As noted by Black and Wiliam (2018):

> *Where the inferences refer to the status of the student, or about their future potential, then the assessment is functioning summatively. Where the inferences relate to the kinds of activities that would best help the student to learn, then the assessment is functioning formatively (p.553)*

Using this conception of assessment, it is possible for a single assessment to serve both formative and summative purposes. This leads some authors (e.g., Bennett, 2011) to posit that the formative/summative distinction is overly simplistic. Wiliam and Black (1996) propose considering formative and summative assessment purposes as a continuum, rather than being inherently at odds. The formative end of the continuum stresses the immediate usability of information obtained from an assessment, while the summative focuses on the consistent meaning of this information. Inferences drawn from a particular assessment may not neatly fall into a formative or summative purpose (Hayward, 2015). In that context, it is interesting to note that some now contend that the seamless integration of formative and summative assessment will be made possible by the use of digital technologies (e.g. Looney, 2019). A recent commentary notes that definitions of formative assessment need to distinguish between assessments that follow psychometric ideals and the evaluative on-the-fly judgements made by teachers in classrooms – the argument being that the latter fall into a category that should not be called formative assessment (Jönsson, 2020). In addition, Looney (2019) notes that interpretations of formative assessment can vary across countries and education cultures. That said, a definition of formative assessment that has been used widely over the past decade and a half is the one developed in the UK by the Assessment Reform Group to guide the work they were doing at the time:

> *[Formative assessment] ...is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. (ARG 2002, p. 1)*

The reader is referred to a recent chapter by Dolin, Black, Harlen and Tiberghien (2018), who provide an in-depth exploration of the relationship between formative and summative assessment. These researchers also highlight the important point that the rationale for the use of formative assessment derives from constructivist and sociocultural (also called situated) theories of learning. They explain:

> *...we recognise that developing understanding requires active participation of learners in constructing their learning. This accords with a cognitive constructivist view of learning, that is, learners making sense of new experiences starting from existing ideas and competencies. Recognition that learning is not entirely, or even mainly, an individual matter but takes place through social interaction is the basis of the sociocultural constructivist perspective of learning. In this view, understanding results from making sense of new experience with others rather than by working individually. (pp. 59-60)*

In addition, Black and Wiliam (2018) argue that the theory underlying formative assessment has to be considered within a broader theory of pedagogy context. All are important in how formative assessment plays out in practice.

The Practice of Formative Assessment

Formative assessment is considered to be a dynamic or cyclical process involving the elicitation and interpretation of evidence that will form the basis of certain actions (Wiliam & Black, 1996). These steps may look different depending on context. In a very short cycle, a teacher might pose a question to a student and ask follow-up questions based on the student's response. A longer cycle might involve the administration of a more formal assessment, where the teacher then critically reviews students' performance and makes adjustments as necessary. As Dolin, Black, et al. (2018) explain, in essence, the process involves:

- Establishing clear goals, and progression steps (including criteria) towards them, that are shared between teachers and students
- Designing or choosing a method for collecting data about student performance
- Using the method for gathering evidence of where students stand in relation to these goals
- Interpreting this evidence to decide what is needed to help students work towards the goals
- Deciding what action can help the student through the next progression step (p. 56)

Ruiz-Primo and Furtak (2006) elaborate on this basic cycle by integrating its steps with classroom discourse theory, proposing a slightly more specific iteration. They propose the ESRU model, where a teacher poses a question (E: Elicit), a student responds (S: Student), the teacher recognises/interprets the response (R: Recognize), and the teacher uses the response (U: Use). These steps can be thought of as occurring within "assessment conversations", which, like the general formative assessment cycle, can look quite different depending upon context (p. 207).

Bennett (2011) highlights that the success of any formative assessment process rests on the accurate interpretation of the evidence elicited. There are many reasons that a student might provide an incorrect answer when a question is posed. A student may make a careless mistake, have a misconception, or simply be lacking in knowledge without having a specific misconception. Interpretations of evidence should also be construct-referenced, implying that the action taken by teachers based on that evidence will improve student understanding beyond response to a single item (Wiliam & Black, 1996).

Shavelson et al. (2008) provide a framework within which to situate classroom formative assessment processes. Specifically, they posit different "degrees" of formative assessment, which are distinguished by their respective levels of formality. At the informal end of the continuum, the authors place "on-the-fly" assessment, which is unplanned and occurs throughout the course of a lesson; these assessment occasions are also sometimes called "teachable moments" (p. 300). Slightly more formal is "planned-for-interaction", which occurs when a teacher purposefully poses questions at key predetermined points within a lesson to evaluate student understanding and inform action. Finally, "embedded-in-the-curriculum" assessment occurs at the formal end of the continuum. This occurs when formal, pre-made assessment occasions are explicitly planned for within the curriculum. These assessments are purposefully administered at occasions where "important sub-goal[s] should have been reached" (p. 301). Results of these assessments can inform teachers as to whether the class is ready to move forward in the curriculum or if some form of re-teaching is required.

Implicit in each of these frameworks is that formative assessment inherently requires an understanding of instructional goals (i.e., the knowledge or skills to be learned), students' current progress in reaching those goals, and what students must do to close the gap between their current attainment and the end goals. Crucially, it also requires some understanding of learning progressions, defined as "the stages or steps that theory suggests most students go through as they progress toward mastering an important competency, like a key concept, process, strategy, practice or habit of mind" (ETS, nd). Developing learning progressions to support the assessment of transversal skills more generally and for STEM education in particular is challenging (ATC21S, 2014; Duschl, 2019), and recent research would suggest that many teachers are not familiar with the concept of learning progressions and, as a consequence, rarely use them as a basis for feedback (Dolin, Black et al., 2018).

Five strategies proposed by Wiliam and Thompson in 2007 to support the formative assessment process have been particularly popular over the past decade or so. Drawing upon the roles of teachers, students, and peers (i.e. other students relative to a single student), the strategies can be articulated as actions to be implemented:

1. Clarify and share learning intentions and criteria for success.
2. Elicit evidence of learning through classroom discussions, questions, and tasks.
3. Provide feedback that moves learners forward.
4. Activate students as instructional resources for one another (peer-assessment)
5. Activate students as the owners of their own learning (self-assessment).

Table 1 provides a visual representation of Wiliam and Thompson's framework (adapted from p. 63).

Table 1. Formative Assessment Framework

	Where the Learner is Going	Where the Learner is Right Now	How to Get There
Teacher	Clarifying and sharing learning intentions and criteria for success	Engineering effective classroom discussions and tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as owners of their own learning	

While formative assessment is perhaps most frequently considered as a way to advance students' cognitive knowledge in various content areas, it can also play a role in developing other capabilities—particularly self-regulation, which is defined as "the degree to which students are motivationally, metacognitively, and behaviorally active participants in their own learning process" (Meusen-Beekman, Joosten-ten Brinke, & Boshuizen, 2015, p. 3). As is evident in Table 1, self-regulation is intimately related to formative assessment; students with self-regulation skills are able to engage in formative assessment activities on their own behalf. Teacher-guided formative assessment can aid students in the development of their self-regulation skills (Meusen-Beekman et al., 2015).

Publications by Lysaght and O'Leary (2017) and by Moss and Brookhart (2019) provide many practical examples of how formative assessment can be operationalised by teachers in the classroom and by instructional leaders in schools. Web-based outputs from the EU funded *Strategies for Assessment of Inquiry Learning in Science* (SAILS) project provide illustrative examples of classroom based assessment practices applied across the sciences **(http://www.sails-project.eu/index.html)**. The publication by Finlayson and McLoughlin (2017) provides an excellent overview of the framework used in SAILS framework to link the range of inquiry skills and competences needed for learning in science, assessment and teacher education/professional development.

DESIGNING FORMATIVE ASSESSMENTS

Reliability and Validity

Validity and reliability are important concerns in the design of any assessment programme. The latest *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014) define validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests" (p. 11) and reliability as "the consistency of scores across replications of a testing procedure" (p. 33). Conventional wisdom within the testing profession maintains that reliability is a "necessary, but not sufficient" condition for validity. The validity of a test for a particular purpose is typically established through a *validity argument,* which draws on many sources of *validity evidence* (AERA, APA, & NCME, 2014). The *Standards* are written in such a way as to generally prioritise summative interpretations and uses of test scores; formative assessments present their own special considerations and challenges.

Some authors argue that unlike summative assessment, formative assessment does not require a high degree of reliability in order to establish validity. Sadler (1989) writes: "Attention to validity of judgements about individual pieces of work should take precedence over attention to reliability of grading in any context where the emphasis is on diagnosis and improvement" (p. 122). This argument is advanced in light of the contexts in which formative assessment takes place. **Unlike summative assessments, many of which are likely to be standardised and must have evidence of reliability in order to justify high-stakes decisions, formative assessments are more likely to be individualised and do not have high-stakes consequences for individual learners. It is therefore much more important that interpretations of evidence gathered in individual formative assessment contexts be accurate in the moment rather than consistent over time.**

The *Standards* (2014) list many different forms of evidence that may contribute to a validity argument for a test. However, not all are weighted equally in formative assessment. Stobart (2006) writes: **"The deceptively simple claim of this chapter is that for formative assessment to be valid it must lead to further learning. The validity argument is therefore about the consequences of assessment"** (p. 133). Other forms of validity evidence (e.g., content, convergent) may play supporting roles, but **if a formative assessment does not support student learning, it cannot be said to be valid for its intended purpose.** Thus, attention to student learning that occurs as a result of formative assessment is an essential part of the assessment's validity argument. Many threats exist to this consequential validity evidence, including lack of time or other resources and the atmosphere of trust and motivation (or lack thereof) in a classroom (Stobart, 2006).

Bennett (2011) frames this issue differently, positing that formative assessments require both a validity argument ("to support the quality of inferences and instructional adjustments") and an efficacy argument ("to support the resulting impact on learning and instruction") (p. 14). Both of these arguments should be supported with logical and empirical evidence, making use of appropriate and rigorous research methods. The validity argument should include traditional forms of validity evidence (e.g., content, convergent), while the efficacy argument should provide evidence that a particular formative assessment was executed effectively and led to improved student learning.

Effective Feedback

Given the importance of consequences for establishing validity in formative assessment practices, and the central role that feedback plays in the same, a brief discussion of what constitutes effective feedback is warranted. Wiliam (2016) notes that **"the only important thing about feedback is what students do with it" (p. 10). While there is no way to guarantee that students will use feedback in a given situation, there are some forms of feedback that stand a greater chance of being effective than others.**

Hattie and Timperley (2007) propose a three-part framework for situating feedback. This involves attention to articulating ultimate goals for students ("Feed Up"), giving students an indication of their progress ("Feed Back"), and showing students where they should move to next ("Feed Forward") (p. 87). As with the formative assessment strategies discussed above, attention to the ultimate goal is essential when providing students with feedback. If feedback is not focused on advancing students' progress towards goals, it cannot improve student learning.

Feedback can fall into one of four categories: 1) Feedback about a specific task; 2) Feedback about a process; 3) Feedback related to self-regulation; or 4) Feedback directed at the personal self. According to Hattie and Timperley (2007), feedback related to process and self-regulation is most helpful in advancing student learning. Limited task-oriented feedback may also be useful; however, feedback directed at the personal self is not helpful because it focuses on the student as a person rather than being directed towards the instructional goal (e.g., "You did a great job!").

Content is not the only feedback-related variable contributing to the use of feedback by students. Time is also essential. Cowie, Moreland, and Otrel-Cass (2013) note that classroom teachers must "plan for and organise time for both the provision of feedback and for *students to make sense of and use the feedback*" (emphasis added, p. 99). If feedback is provided to a student, but the lesson immediately moves on, there is no real opportunity for the student to learn from the feedback that was provided.

Intended Content Domains (STEM Focus)

Intended content domain is a key consideration for any assessment development activity; however, it is particularly important for formative assessments that are developed externally (i.e., not by teachers themselves). Bennett (2011) asserts that "...to be maximally effective, formative assessment requires the interaction of general principles, strategies, and techniques with reasonably deep cognitive-domain understanding" (p.15). Wiliam (2019) echoes this sentiment, but notes that discussions of formative assessment principles should also be broad enough as to apply across multiple disciplines. This implies that formative assessments in different domains will require their own instruments and their own feedback mechanisms. A one-size-fits-all approach is unlikely to be useful. **Additionally, teachers making use of an externally developed formative assessment must have adequate content domain knowledge in order to interpret evidence obtained from the instrument accurately and carry out informed actions.**

Formative assessment in STEM can present its own special challenges with respect to content domain definition. While the individual STEM subjects are taught separately in most curricular contexts, there is evidence to support the consideration of STEM as a single, unitary domain. Knowles and Kelley (2016) argue for the unification of STEM, although they also acknowledge the difficulty of this task. To address this difficulty, the authors suggest approaching STEM through a lens of situated learning cognition, which "...recognizes that the contexts, both physical and social elements of a learning activity, are critical to the learning process... Often when learning is grounded in a situated context, learning is authentic and relevant, therefore, representative of an experience found in actual STEM practice" (Knowles & Kelley, 2016, p. 4). This argument is supported by Barrett, Moran, and Woods (2014), who describe the implementation of a STEM module integrating meteorology and engineering. Barrett et al. (2014) argue that consideration of these subjects together is essential given their interaction in the realm of public safety (e.g., knowing conditions that can lead to tornado formation as well as design principles that may allow buildings to withstand tornadoes). Some empirical support also exists for the unification of STEM. Bicer, Capraro, and Capraro (2017) used a structural equation modelling approach to establish support for a single STEM construct contributing to Texan students' achievement in both mathematics and science.

Moving beyond how STEM itself is conceptualised, the specifications surrounding specific STEM content areas can also have implications for formative assessment. Burkhardt and Schoenfeld (2019), while discussing mathematics, note that although the overarching goal of formative assessment (i.e., improvement of student learning) is relatively straightforward, this goal is complicated by how the content domain in question is defined. For example, the authors note that "...in mathematics in particular, the goals of instruction have changed significantly in the past decades, shifting from a content focused perspective... to one that gives equal attention to such content and to practices including reasoning, proving, conjecturing, representing, etc." (p. 38). Similar content shifts have recently happened in the field of science, which have begun to highlight the social nature of science, as well as disciplinary practices (Furtak, Heredia, & Morrison, 2019). **This highlights the importance of reaching a shared understanding of the key content and, perhaps, even more importantly, learning progressions, that underpin all STEM subjects and the concept of STEM as a unitary domain.**

For a more detailed discussion of defining STEM within the context of this project, see Report #1 in this series. The reader is also referred to a key output from a European research project that investigated assessment methods aimed at supporting and improving inquiry-based approaches in European science, technology and mathematics (STM) education – see Dolin and Evans (2018).

FORMATIVE ASSESSMENT USE AND RELATIONSHIP WITH STUDENT OUTCOMES

This section provides an overview of research establishing relationships between various forms of formative assessment and student outcomes. In addition to examining cognitive or achievement-related outcomes, this body of research also touches upon other student skills or attributes, including self-regulation and motivation.

Researchers were quick to acknowledge the benefits of formative assessment following the release of a seminal review by Black and Wiliam (1998). Since then, the review, as well as subsequent interpretations of the review, have been critiqued (e.g., Dunn & Mulvenon, 2009; Bennett, 2011). Because of the lack of standardization within formative assessment, it is difficult to draw definitive conclusions about the practice as a whole (Wiliam Lee, Harrison, & Black, 2004). Some researchers (e.g., Bennett, 2011) claim that we cannot unequivocally declare that use of formative assessment leads to increased student learning. In light of this, this section reviews recent research concerning the implementation of formative assessment systems and their relationship with STEM-related student outcomes.

Burns, Klingbeil, and Ysseldyke (2010) conducted a study examining the relationship between the use of Accelerated Math and achievement on state mathematics exams. The authors refer to Accelerated Math as a "technology-enhanced formative evaluation" tool that uses a data-driven decision-making framework in which computer-adaptive tests are used to collect data, determine appropriate instructional targets, and monitor student progress" (p. 583). Accelerated Math enables teachers to print targeted worksheets for students to attain specific objectives and notifies teachers when students are ready for a mastery test. The authors found that achievement on state tests was higher for schools using Accelerated Math than schools in comparison groups for each state. Four US states were represented in this study, each of which had their own mathematics assessment.

Decristan et al. (2015) used a cluster randomised trial in German elementary schools to evaluate the effects of formative assessment on student learning in a unit related to the topic of 'sinking and floating'. Additionally, the authors examined the interaction of formative assessment with classroom process quality, which includes cognitive activation of students, supportive classroom climate, and effective classroom management. It was hypothesised that effective use of formative assessment might have a positive relationship with these classroom process quality indicators. The formative assessment consisted of open-ended assessments accompanied by feedback sheets. Using hierarchical linear modelling, the authors found that formative assessment was positively related to student learning and that this relationship was strengthened by high cognitive activation of students and supportive classroom climate. There was no interaction effect for classroom management.

Timmers, Walraven, and Veldkamp (2015) examined the effects of self-regulation related feedback on adolescent students' performance on an information seeking task. Fifty students from the Netherlands completed two information seeking tasks requiring internet searches. When finished with the first tasks, students were prompted to reflect upon what they might do differently for the second task. Students who articulated specific goals or actions for the second task (for example, accessing websites to gather information) tended to follow through, visiting more unique websites than those who did not articulate specific goals and actions. The researchers found that students' performance upon the second task improved, although still generally fell short of expert graders' expectations.

Maier, Wolfe, and Randler (2016) studied the effect of different types of automated feedback provided by a computerized Concept Test in biology in 10 Bavarian high schools. Students were assigned to one of three conditions: 1) no feedback; 2) verification-only (i.e., correct or incorrect) feedback; and 3) elaborated feedback. The authors found that verification-only feedback was generally associated with higher achievement than the other two conditions after controlling for prior achievement and motivation. However, elaborated feedback was associated with higher achievement than verification-only feedback when students indicated that they found the feedback useful. The authors posit that the elaborated feedback in the formative assessment system may have been too detailed, leading some students to ignore it.

Faber, Luyten, and Visscher (2017) examined the effects of a digital formative assessment tool for mathematics achievement and motivation in a randomized control trial with primary school students in the Netherlands. Forty classrooms made use of the digital formative assessment tool, *Snappet*, while fifty classrooms engaged in regular teaching practices. *Snappet*, accessed using a tablet, provided verification feedback to students, summary and individual feedback to teachers, and adaptive lessons based on student performance. Using a multilevel regression model, Faber et al. (2017) found that use of *Snappet* positively impacted both mathematics achievement and motivation.

Hooker (2017) conducted a qualitative study of teachers' and parents' perceptions of ePortfolios in an Australian special education context. Six families and associated key teachers were interviewed in two rounds of data collection. The teachers were using traditional paper learning portfolios during the first round of data collection and transitioned to ePortfolios for the second round. Teachers found that they were able to contribute to the ePortfolios on a more regular basis, which parents appreciated. Both groups also appreciated the consistency of formatting with respect to the ePortfolios, which made it easier for teachers to organize information and facilitate parents' consumption of information entered by the teachers. Hooker (2017) concluded that the introduction of ePortfolios was beneficial, but cautioned that such benefits could only be realized "if the platform is thoughtfully and meaningfully constructed using a sound theoretical base, evidence from practice and is context specific" (p. 450).

Vogelzang and Admiraal (2017) conducted an action research study concerning the incorporation of formative assessment into two sections of the curriculum of a chemistry class in the Netherlands. A chemistry unit on lactic acid was divided into two sub-units, each of which had a pre- and a post-test. Formative assessment in the form of extensive discussion and feedback was incorporated into one sub-unit for each of the course sections. Groups did not differ at the pre-test for either sub-unit and in both cases the group engaging in formative assessment scored higher on the post-test.

Hondrich, Decristan, Hertel, and Klieme (2018) sought to determine the effects of formative assessment on students' motivation and perceptions of their own competence in a unit related to relative density and buoyancy. The study took place in a German primary school and made use of a pre-/post-test design with a comparison group. The formative assessment intervention included short, open-ended written tasks accompanied by written feedback and adaption of instruction based upon results (including differentiated worksheets). Teachers were told to stress the formative nature of these activities. Using multilevel modelling, the authors found that the treatment and comparison students did not differ in perceptions of their own competence or motivation prior to the intervention; however, the group engaging in formative assessment scored higher on both post-measures.

Pinger, Rakoczy, Besser, and Klieme (2018) examined the relationship between the quality of formative assessment practice, student achievement, and student interest in the context of a German 9th grade unit on the Pythagorean Theorem. Teachers were instructed to embed process-oriented feedback (the formative assessment mechanism) at three junctures in the unit. A pre-/post-test design was used. The authors coded written process-oriented feedback provided by the teachers and videotaped a single lesson to evaluate the embeddedness of the formative assessment into the class. The following were coded: the number of comments provided, the specificity of feedback provided, feedback related to the personal self, and feedback using social reference norms. Pinger et al. (2018) found that teachers generally did not provide feedback related to the personal self or use social reference norms. Using multilevel modelling, they also found that both the number of comments provided as well as feedback specificity had negative relationships with post-test performance (however, these students also tended to score lower on the pre-test) and no relationship with interest. The embeddedness of formative assessment and repeated feedback emphasis was positively associated with interest at the class level. However, the coding of formative assessment embeddedness from the videotaped lessons had low inter-rater reliability.

Taken together, these results suggest that formative assessment can be effective for bolstering student achievement in STEM fields, as well as other affective domains and student skills. However, they do not guarantee that formative assessment *will* be effective in these areas. It was noted earlier that a diverse array of formative assessment techniques exist and, as will become clear, this multitude is only increased when digital technology enhancements are incorporated.

DIGITAL FORMATIVE ASSESSMENT

Digital formative assessment includes all features of the digital learning environment that support assessment of student progress and which provide information to be used as feedback to modify the teaching and learning activities in which students are engaged (Looney, 2019, p. 10).

Interest in digital formative assessment has grown rapidly in the past few decades (Shute & Rahimi, 2017). One reason for this **is the potential of digital technology to either deploy or enable the provision of feedback in a timelier manner** compared to a teacher unaided by technology (Spector et al., 2016). Other benefits include greater opportunities for self/peer assessment and student collaboration, support for learner choice in assessment, access for students with disabilities, the assessment of anywhere/anytime learning, and the seamless integration of formative and summative assessment (Looney, 2019). In addition, **the use of such technologies allow for the measurement of constructs and processes that were previously inaccessible, rendering them potentially transformative for the science of assessment** (Shute, Leighton, Jang, & Chu, 2016). Some models for thinking about the integration of digital formative assessment into classroom practice, as well as current applications of technology-enhanced formative assessments, are introduced in the sections to follow.

Frameworks and Models

Three models are useful for framing digital formative assessment. One of these models, Technological Pedagogical Content Knowledge (TPACK) (Koehler & Mishra, 2009) presents a general way of thinking about integration of technology with teaching and learning. Another Formative Assessment in Science and Mathematics Education (FaSMEd) (European Commission, 2016), integrates technological functions with formative assessment. Thirdly, Ng's (2012) digital literacy framework illustrates different skills and competences that may be required of teachers and students for the effective implementation of a technology-enhanced formative assessment program.

The TPACK model acknowledges upfront that there is "no 'one best way' to integrate technology into the curriculum" (Koehler & Mishra, 2009, p. 62). Teaching is an activity that is highly context-dependent, and there are many different forms of knowledge required by teachers to adapt as contexts change over time. Specifically, the TPACK sets forth three types of knowledge teachers must have to effectively integrate technology into their classrooms; these three general types of knowledge also intersect with each other. These knowledge types and their intersections are presented in Figure 1.

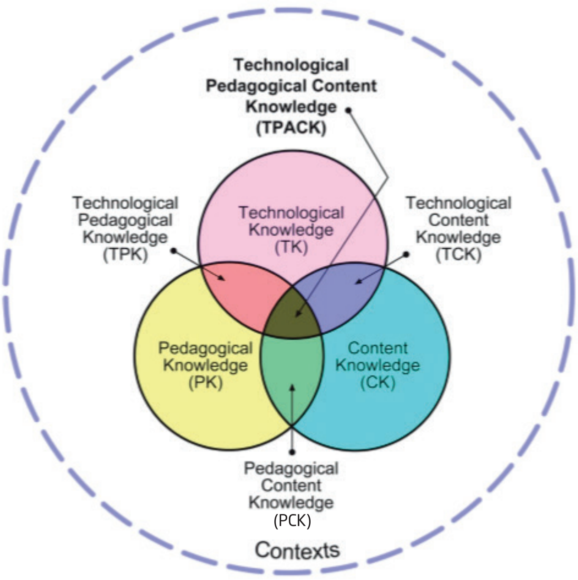

Figure 1. The TPACK Model (Reproduced from Koehler & Mishra, 2009, p. 63).

TPACK represents the intersection of content, pedagogical, and technological knowledge for a teacher. Each type of knowledge is described below:

- Content: Subject-specific knowledge about what is to be learned.
- Pedagogical: Knowledge about the processes of teaching and learning.
- Technological: Sufficient broad knowledge of technological tools to use them in everyday life.

Each of the knowledge types interacts with and can inform the other. For example, technological pedagogical knowledge requires teachers to understand how to best incorporate technological tools into classrooms for the purpose of enhancing student learning. TPACK, where all knowledge forms overlap, is considered more than the sum of its individual parts, and involves:

> *...an understanding of the representation of concepts using technologies; pedagogical techniques that use technologies in constructive ways to teach content; knowledge of what makes concepts difficult or easy to learn and how technology can help redress some of the problems that students face; knowledge of students' prior knowledge and theories of epistemology; and knowledge of how technologies can be used to build on existing knowledge to develop new epistemologies or strengthen old ones (Koehler & Mishra, 2009, p. 66).*

Over the years, TPACK has been subject to considerable comment and critique (Harris, Phillips, Koehler, & Rosenberg, 2017). Cherner and Smith (2017) argue that a weakness of the original model is that it was focused on teacher knowledge rather than on student learning. In reconceptualising the model, they center it specifically on student learning of 21st century skills with teachers being required to take account of the contextual factors that affect and define the student, while simultaneously teaching academic content and technology use.

Focusing specifically on formative assessment, the FaSMEd initiative, sponsored by the European Commission, builds on the framework of Wiliam and Thompson (2007). The FaSMEd framework proposes a third dimension along with the actors and strategies of formative assessment. This third dimension is referred to as functionalities of technology. It includes three ways that technology might be integrated into the formative assessment process (European Commission, 2016, p. 5-6):

1. Sending and Displaying: These are actions that facilitate communication between the different actors in the formative assessment process; they can be thought of as facilitating the elicitation and student response processes. A classroom response system where students reply to items using phones or tablets and where results are displayed for the class would be an example of this.
2. Processing and Analysing: These are actions where technology supports the interpretation phase of formative assessment, such as extracting or summarizing relevant data. An example of this would be a data dashboard summarizing student performance.
3. Providing an Interactive Environment: These are actions that enable students to work individually or collaboratively to explore content and may include features from the other two categories. Examples of this are specialized software for allowing students to explore geometrical drawings or other specific topics.

A visual representation of this model is presented in Figure 2.

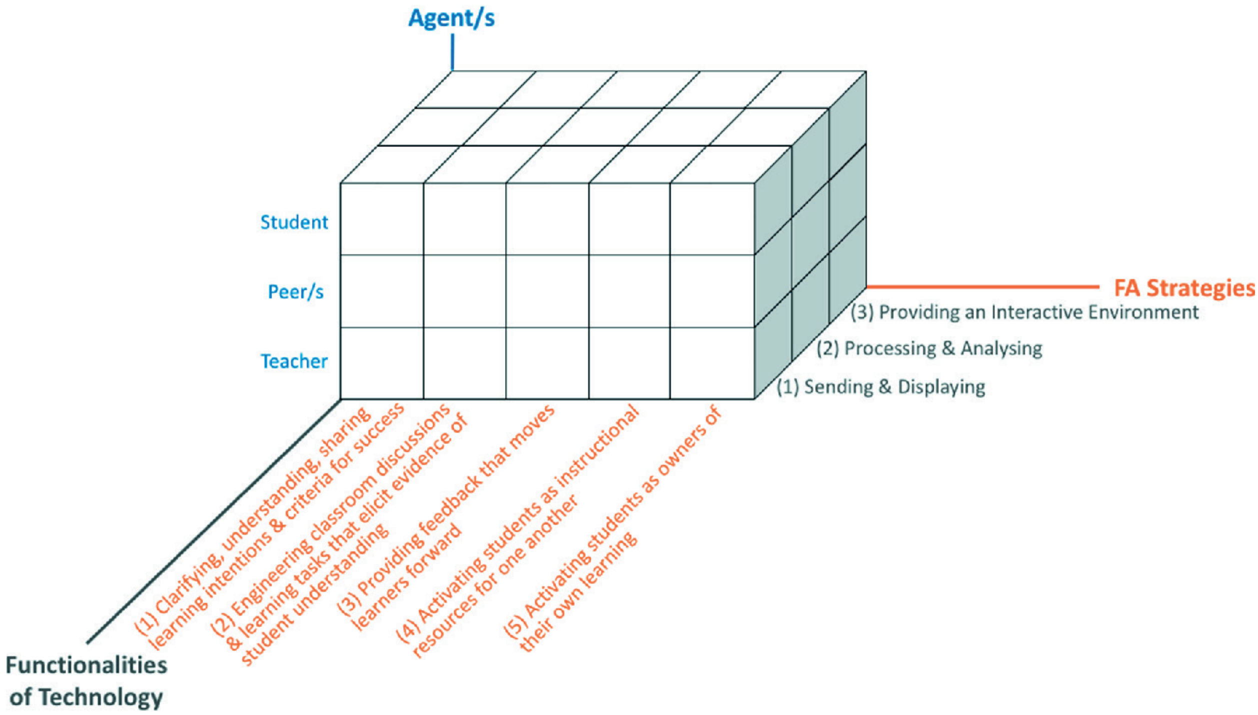

Figure 2. FaSMEd Technology Enhanced Formative Assessment Model (Reproduced from European Commission, 2016, p.5)

Each of the three technology functionalities may be useful for any of the five formative assessment strategies outlined in Figure 2. Much of the existing work regarding technology-enhanced assessment in STEM concerns large - scale assessments to inform summative inferences (e.g., Quellmaz et al., 2013); the functionalities of technology in these contexts (e.g., allowing improved fit to a multidimensional IRT model) are not likely to be as relevant to teachers using technology in formative contexts.

Together, these two frameworks provide a useful way of thinking about the development of a technology-enhanced formative assessment programme. The FaSMEd framework highlights specific technological functions that may be incorporated into formative assessment practices, while the TPACK model explicates the forms of knowledge needed by teachers to effectively capitalise on these functions to improve student learning. At this juncture, a third model is also useful: Ng's (2012) framework of digital literacy. This framework is useful for guiding consideration of how the digital competences or literacies of students and teachers may come into play over the course of this technology-enhanced formative assessment initiative.

The exact nature of technological competences or literacies required by teachers and students for any particular initiative will ultimately be determined by the design of the formative assessment program. **Digital competence and digital literacy are not terms with widely agreed-upon definitions** (McGarr & McDonagh, 2019), although as Spante, Hashemi, Lundin and Algers (2018) note, they are often used synonymously and underpin each other.

Ng (2012) proposes three general components of digital literacy: technical, cognitive, and social-emotional. The technical dimension refers to the routine processes of using digital tools (e.g., set-up, operating, troubleshooting). This digital literacy dimension is clearly important for technology-enhanced formative assessment initiatives; both teachers and students must be able to effectively make use of whatever technological tools are involved. The cognitive dimension involves evaluating and selecting appropriate digital tools for a given goal or process. This dimension is important to consider if the technology-enhanced formative assessment programme has multiple components; teachers must be able to evaluate the utility of different programme components at different junctures within a lesson. Finally, the social-emotional dimension involves the ability to critically navigate technologies that allow for communication among individuals. Given the potentially collaborative nature of technology-enhanced formative assessment, this dimension also merits consideration.

POTENTIAL USES OF DIGITAL FORMATIVE ASSESSMENT

Across the literature, four types of learning stand out as being important when building critical skills for STEM. These must also be considered when technology-enhanced formative assessment is being addressed:

- Problem and research-based learning (PBL): PBL is a constructivist approach to learning, where students' activities are "organized around the investigation, explanation, and resolution of meaningful problems" (Hmelo-Silver, 2004, p. 236). Students are provided with a meaningful problem context and learn through the process of solving the problem. Although conceptually attractive, PBL has been criticized for its lack of efficiency; students may complete problem solving activities without gaining understanding of the intended content (Kirschner, Sweller, & Clark, 2006).
- Enquiry based learning: Enquiry based learning is similar to PBL, with the important distinction that students drive the enquiry process. The enquiry learning process requires students to pose questions, carry out investigations to answer those questions, and communicate knowledge with others (Jennings, 2010). Enquiry based learning faces similar critiques to those of PBL (Kirschner et al., 2006).
- Collaborative learning: Collaborative learning occurs when students work in small groups to achieve a learning goal, challenging students socially and emotionally in addition to cognitively (Laal, 2013).
- Mobile learning: Mobile learning occurs when students use mobile devices in the context of a learning environment (Orr, 2010; Crompton, Burke & Lin, 2019). Unlike the previous three areas, mobile learning is inherently based on the use of technology and facilitates anywhere/anytime learning and assessment.

One of the most-heralded benefits of technology-enhanced assessment is the incorporation of new item types capable of providing more nuanced information than simple multiple-choice or constructed-response items (Sireci & Zenisky, 2006). For example, students may carry out tasks in extended and dynamic contexts such as the *problem-solving* and inquiry tasks incorporated into the 2019 eTIMSS assessment (Martin, Mullis, & Foy, 2017). Problem- and inquiry-based tasks at the classroom level may also be enhanced by incorporation of technological tools, such as those developed by NASA's *Classrooms for the Future* initiative (Hickey, Taashoobshirazi, & Cross, 2012). Embedded within a meaningful ongoing context, such assessment items may have the potential to realise Knowles and Kelley's (2016) unified framework for STEM.

In addition to providing a meaningful context with engaging items, technology-enhanced assessment can aid inquiry in other ways. **For example, incorporation of technology-enhanced assessment can automate the feedback process for students completing an inquiry-based task.** A variety of different automated feedback conditions may be programmed into the assessment, allowing differentiation based on students' specific outcomes (e.g., Ryoo & Linn, 2016).

Incorporation of technology into classroom assessment practice can also enhance collaborative learning among students in a myriad of ways. Dukuzumuremyi and Siklander (2018) report on student-to-student interaction over laptops in a Finnish primary school context. They found that students interacted verbally, nonverbally (e.g., typing notes into a shared document), and kinaesthetically (e.g., shaking hands in agreement) around the laptops while engaging in collaborative computer instruction. van Dijk and Lazonder (2016) discuss peer evaluation of inquiry-based concept maps in a technology-enhanced learning environment for middle school students. ePortfolios have also been used for collaborative formative assessment in university courses concerning design and technology. After assembling ePortfolios, students engaged in an adaptive comparative judgement task, helping them to clarify their own understandings of design quality by holistically evaluating the work of others (Canty, Seery, Hartell, & Doyle, 2017; Seery, Canty, & Phelan, 2012).

A significant project drawing together many of the key ideas underpinning problem/research-based learning and collaborative learning was the *Assessment and Teaching of 21st Century Skills* (ATC21S™) - an initiative supported by CISCO, Intel, Microsoft and the governments of Australia, Singapore, Finland, United States, the Netherlands, and Costa Rica. With a focus on collaborative problem solving and learning in digital networks, the research outcomes are significant in that they (a) provide a roadmap for how transversal skills can be conceptualised and taught, (b) provide validity evidence for tasks developed to generate data for formative feedback and (c) provide guidance for how digital technology can be leveraged to support the assessment process. That said, the research also points to the sheer complexity involved in assessing a transversal skill such as collaborative problem solving (see, Griffin, McGaw, & Care, 2012; Griffin & Care, 2015; Care, Griffin, & Wilson, 2018).

With respect to *mobile learning*, Nikou and Economides (2018) reviewed 43 pieces of literature related to mobile based assessment. They found that mobile based assessment was most commonly used by primary school students for STEM subjects and that most studies came from Taiwan, China, the United States, and Spain. They conclude that findings regarding mobile based assessment are generally positive, noting that 60% of reviewed studies found positive relationships between mobile assessment and achievement (and 33% of studies did not explicitly discuss student achievement). Additionally, mobile based assessment seemed to be perceived positively by students, although this finding was extrapolated from student comments in the studies rather than statistical analysis. The researchers found that perceived ease of use was the most important factor for teachers when deciding whether or not to implement mobile-based assessment. Lai (2019) provides a comprehensive overview of trends in research on mobile learning more generally.

As noted above, the four learning types (problem based, inquiry based, collaborative, and mobile) are not mutually exclusive. This is demonstrated in the form of project e-scape (Kimbell, 2012), where personal digital assistants (PDAs, comparable to personal mobile devices) were used to aid in collaboration among students. This deliberate integration of technology and student collaboration (discussed further below) may be informative for formative assessment efforts.

Incorporation of technology into formative assessment may also assist with enabling access for students with disabilities. Scalise et al. (2018) reviewed literature related to accessibility and accommodations for technology-enhanced STEM tasks. They began by noting that there are several frameworks available that can guide assessment developers in ensuring accessibility of technology-enhanced items, including the Universal Design framework, the Web Content Accessibility Guidelines, and the Accessible Portable Item Protocol. Missing data is a major concern for students with disabilities, which compromises assessment validity. Technology provides an opportunity to employ many strategies that may reduce missing data among students with disabilities, including language simplification, altering presentation modalities, and providing a targeted interface, time, or tools for specific students.

Finally, incorporation of technological tools presents new opportunities for evaluating the impact of formative assessment on students' non-content-specific skills, such as self-regulation. One such opportunity is presented in the form of moment-by-moment learning curves (MBMLCs) (Baker, Goldstein, & Heffernan, 2011). Data mining techniques are used to calculate the probability that a student has learned a particular skill at a given moment in time based upon factors such as correctness of a response, likelihood of a lucky guess, and likelihood of incorrect answers despite understanding of the skill. This is then represented visually in an MBMLC. Molenaar, Horvers, and Baker (2019) examined the relationships between different MBMLC shapes and students' self-regulation in a mathematics assessment. They detected four MBMLC patterns, each of which was related to different patterns of accurate responses and pre/post-test score changes. MBMLCs represent one way in which process data generated by technology-enhanced formative assessments might be analysed.

IMPLEMENTATION OF DIGITAL FORMATIVE ASSESSMENT

This section provides an overview of literature related to the implementation of various technology-enhanced formative assessments. Some of these studies are descriptive, simply reporting what this implementation "looks like". Others specifically highlight teachers' perceptions of technology-enhanced formative assessment, noting any particular difficulties they faced. The specific forms of technology-enhanced formative assessments vary across studies; however together, these studies highlight the importance of adequate support and training for teachers in the implementation of technology-enhanced formative assessment systems.

Feldman and Capobianco (2008) examined US teachers' processes of integrating a personal response system (PRS) for formative assessment into physics classes. The PRS comprised hardware and software including "remotes that students use to send answers to a computer that then displays the results in a histogram" (p. 82). The system came pre-programmed with physics questions; teachers could also create their own. Eight physics teachers were interviewed three times over the course of PRS adoption; classes were also observed 3-5 times. The authors found that teachers (particularly novice teachers) struggled to manage the logistics of implementing the PRS, but teachers did grow in their use and understanding of the PRS over time. Teachers tended not to use the pre-programmed physics items because they were not well-aligned with what was being taught. Teachers also used the PRS in unexpected ways, such as polling students' attitudes about issues in science.

Lee, Feldman, and Beatty (2012) conducted a survey of primary and secondary teachers in the United States in order to capture perceptions about the challenges of incorporating classroom response systems (CRS) into formative assessment. The authors found that many factors influenced teachers' use of CRS, including logistical issues (e.g., time, difficulty operating technology), classroom or pedagogical issues (e.g., difficulty writing effective questions, lack of skill conducting formative assessment, student behavior, difficulty facilitating whole-class discussions), and broader contextual factors (e.g., non-school commitments, teachers' beliefs about the technology).

Kimbell (2012) provides an overview of the implementation of *project e-scape*. This digital portfolio assessment system involves students (initially in design/technology courses, later expanded to science and geography) using PDAs to create ePortfolios documenting their learning processes. Student-to-student feedback is integrated as part of the process. Kimbell (2012) notes the importance of providing appropriate training to teachers throughout the implementation process, covering topics such as the general structure of the project e-scape activities and how best to facilitate them. Other key issues that are highlighted include the manageability of integrating the system into classrooms and the pedagogical issue of "the extent to which the use (for assessment purposes) of such a system can support and enrich the learning experience of design and technology" (Kimbell, Wheeler, Miller & Pollitt, 2007, p. 21).

Aldon, Cusi, Morselli, Panero, and Sabena (2015) conducted observational case studies of individual teachers involved with the FaSMEd project in Italy and France. In both contexts, the authors found some disconnect in how teachers spoke about formative assessment versus how it manifested in their classrooms. While teachers spoke about using formative assessment strategies in interviews, "...most of the time FA [formative assessment] was not developed over time and appeared occasionally in the classroom more as a reassuring method than a teaching strategy" (p. 640).[1]

1 This finding is echoed in the partner-provided resource Hirsch & Lindberg (2015), which is not formally presented here because it is not written in English. Citation: Hirsh, Å., & Lindberg, V. (2015). Formativ bedömning på 2000-talet- en översikt av svensk och internationell forskning. Stockholm, Sweden.

The *Assessment of Transversal Skills 2020* (ATS2020), a large scale European funded research study involving 250 schools, 10,000 students (10-15 years old) and 1000 teachers from 10 countries (Belgium, Croatia, Cyprus, Estonia, Greece, Finland, Ireland, Lithuania, Slovenia and Spain). This study was designed principally to investigate the impact of innovative teaching and learning approaches allied to the use of ePortfolios and electronic journaling on the development of transversal skills across four competence areas: Information Literacy, Autonomous Learning, Collaboration and Communication, and Creativity and Innovation **(see, http://www.ats2020.eu/)**. While findings were generally positive in terms of student learning, many participants indicated that the use of the ePortfolio was time consuming and somewhat confusing to use. Time was also an issue in respect to the use of journals. Implementing both approaches was complicated by the fact that participants found self-assessment to be challenging (Economou, 2018). Crucially it was noted that "implementing an innovative learning model and its critical elements, such as ePortfolio, assessment of, for and as learning, development of transversal skills, and technology-enhanced learning design, is a complex process for both teachers and students, and it needs time to be adopted" (Economou, 2018, p. 28).

Haßler, Major, and Hennessy (2016) conducted a review of studies examining the use of tablets in schools. The authors did not focus explicitly upon formative assessment; however, this is likely to have been an intended tablet use for many studies reviewed. Sixteen of the twenty-three studies they reviewed reported positive relationships between student outcomes and the use of tablets, five reported no difference, and two reported negative relationships. The authors identified several factors contributing to positive outcomes: ease of tablet use, portability, use of a touchscreen (rather than a stylus), and effective integration into the curriculum.

Panero and Aldon (2016), also part of the FaSMEd study, carried out a case study of a 9th grade mathematics class that integrated tablets into instruction for formative assessment purposes. Three classroom observations were conducted over the course of a year. During the first observation, students tended to complete their work using paper and pencil and used the tablets only for submitting answers. By the third observation, the tablets were more integrated into students' learning processes. The most common uses of the tablet system by the teacher were displaying and discussing the work of a particular student or surveying the class.

Geer, White, Zeegers, Au, and Barnes (2017) conducted a mixed methods study to examine the integration of iPads into instruction in an Australian context. The authors found that students and teachers both found the iPads to be useful for giving or receiving feedback.

Shute and Rahimi (2017) conducted a review to determine how computer-based assessment for learning (CBAfL) is used in primary and secondary schools. Examining nine review papers and eight empirical studies, the authors found three prevalent general categories of use: CBAfL as a supplement to classroom instruction, web-based CBAfL, and data-driven or continuous CBAfL. As a supplement to classroom instruction, CBAfL was most frequently used as a way to provide prompt feedback to students or to personalize learning. Although most web-based assessments are summative in nature, web-based CBAfL was typically used to promote student engagement and enable students to monitor their own learning. Finally, data-driven or continuous assessment refers to the use of learning analytics or game-based assessment.

Nikou and Economides (2019) examined the utility of a model for capturing teachers' acceptance of mobile based assessment. The proposed technology acceptability model included ease of use, perceived utility, teacher mobile self-efficacy, social circumstances, facilitating school conditions, and perceptions of how well mobile assessment systems perform their intended functions. In all, 161 teachers from different European countries were surveyed regarding how each piece of the model related to their intention to use mobile based assessment.

Several findings emerge from this selection of studies:

1. Teachers require appropriate support and continuous professional development (CPD) in order to successfully integrate technology-enhanced formative assessment systems and/or platforms into their classrooms (here a distinction should be made between CPD to use technology and CPD for better use and integration of formative assessment more generally). This finding holds across different technological tools.
2. Perceived benefits and ease of use are important factors in teachers' incorporation of technology-enhanced formative assessment.
3. Even though teachers appreciate the importance of formative assessment, they may struggle to engage in it effectively.

These three findings should inform the development of any formative assessment initiative; more specifically, teacher input must be solicited early on to ensure clear benefits and user-friendliness. Additionally, support and technical training and ongoing professional development on the pedagogical integration of technology must be provided.

While the exact form of initial training will be determined by the nature of the initiative, one possible piece of ongoing support is the creation of a pedagogical pattern language to guide teachers' technology-enhanced formative assessment use. Originating in the field of architecture, patterns "describe[s] a problem which occurs over and over again in our environment, and then describes the core of a solution to that problem in such a way that you can use this solution a million times over, without doing it in the same way twice" (Alexander, Ishikawa, & Silverstein, 1977). Examples of pedagogical patterns are presented in Table 2.

Table 2. Example Pedagogical Patterns

Pattern Name	Pattern Source	Pattern Description
Assessment Diversity	Bergin et al. (2015)	"Use a variety of assessment techniques in each course to account for different learning modalities and increase the richness of student experience" (p. 9)
Three Stars and a Wish	Larson, Trees, & Weaver (2008)	"Find three positive things to say about a students' work before giving a criticism" (p. 3)

Grundschober, Ghoneim, Baumgartner, and Gruber-Muecke (2018) describe the process of "mining" pedagogical patterns in different areas of formative assessment and creating a concept map to show the relationships among them. The concept map is intended to clarify these relationships for teachers so as to increase the utility of the pedagogical patterns. The concept map and resources related to individual patterns are available freely online. This is an example of one form of resource this formative assessment project may consider creating for teachers. Grundschober et al.'s (2018) work may serve as a valuable starting point, with additional patterns incorporated as needed based on the nature of the technology-enhanced formative assessment programme developed.

Explicit consideration of teacher professional development related to the formative assessment initiative is also warranted. Beesley, Clark, Dempsey, and Tweed (2018) note that one-time professional development sessions are unlikely to change meaningfully or alter teachers' practices of formative assessment. Rather, they propose the Assessment Work Sample Method (AWSM) professional development model, where teachers receive an initial two-day introduction to mathematics formative assessment, followed by eight shorter sessions focused on analysing student work and crafting effective feedback. Irrespective of the method, the key point is that professional development must be ongoing and challenge teachers to implement and critically reflect new practices, where appropriate.

Some Caveats

Some have argued for the incorporation of technology into classrooms because it is attractive to students who have grown up as *digital natives* with internet access and other technological innovations. Kirschner and De Bruyckere (2017) problematise this characterization, arguing that *digital native* is a misnomer, and that it is incorrect to assume that simply because a generation has grown up with technology that they know how to use it effectively or that it will enhance their engagement and learning. In fact, income may have a much stronger relationship to technological savviness than age (e.g. Czaja et al., 2006) – something that has the potential to be problematic in the field of classroom based formative assessment. This finding suggests that effective incorporation of technology into formative assessment practice should not introduce construct-irrelevant variance for low-income students; this would introduce a serious problem with respect to validity.

Using a sample of lower secondary Norwegian students, Hatlevik, Guðmundsdóttir, and Loi (2015) found that the number of books in the home and speaking Norwegian at home were both associated with higher digital competence. The study's measure of digital competence included items that could be considered aligned with all three of Ng's (2012) components of digital literacy (technical, cognitive, social-emotional). This again highlights the importance of providing appropriate orientations for students (as well as teachers) when implementing technology-enhanced formative assessment programmes. It is also important to be mindful that different countries may have different cultural norms or expectations surrounding the use of technology in the classroom.

Crucially, designing learning environments with digital formative assessment in mind will require that teachers are both technologically literate and assessment literate, i.e., teachers who know how to leverage technology to improve teaching and learning through formative assessment. This is no easy task. The challenge is highlighted by Looney (2019) who notes that "a range of surveys and evaluations have found that a majority of teachers tend to use new technologies to reinforce traditional approaches to learning and assessment (p. 47). There is the onerous task of teacher professional development to consider.

As noted above, **incorporation of technology into formative assessment is a means to an end, not a goal in its own right.** Formative assessment validity evidence is seriously compromised if a formative assessment practice does not yield increased student learning. To keep the goal of student learning, authors (e.g., Spector et al., 2016; Shute et al., 2016) argue that some form of principled assessment design (e.g., evidence - centered design, assessment engineering) must be used when developing technology-enhanced formative assessment. Technological advances should be incorporated into formative assessment only as much as they enhance student learning, and the mechanisms for this enhancement should be a key part of the assessment's validity argument. The conclusion at the end of a 2018 review of the state of the art in digital technology-based assessment is worth reiterating:

> *It is clear that the field of assessment is undergoing great changes with the influence of digital technology. From a practical viewpoint, technology has improved the efficiency of many aspects of assessment delivery and scoring; and more recently, in parallel with advances in computing and artificial intelligence, it has opened up possibilities for increasingly complex, sophisticated and intellectually-challenging assessments. That said, it is also clear that we are only on the cusp of realising its full potential…references to 21st century skills are now firmly established in curricular frameworks and policy documents worldwide, but in reality, these skills are heterogeneous and practical efforts to assess them still lag behind. This is particularly true in the case of less cognitively-oriented skills, such as citizenship and personal and social responsibility. In order to ensure that future developments in technology-enhanced assessment take positive steps towards narrowing this gap, it is important to critically evaluate the contribution of each new innovation. Ultimately, those involved in assessment design would do well to bear in mind Bennett's description of third generation technology-based assessment as 'driven by substance'. It is imperative that technology does not become the primary focus of 21st century assessment. Emphasis must remain on reliability, validity, authenticity and underlying pedagogical purpose. (O'Leary, Scully, Karakolidis, & Pitsia, 2018, p. 11-12)*

KEY QUESTIONS AND CONSIDERATIONS

In conclusion, a number of key questions arise from this report. What an effective formative assessment system or platform might look like in practice is a particularly pertinent one and will be addressed in Report #4, specifically written to provide guidance for the identification of digital tools and associated architectures that can be used or adapted across various European contexts. The extant literature suggests one thing for certain -- developing an effective technology-enhanced formative assessment platform to support formative assessment of transversal skills in STEM will be no easy task. Assuming the development of a curriculum embedded formative assessment system (Shavelson et al., 2008), there are many questions that must be addressed meaningfully to maximise the chance of achieving the ultimate goal of improving student learning in STEM fields. The following series of questions presents some points for consideration.

What is meant by STEM?

A compelling case has been made for the consideration of STEM as a unitary construct, rather than having separate science, technology, engineering, and mathematics components. What vision of STEM will this (and future) initiatives adopt and how will that vision correspond to how STEM and its complementary subjects are currently covered in countries' curricula?

How will the ultimate improvement of STEM learning be assessed throughout the initiative?

As the goal of formative assessment is to improve student learning, a decision must be made about how "student learning" will be operationalised. Will summative assessments be developed in conjunction with the formative assessment system? Will individual countries or schools utilize their own summative assessments to this end? How will the validity and comparability of this measure be assured across contexts? This report does not have the capacity to answer these questions; however, they must be considered by stakeholders wishing to harness the potential of technology-enhanced formative assessments.

How will validity evidence for the formative assessment system be gathered and assembled into a validity argument?

The importance of validity within formative assessment cannot be overstated. Plans for assembling validity evidence and articulating a validity argument should be considered from the beginning of the process. This endeavor could be supported through the use of a principled assessment design framework such as evidence - centered design or assessment engineering.

How can the quality of delivery for the formative assessment system be assured?

Research has shown that the quality of delivery for formative assessments impacts the effects of those assessments, and that teachers sometimes struggle to successfully execute formative assessment processes. Once a formative assessment has been selected or created, plans must be made to ensure implementation fidelity of the system. Appropriate professional development must be provided to teachers and other parties as needed to ensure fidelity.

How will the principles of effective feedback be incorporated into the formative assessment system?

It is clear from the literature that some types of feedback are more effective than others. It is imperative that the formative assessment system developed in this project is process-oriented with self-regulation of learning as a key priority. With that in mind, future work must consider how different digital tools can provide these forms of feedback. Additionally, if feedback is to be elaborated, care must be taken that it is not too overwhelming for students to use. Finally, students must be provided with adequate time to make use of any feedback they are given. *Cognitive labs* (i.e., questioning students or teachers during their use of specific tools) while the assessment development process of the project is ongoing, might be a useful approach to gathering data on how all of these variables play out in practice.

How will the formative assessment system reflect existing frameworks of formative assessment actors, strategies, and technology-enabled actions?

Identifying how the formative assessment system maps to an existing framework, such as the technology modified version of Wiliam and Thompson's (2007) model, can serve as a foundation for the logical component of the assessment system's validity argument. It can also assist in raising important questions about implementation.

How will the formative assessment system ensure that technology is used purposefully to advance student learning, and not simply for its own sake?

Incorporation of technology into formative assessment practice should only be done in service of improving student learning. A clear theory of action for the formative assessment system should be articulated. This theory of action can be informed by the existing frameworks of formative assessment noted above.

Thinking more broadly, many other questions will arise when local and pan-European educational policy directions have to be considered. Drawing on the work of Looney (2019) four seem particularly pertinent: how to develop technology and assessment literacy in teachers, how to ensure plans for digital formative assessment cohere with other educational priorities, how to ensure equity in access to digital tools and programmes and how to encourage cooperation and investment in research and development for digital formative assessment.

REFERENCES

Aldon, G., Cusi, A., Morselli, F., Panero, M., & Sabena, C. (2015). Which support technology can give to mathematics formative assessment? The FaSMEd project in Italy and France. *Quaderni di Ricerca in Didattica (Mathematics), 25*, 631-641.

Alexander, C., Ishikawa, S. & Silverstein, M. (1977). *A pattern language: towns, buildings, construction*. New York, NY: Oxford University Press.

American Educational Research Association; Joint Committee on Standards for Educational and Psychological Testing (U.S.); American Psychological Association; National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

ARG (Assessment Reform Group). (2002). *Assessment for learning: Ten principles*. http://www.hkeaa.edu.hk/DocLibrary/SBA/HKDSE/Eng_DVD/doc/Afl_principles.pdf Accessed 4 February, 2020.

ATC21S. (2014). *Assessment and teaching of 21st century skills: Collaborative problem solving empirical progressions.* http://www.atc21s.org/uploads/3/7/0/0/37007163/collaborative_problem_solving_emprical_progressions_v1.1.pdf Accessed 13 February, 2020.

Baker, R.S., Goldstein, A.B., & Heffernan, N.T. (2011). Detecting learning moment by moment. *International Journal of Artificial Intelligence in Education, 21*(1-2), 5-25.

Barrett, B.S., Moran, A.L., & Woods, J.E. (2014). Meteorology meets engineering: An interdisciplinary STEM module for middle and early secondary school students. *International Journal of STEM Education, 1*(6).

Beesley, A.D., Clark, T.F., Dempsey, K., & Tweed, A. (2018). Enhancing formative assessment practice and encouraging middle school mathematics engagement and persistence. *School Science & Mathematics, 118*(1), 4-16.

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

Bergin, J., Kohls, C., Köppe, C., Mor, Y., Portier, M., Schümmer, T., & Warburton, S. (2015). Assessment-Driven Course Design - Fair Play Patterns. In *Proceedings of the 22nd Conference on Pattern Languages of Programs*. The Hillside Group. http:// www.hillside.net/plop/2015/papers/panthers/4.pdf Accessed 13 February, 2020.

Bicer, A., Capraro, R.M., & Capraro, M.M. (2017). Integrated STEM assessment model. *EURASIA Journal of Mathematics Science and Technology Education, 13*(7), 3959-3968.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice,* 5(1), 7-74.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: principles, policy, & practice, 25*(6), 551-575.

Burkhardt, H., & Schoenfeld, A. (2019). Formative assessment in mathematics. In H. Andrade, R.E. Bennett, & G. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 35-67). New York, NY: Routledge.

Burns, M.K., Klingbeil, DA., & Ysseldyke, J. (2010). The effects of technology-enhanced formative evaluation on student performance on state accountability math tests. *Psychology in the Schools, 47*(6), 582-591.

Canty, D., Seery, N., Hartell, E., & Doyle, A. (2017). *Integrating peer assessment in technology education through adaptive comparative judgement.* Paper presented at Pupils' Attitudes Toward Technology (PATT-34) Conference. Philadelphia, PA.

Care, E., Griffin, P., & Wilson, M. (Eds.). (2018). *Assessment and teaching of 21st century skills. Research and applications.* Dordrecht: Springer.

Cherner, T., & Smith, D. (2017). Reconceptualizing TPACK to meet the needs of twenty-first-century education. *The New Educator, 13*(4), 329-349.

Cowie, B., Moreland, J., & Otrel-Cass, K. (2013). *Expanding notions of assessment for learning inside science and technology primary classrooms*. Rotterdam, The Netherlands: Sense Publishers.

Crompton, H., Burke, D., & Lin, Y. (2019). Mobile learning and student cognition: A systematic review of PK-12 research using Bloom's Taxonomy. *British Journal of Educational Technology, 50*(2), 684–701.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A., & Sharit, J. (2006). Factors predicting the use of technology: findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). *Psychology and Aging, 21*(2), 333–352.

Decristan, J., Klieme, E., Kunter, M., Hochweber, J. Büttner , G., Fauth, B., Hondrich, A.L., Rieser, S., & Hardy, I. (2015). Embedded formative assessment and classroom process quality: How do they interact in promoting science understanding? *American Educational Research Journal, 52*(6), 1133-1159.

Dolin J., Black P., Harlen W., & Tiberghien A. (2018). Exploring relations between formative and summative assessment. In: Dolin J., Evans R. (Eds.), *Transforming assessment through an interplay between practice, research and policy. Contributions from science education research* (pp. 53-80). Cham, Switzerland: Springer.

Dolin, J., & Evans, R. (Eds.). (2018). *Transforming assessment through an interplay between practice, research and policy. Contributions from science education research*. Cham, Switzerland: Springer.

Dukuzumuremyi, S., & Siklander, P. (2018). Interactions between pupils and their teachers in collaborative and technology-enhanced learning settings in the inclusive classroom. *Teaching and Teacher Education, 76*, 165-174.

Dunn, K.,. & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research, & Evaluation, 14*(7), 1-11.

Duschl, R. (2019). Learning progressions: framing and designing coherent sequences for STEM Education. *Disciplinary and Interdisciplinary Science Education Research*, 1(4). https://doi.org/10.1186/s43031-019-0005-x Accessed 4 February, 2020.

Economou, A. (2018). *ATS2020-Assessment of transversal skills: Reflections and policy recommendations on transversal skills development and assessment.* http://www.ats2020.eu/images/documents/ATS2020-Reflections.pdf Accessed 13 February, 2020.

ETS (Educational Testing Service) (nd.). *What are learning progressions and why are they important?* https://news.ets.org/stories/what-are-learning-progressions-and-why-are-they-important/ Accessed 4 February, 2020.

European Commission. (2016). *FaSMEd summary report*. https://cordis.europa.eu/docs/results/612/612337/final1-final-fasmed-summary-report-final.pdf Accessed 4 February, 2020.

Faber, J.M., Luyten, H., & Visscher, A.J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education, 106*, 83-96.

Feldman, A., & Capobianco, B.M. (2008). Teacher learning of technology enhanced formative assessment. *Journal of Science Education Technology, 17*, 82-99.

Finlayson ,O., & McLoughlin, E. (2017) Building teacher confidence in inquiry and assessment: Experiences from a Pan-European Collaboration In M. Peters, B. Cowie, & I. Menter (Eds.). *A companion to research in teacher education* (pp. 825-838). Singapore: Springer.

Furtak, E.M., Heredia, S.C., & Morrison, D. (2019). Formative assessment in science education: Mapping a shifting terrain. In H. Andrade, R.E. Bennett, & G. Cizek (Eds.). *Handbook of formative assessment in the disciplines* (pp. 97-125). New York, NY: Routledge.

Geer, R., White, B., Zeegers, Y., Au, W., & Barnes, A. (2017). Emerging pedagogies for the use of iPads in schools. *British Journal of Educational Technology, 48*(2), 490-498.

Griffin, P., & Care, E. (Eds.). (2015). Assessment and teaching of 21st century skills: Methods and approaches. Dordrecht: Springer.

Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Dordrecht: Springer.

Grundschober, I., Ghoneim, A., Baumgartner, P., & Gruber-Mücke, T. (2018). A pattern language remix for ATS2020. Using existing pedagogical patterns to create a new language for formative assessment within the ATS2020 learning model. In R. Sickinger, P. Baumgartner, & T. Gruber-Mücke (Eds.). *Pursuit of pattern languages for societal change. A comprehensive perspective of current pattern research and practice* (pp. 288–317). Krems: Tredition.

Harris, J., Phillips, M., Koehler, M., & Rosenberg, J. (2017). TPCK/TPACK research and development: Past, present, and future directions. *Australasian Journal of Educational Technology, 33*(3), i-viii.

Haßler, B., Major, L., & Hennessy, S. (2016). Tablet use in schools: A critical review of the evidence for learning outcomes. *Journal of Computer Assisted Learning*, 32(2), 139-156.

Hatlevik, O. E., Guðmundsdóttir, G. B., & Loi, M. (2015). Examining factors predicting students' digital competence. *Journal of Information Technology Education: Research, 14*, 123-137.

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

Hayward, L. (2015). Assessment is learning: the preposition vanishes. *Assessment in Education: principles, policy & practice, 22*(1), 27-43.

Hickey, D.T., Taasoobshirazi, G., & Cross, D. (2012). Assessment as learning: enhancing discourse, understanding, and achievement in innovative science curricula. *Journal of Research in Science Teaching, 49*(10), 1240-1270.

Hmelo-Silver, C.E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*(3), 235-266.

Hondrich, A.L., Decristan, J., Hertel, S., & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift für Erziehungswissenschaft*, 21, 717-734.

Hooker, T. (2017). Transforming teachers' formative assessment practices through ePortfolios. *Teaching and Teacher Education, 67*, 440-453.

Jennings, L.B. (2010). Inquiry-based learning. In T.C. Hunt, J.C. Carper, T.J. Lasley II, & D. Raisch. *Encyclopedia of Educational Reform and Dissent.* Thousand Oaks, CA: SAGE Publications.

Jönsson, A. (2020). Definitions of formative assessment need to make a distinction between a psychometric understanding of assessment and "evaluate judgements." *Frontiers In Education, 5*(2), 1-4.

Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education, 22*(2), 135-155.

Kimbell, R., Wheeler, T., Miller, S., & Pollitt, A. (2007). *E-scape portfolio assessment: Phase 2 report*. Goldsmiths, University of London.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.

Kippers, W.B., Wolterinck, C.H.D., Schildkamp, K., Poortman, C.L., & Visscher, A.J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199-213.

Kirschner, P.A., & De Bruyckere, P. (2017). The myths of the digital natives and the multitasker. *Teaching and Teacher Education, 67*, 135-142.

Kirschner P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86.

Knowles, T.R., & Kelley, J.G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education, 3*(11), 1-11.

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60-70.

Laal, M. (2013). Collaborative learning; elements. *Social and Behavioral Sciences, 83*, 814-818.

Lai, C. (2019). Trends of mobile learning: A review of the top 100 highly cited papers. *British Journal of Educational Technology*, 50. doi:10.1111/bjet.12884 Accessed 4 February, 2020

Larson, K.A., Trees, F.P., & Weaver, D.S. (2008). Continuous Feedback Pedagogical Patterns. In *PLoP'08 Proceedings of the 15th Conference on Pattern Languages of Programs*. New York, NY: ACM. Retrieved from: http://hillside.net/plop/2008/papers/PLoP2008_26_Larson+Trees+Weaver.pdf

Lee, H., Feldman, A., & Beatty, I.D. (2012). Factors that affect science and mathematics teachers' initial implementation of technology-enhanced formative assessment using a classroom response system. *Journal of Science Education Technology*, *21*, 523-539.

Looney, J. (2019). Digital formative assessment: A review of the literature. http://www.eun.org/documents/411753/817341/Assess%40Learning+Literature+Review/be02d527-8c2f-45e3-9f75-2c5cd596261d Accessed 4 December, 2019.

Lysaght, Z., & O'Leary, M. (2017). Scaling up, writ small: using an assessment for learning audit instrument to stimulate site-based professional development, one school at a time. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 271-289.

Maier, W., Wolfe, 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.

Martin, M.O., Mullis, I.V.S., & Foy, P. (2017). TIMSS 2019 Assessment Design. In I.V.S. Mullis & M. O. Martin (Eds.), *TIMSS 2019 Assessment Frameworks* (pp. 79-91).

McGarr, O., & McDonagh, A. (2019). Digital competence in teacher education, *Output 1 of the Erasmus+ funded Developing Student Teachers' Digital Competence (DICTE) project*. https://dicte.oslomet.no/ Accessed 13 February, 2020.

Meusen-Beekman, K., Joosten-ten Brinke, D., & Boshuizen, H.P.A. (2015). Developing young adolescents' self-regulation by means of formative assessment: A theoretical perspective. *Cogent Education, 2*(1), 1071233.

Mishra, P., & Koehler, M.J. (2006). Technological pedagogical content knowledge: A framework for integrating technology in teacher knowledge. *Teachers College Record, 108*(6), 1017-1054.

Molenaar, I., Horvers, A., & Baker, R. S. (2019). What can moment-by-moment learning curves tell about students' self-regulated learning?. *Learning and Instruction*, 101206. https://doi.org/10.1016/j.learninstruc.2019.05.003. Accessed 4 February, 2020.

Moss, C.M., & Brookhart, S. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders.* Alexandria, VA: ASCD Publications.

Ng, W. (2012). *Empowering scientific literacy through digital literacy and multiliteracies*. Hauppauge, NY: Nova Science Publishers, Inc.

Nikou, S.A., & Economides, A.A. (2018). Mobile-based assessment: A literature review of publications in major refered journals from 2009 to 2018. *Computers & Education, 125*, 101-119.

Nikou, S.A., & Economides, A.A. (2019). Factors that influence behavioral intention to use mobile-based assessment: A STEM teacher's perspective. *British Journal of Educational Technology, 50*(2), 587-600.

O'Leary, M., Scully, D., Karakolidis, A., & Pitsia, V. (2018). The state of the art in digital technology based assessment. *European Journal of Education, 53*(2), 160-175.

Orr, G. (2010). *A review of literature in mobile learning: Affordances and constraints*. Paper presented at the 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education, Kaohsiung, Taiwan. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5476544 Accessed 13 February, 2020.

Panero, M., & Aldon, G. (2016). How teachers evolve their formative assessment practices when digital tools are involved in the classroom. *Digital Experiences in Mathematics Education, 2*, 70-86.

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment – effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice, 25*(2), 160-182.

Quellmalz, E.S., Davenport, J.L., Timms, M.J., DeBoer, G.E., Jordan, K.A., Huang, C., & Buckley, B.C. (2013). Next-generation environments for assessing and promoting complex science learning. *Journal of Educational Psychology, 105*(4), 1100-1114.

Ruiz-Primo, M.A., & Furtak, E.M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3), 205-235.

Ryoo, K., & Linn, M.C. (2016). Designing automated guidance for concept diagrams in inquiry instruction. *Journal of Research in Science Teaching, 53*(7), 1003-1035.

Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144.

Scalise, K., Irvin, P.S., Alresheed, F., Zvoch, K., Yim-Dockery, H., Park, S., Landis, B., Meng, P., Kleinfelder, B., Halladay, L., & Partsafas, A. (2018). Accommodations in digital interactive STEM assessment tasks: Current accommodations and promising practices for enhancing accessibility for students with disabilities. *Journal of Special Education Technology, 33*(4), 219-236.

Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology Design Education, 22*, 205-226.

Shavelson, R.J., Young, D.B., Ayala, C.C., Brandon, P.R., Furtak, E.M., Ruiz-Primo, M.A., Tomita, M.K., & Yin,Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21*(4), 295-314.

Shute, V.J., Leighton, J.P., Jang, E.E., & Chu, M. (2016). Advances in the science of assessment. *Educational Assessment, 21*(1), 34-59.

Shute, V.J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*, 1-19.

Sireci, S.G., & Zenisky, A.L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp. 329-348). New York, NY: Routledge.

Spante, M., Hashemi, S., Lundin, M., & Algers, A. (2018). Digital competence and digital literacy in higher education research: Systematic review of concept use. *Cogent Education, 5*, 1-21.

Spector, J.M., Ifenthaler, D., Sampson, D., Yang, L., Mukama, E., Warusavitarana, A., Dona, K.L., Eichorn, K., Fluck, A., Huang, R., Bridges, S., Lu, J., Ren, Y., Gui, X., Deenan, C.C., San Diego, J., & Gibson, D.C. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology and Society, 19*(3), 58-71.

Stobart, G. (2006). The validity of formative assessment. In Garder, J. (Ed.), *Assessment and Learning.* Thousand Oaks, CA: SAGE Publications Ltd.

Timmers, C.F., Walraven, A., & Veldkamp, B.P. (2015). The effect of regulation feedback in computer-based formative assessment on information problem solving. *Computers & Education, 87*, 1-9.

van Dijk, A.M., & Lazonder, A. (2016). Scaffolding students' use of learner-generated content in a technology-enhanced inquiry learning environment. *Interactive Learning Environments, 24*(1), 194-204.

Vogelzang, J., & Admiraal, W.F. (2017). Classroom action research on formative assessment in a context-based chemistry course. *Educational Action Research, 25*(1), 155-166.

Wiliam, D. (2019). Why formative assessment is always both domain-general and domain-specific and what matters is the balance between the two. In H. Andrade, R.E. Bennett, & G. Cizek (Eds.), *Handbook of formative assessment in the disciplines* (pp. 243-264). New York, NY: Routledge.

Wiliam, D. (2016). The secret of effective feedback. *Educational Leadership, 73*(7), 10-15.

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal, 22*(5), 537-548.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning.* New York, NY: Lawrence Erlbaum Associates.

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

APPENDIX A

Assessment of Transversal Skills in STEM (ATS STEM)
Erasmus+ Call reference: EACEA/28/2017

Terms of Reference for Work Package 1, Task 4 and Work Package 2, Task 1

Excerpts from the Original Proposal (See pages 61-65)

WP 1 - STEM Conceptual Framework

Work package 1 (WP1) sets the baseline for this project providing the theoretical and operational frameworks for the policy experimentations. It will result in a set of sharable outputs that illustrate a pathway to the improvement and modernisation of STEM Education in schools in Europe within the partner countries to develop the skills of learners in the key areas of Science, Technology, Engineering and Mathematics.

Task 4: Review of digital assessment approaches: Digital Assessment of Learning of STEM Skills

This will involve a review of relevant digital assessment approaches to determine which contemporary technology-enhanced approaches are best suited to the teaching and learning of STEM. In particular, it will analyse and report on which approaches can enable:

- Problem-based and research-based learning
- Enquiry-based learning
- Collaborative learning
- Mobile learning

Output: A report that highlights best practice in digital assessment of core STEM Skills and competencies. This report will primarily be targeted at the STEM researchers in higher education, policy makers and those in ICT leadership roles in schools.

WP 2 - STEM Formative Digital Assessment Approach

Work package 2 (WP2) is focused on digital assessment and provides an evidence-based platform for the formative assessment of STEM learning tasks. It will result in a carefully selected STEM formative assessment digital tool package that fits the development and assessment of transversal skills as agreed upon in WP1. The outcomes of the comparison or adaptability of tools for STEM formative assessment will raise awareness of the didactic implications of formative assessment in the teaching and learning process. The development and/or adaptation of a tool package will be carried out based on careful review of existing solutions and in close cooperation with key users in order to suit the needs of the piloting partner's schools and support the didactic purpose of the chosen assessment as well as suit considerations regarding storage of evidence and quality assurance of the assessment operation and outcome.

Task 1: Formative Assessment Design: Building Critical Skills in STEM: How Digital Assessment Can Give Learners Feedback

- The initial task will involve: A review and synthesis of the research literature on STEM formative digital assessment with particular respect to schools.
- A mapping of the current state of the art of STEM formative digital assessment to show the state of the art in this area. It will highlight how students can best be scaffolded towards the development of key STEM skills and how digital tools can capture the evidence for this and augment teaching practices to help provide constructive feedback on student progress.

Output: A review and synthesis of state of the art on STEM formative digital assessment with particular respect to schools.

ATS STEM