Assessment practices for 21st century learning: review of evidence

Analytical report

More information on the European Union is available on the Internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2017

ISBN 978-92-79-70077-4
doi: 10.2766/71491

Please cite this publication as:

Siarova, H.; Sternadel, D.; Mašidlauskaitė, R., 'Assessment practices for 21st century learning: review of evidence', *NESET II report*, Luxembourg: Publications Office of the European Union, 2017. doi: 10.2766/71491.

AUTHORS:

- **Hanna Siarova,** Public Policy and Management Institute, Lithuania
- **Dalibor Sternadel,** Public Policy and Management Institute, Lithuania
- **Rūta Mašidlauskaitė,** Public Policy and Management Institute, Lithuania

PEER REVIEWERS:

- **Prof. Dr Therese N. Hopfenbeck,** Oxford University Centre for Educational Assessment (UK)
- **Dr Jana Huttova,** NESET II Scientific coordinator

ACKNOWLEDGMENTS:

The authors gratefully acknowledge the advice and useful comments from Therese Hopfenbeck and Jana Huttova on earlier versions of this report. The authors are also thankful for feedback from Susanne von Below and Vladimir Garkov that helped to improve this report.

CONTRACTOR:

Public Policy and Management Institute
Gedimino ave. 50, LT - 01110 Vilnius, Lithuania
Phone: +370 5 2620338 Fax: +370 5 2625410
www.ppmi.lt

Director Haroldas Brožaitis

ABOUT THE NESET II

NESET II is an advisory network of experts working on the social dimension of education and training. The European Commission's Directorate-General for Education and Culture initiated the establishment of the network as the successor to NESSE (2007-2010) and NESET (2011-2014). The Public Policy and Management Institute (PPMI) is responsible for the administration of the NESET II network.

CONTENTS

TABLES AND FIGURES

Figures

Tables

Boxes

EXECUTIVE SUMMARY

The social and economic transformations of contemporary knowledge societies demand new ways of thinking and learning. Knowledge alone is not sufficient; a broader range of skills and abilities is needed to navigate a moving landscape characterised by the increasing importance of information and communications technologies (ICTs), the decline of functional skills-based professions, and increasing competition. These skills and competences are no longer associated with certain academic disciplines, but are transversal and multi-dimensional in nature. The European key competence framework identified eight key competences supported by seven transversal skills necessary for personal fulfilment, active citizenship and all-rounded development in the 21st century, which are currently being further revised.

While the implementation of the key competence framework has been on the agenda of all Member States, policies and practices for the assessment of these competences have yet to be fully implemented in EU countries' efforts to integrate competence-based education into school curricula.

Assessing key competences and transversal skills is a challenging task, as they refer to complex constructs that are not easily measurable. Although assessment policies that support the assessment of traditional key competences such as maths, languages and science have largely been implemented, they are often limited to the contexts provided by the subject matters with which they are most closely associated, and rarely assess related attitudes. Cross-curricular competences and transversal skills are harder to associate with individual subjects and to reflect in specific learning outcomes. An innovative approach to assessment practices is needed to grasp the complexity and multiple roles of modern learning.

This report reviews international research to demonstrate how European education systems can improve their assessment practices to measure and support students' acquisition of key competences and transversal skills. It addresses the following questions:

- What are the different roles and purposes of classroom assessment?
- How can classroom assessment better support '21st century learning and teaching', and what are the key conditions for aligning current assessment practices with the development of key competences?
- What specific classroom assessment practices are effective to assess the full range of students' abilities and outcomes?
- What key recommendations can be made that can serve as important (first) steps to improve current policies on educational assessment?

One of the limitations of this review is the scarcity of European-based research that looks into the effectiveness of different assessment approaches when assessing non-traditional competences and transversal skills. That said, there are many innovative practices and inspiring approaches to assessing students' learning, both emerging and being tested, which the authors document in this review. However, we also acknowledge the limitations of the examples provided and the lack of empirical evidence on their effectiveness, where applicable.

Key findings

- **The field of educational assessment is currently divided and fragmented** into differing and often competing paradigms, methods and approaches: formative versus summative, norm-referenced

versus criterion/standards-referenced, internal versus external, measurement versus judgement, etc. However, at the same time more and more education stakeholders realise that assessment is a process which aims to document learning as well as to feed and improve it, and therefore needs to be guided by theories, models, and evidence (see section 2.1).

- Apart from conceptual and instructional considerations, all assessment practices used in the classroom have to adhere to overall technical requirements, such as **being valid and reliable**. Assessments and the qualifications they underpin are increasingly important because they may affect individuals' access to life chances – especially in employment and continuing levels of education. There is growing awareness of the importance of assessment and certification processes to be equitable, fair and inclusive. Hence, in addition to validity and reliability, **transparency, equity and freedom from bias** are key requirements of effective assessment (see section 2.2).
- **The successful implementation** of assessment methods covering a broad range of key competences **depends on a number of separate but interlinked conditions** at classroom, school and system level:
  - **Operationalising key competences in terms of learning outcomes** is a necessary foundation for consistent assessment practices.
  - The effective use of assessment results and assessment practices depends on **teachers' assessment literacy and ability to appropriately integrate assessment data in their teaching**, which in turn requires teacher education programmes (both initial teacher education and continuing professional development) ensure comprehensive preparation of teachers in this field (see chapter 3).
- Research demonstrates that **there is no single method that would fully measure key competences and transversal skills, nor serve as a best practice for student assessment**. Several methods and types of assessment need to be used to assess various skills comprehensively. The effectiveness of a method depends on its purposes and design, as well as on schools' and teachers' capacity to use it. At the same time, there is no universal combination of methods that would serve as a recipe. Teachers can be rather flexible in their choice of methods as long as these assessment approaches serve multiple purposes and follow the principles of validity, reliability and equity.
- Our review demonstrated that **standardised assessment methods such as tests and multiple-choice assessments are often used for the assessment of certain key competences, such as digital competence, languages, mathematics, and science**. Well-designed tests and multiple-choice questionnaires can also assess higher order skills, such as critical thinking or social and emotional competences.
- **Performance-based assessment has the potential to measure and foster wide-ranging competences and higher-order skills**, since it encompasses different assessment techniques and integrates a feedback mechanism. The key strengths of performance-based assessment include its focus on the learners' personalised needs, clear definitions of the learning goals, and timely feedback.
- Due to their collaborative nature, **peer- and self-assessments** can be effective in enabling students' 'deep learning'[1] and self-regulation, and **are important elements of an integrated assessment framework to be used in classrooms.** Research suggests that peer- and self-assessments are particularly useful in developing non-traditional competences, such as initiative and entrepreneurship,

[1] Deep learning refers to collaborative learning through reflection in action and on action (Fullan and Langworthy, 2014). The goals of deep learning are is to provide students with competences and dispositions necessary to become 'creative, connected, and collaborative life-long problem solvers and healthy, holistic human beings' (Ibid, p. 2).

learning to learn and social competence, as well as transversal skills such as critical thinking, creativity, problem-solving, risk assessment, decision-taking, and constructive management of feelings.

- **The use of ICT in assessment** allows to **deliver traditional assessment faster and more effectively** and at the same time offers opportunities to **change the way competences are assessed**, finding effective solutions for assessing non-traditional competences.

Key policy implications and recommendations

Our review suggests several steps, the implementation of which can help improve current education assessment policies and ensure that they grasp the full range of students' abilities and outcomes (see Chapter 5 for more details).

The main lessons and recommendations of this report are listed below.

- **The different functions of classroom assessment must be seen in synergy** with each other and be built into comprehensive assessment frameworks. Integrating different approaches to assessment can allow finding a balanced and consistent practice for assessing key competences.

Recommendations

- Policy-makers and schools should promote and employ an **integrated approach towards classroom assessment**, capitalising on the benefits and opportunities of summative, formative and diagnostic assessment to fit the multi-dimensional nature of key competences and transversal skills.
- Assessment practices need to **document learners' competences and help develop them** informing teachers' practices and curricula focus.
- When designing an integrated assessment framework, teacher and policy-makers should **find a balance between reliability and validity of assessment approaches according to the assessment purpose.** In this respect, employing both summative methods (which have stronger reliability), in combination with formative methods that strengthen the overall validity of the assessment approach, is important.

- **Defining key competences in terms of detailed and concrete learning outcomes is necessary for consistent assessment practices**. In an attempt to emphasise key competences and learning outcomes in education policy across Europe, many education stakeholders initially focused on 'traditional' key competences in a limited range of subject areas rather than looking at higher order skills throughout the curriculum. Nevertheless, learning outcomes defined in curricular documents, education standards or legislation increasingly cut across subject boundaries and refer to key competences. Making learning outcomes specific can help to adapt learning and assessment practices.

Recommendations

- **Key competences should be clearly defined in relevant policy documents and/or school curricula**, including a wide range of clear, concrete and detailed learning outcomes linked to teaching practice, taking into account cognitive and non-cognitive components.
- There is a need to transform teachers' attitudes and practices in the classroom in order to 'unpack' the competences to be taught and assessed.

- Teacher education (at ITE, induction and CPD level) should provide teachers with a common understanding of key competences and relevant assessment practices to assess student learning. **Integrating assessment practices at the ITE level and during practice** in schools can help increase teachers' assessment literacy and their skills to integrate assessment data into their teaching. **Collaborative learning environments can be effective tools to support teachers** at all stages of the teacher education continuum. School leaders and staff, school networks, municipalities, regional and national authorities can all be influential in supporting and monitoring teacher learning communities (TLCs) to reflect upon and enhance the effectiveness of assessment practices.

Recommendations

- **Teacher education systems should support teachers** at all stages of their careers **to effectively put into practice classroom assessment methods** that help to assess key competences and improve student learning.
- ITE, induction and CPD programmes should focus on a **broad range of complementary assessment methods for formative and summative purposes** to enhance teachers' assessment competences.
- **Collaborative learning environments, such as teacher learning communities (TLCs), should be promoted at all stages of the teacher education continuum** to support teachers' professional development and assessment competences.

- At the policy level, **there is a need for a clear vision and strategy for educational assessment where national and local approaches serve clearly defined purposes** with aligned assessment formats. There is a need to develop clear goals and reference points to guide student assessment at the classroom level.

Recommendations

- Schools should be provided with comprehensive guidance concerning valued learning outcomes in national curricula and standards.
- Education policy needs to strike a balance between formative and summative assessment, utilising the benefits of both, without over-relying on one particular method.

- Although standardised assessments have proved to be effective in assessing a number of key competences, they **can face challenges in evaluating non-traditional competences and transversal skills if narrowly focused solely on the content of the subject being assessed**. New technology has the potential to enhance the functionality of standardised assessments and their effectiveness in assessing cross-curricular competences.

Recommendations

- The assessment framework developed at school level should allow teachers to **draw on multiple sources of evidence** in order to form the best judgment on students' achievement level.
- Schools' capacity to use **the potential of ICTs for developing sophisticated assessment instruments should be better supported,** where standardised assessments could also be used for formative purposes.
- For standardised assessments to be effective, they should include the following items: **structure and content that reproduce real-life contexts authentically; multiple steps requiring a chain of reasoning and a range of competences; and a range of formats allowing responses that require different competences.**

- Performance-based assessment can serve both summative and formative purposes. They are seen as more effective than standardised tests in capturing more complex performances and processes. By employing a variety of techniques, such as holistic scoring rubrics, project-based assessment and portfolios[2], this approach can help assess a larger variety of competences and skills and better respond to individual learners' needs evaluating a learner's progress from his or her starting position.

Recommendations

- **Portfolios, holistic scoring rubrics and formative feedback can be helpful for formulating goals, monitoring student progress and assess broad competences**. The systematic development of these methods should be supported.
- **Clear definitions and a scale for competence development need to be included in performance-based assessment methods**, taking into account the specific context to ensure its effectiveness.

- Peer and self-assessment is considered to be a powerful tool to engage students in active learning, reflection and raise their motivation and academic standards. However, for these methods to be effective, there is a need for careful planning and accurate teachers' feedback to enhance students' self-regulated skills for them to meaningfully engage into assessment practices.

Recommendations

- **Peer-and self-assessment as both assessment and learning process** should become an essential component of classroom's integrated assessment practices.
- Criteria to judge performance in relation to national goals and learning outcomes should be better clarified and illustrated when using formative assessment methods, in particular peer- and self-assessment.
- There is a need for further **policy and research support for the development of toolkits on self-assessment practices, especially for non-traditional competences and transversal skills.**

- Although technology-based assessment is receiving more and more interest from researchers, practitioners and policy makers, **only first and second-generation e-assessments such as computer-based standardised tests, multiple-choice assessments and adaptive tests are widely applied at classroom level**. Not yet systematically implemented, innovative personalised embedded assessment practices, using complex tools such as learning analytics, game-based assessment and intelligent tutors provide ample opportunities to assess transversal skills, **such as problem-solving, critical thinking and creativity**.

[2] Portfolio assessment is a systematic and longitudinal collection of student work that shows his or her learning process, progress and performances (Frejd, 2013; Pepper, 2013).

Recommendations

- Although technologies play an important role in contemporary classrooms, they should be carefully implemented. **Using technology-based assessments for formative assessment purposes should be accompanied by effective feedback and scaffolding mechanisms**.
- **Policy makers should promote the introduction of innovative tools** in the classroom and integrate specific training into teacher education programmes. Relevant teacher training and transformation of traditional teaching and assessment approaches is crucial, for ICT tools to be effectively embedded into classroom practices and positively affect learning.
- Innovative assessment tools should build on **active students' engagement** in the assessment process for them to be effective.
- **Apart from standardised e-assessment methods, the pool of e-assessment toolkits in Europe is rather fragmented and their effectiveness needs further research.** Practitioners would benefit from a more systematic application of technology-based tools into the curricula across EU Member States.

INTRODUCTION

1.1. Background

1.1.1. Shift towards competence-based learning

In recent decades, education systems have shifted from a traditional content-based approach towards a more comprehensive and cross-cultural competence-based approach of education and training programmes (Camacho and Legare, 2016). Traditionally, the notion of competence was seen as being more pertinent to vocational education and training due to its direct link with the labour market (Halász and Michel, 2011; Tchibozo, 2011). Since the 1990s, however, the term 'competence' has also become increasingly used in school education. In many countries, curricula have been increasingly defined not only in terms of knowledge in different academic subjects, but also attitudes, skills, behaviours and values (Halász and Michel, 2011).

EU and OECD countries, and international organisations such as UNESCO, share the view that the globalisation of the economy, the rapid advancement of information and communication technologies, and the increasing role of knowledge and the associated social and cultural changes, are all fundamentally altering the way people live, work and learn (Takayama, 2013; Busca Donet et al., 2017). The 1996 UNESCO Delors report called for the re-orientation of school education to emphasise the all-rounded development and realisation of the human potential of individual learners (UNESCO, 1996). It proposed four pillars on which education should build: learning to know, learning to do, learning to live together, and learning to be. Tawil and Cougoureux (2013) argued that the pillar of 'learning to be' encourages more balanced school curricula, which takes into account not only individual cognitive-intellectual competences but also spiritual, moral, social skills and values. In addition, the pillar of 'learning to live together' called for the strengthening of a learning area devoted to promoting skills and values, based on the principles of respect for life, human dignity and cultural diversity.

Apart from explanations based on economic globalisation and employability, researchers have highlighted other factors that contributed to the development of this new education paradigm:

- the rapid pace of change and obsolescence of knowledge and skills prompted a need to prepare learners to cope with the change, and to question the consequences of these changes;
- the digitalisation of many spheres of social life prompted a need to prepare learners to use rapidly changing ICTs in a relevant way, and to raise awareness of the new ethical challenges brought about by social networks, while also adapting teaching/learning practices to young people's digital culture;
- growing inequalities, new forms of social exclusion, and the resurgence of xenophobia, racism and intolerance, prompted a need to increase the awareness of learners to the consequences of these trends, and to develop tolerance and openness for diversity for the sake of social cohesion, peace and democracy among and within countries (Halász and Michel, 2011; Tchibozo, 2011);
- the development of international education surveys, such as PISA, paved the way for renewed governance in education, and influenced domestic policy processes and debates, albeit to varying degrees depending on the political and cultural contexts of each education system (Michel and Pons, 2017).

In light of these explanations, the benefit of competence-based learning lies in its emphasis on the constructivist vision of learner development, in which each learner develops his or her own abilities and is able to mobilise and critically reflect on his or her knowledge (Tchibozo, 2011). Competence-based education focuses on student-centred learning, and represents a shift from the content of teaching and learning paths to 'learning outcomes' that are likely to be achieved and which can be used either in further educational pathways, or in the world of work and any other social context (Ravotto, 2011; Williams, 2015). Furthermore, competence-based education emphasises 'problem-solving, study and reflection skills, the use of tacit knowledge', beyond merely basic skills and competences (Travers, 2012, p. 45). In other words, it focuses not only on the content students are expected to become acquainted with, but also pays attention to students' ability to bring taught content to use (Rasmussen, 2013). Boillard argues that this approach is more robust than a traditional learning system, as it 'focuses on whether what needed to be learned was, in fact, learned' (2011, p. 57). In this respect, competence becomes a key concept.

1.1.2. Key competences for the 21st century

There is a consensus in Europe that individuals need to be prepared for the social and economic transformations of contemporary knowledge societies. Knowledge alone is not sufficient; a broader range of skills and abilities are needed to navigate a moving landscape characterised by the increasing importance of ICT, the decline of functional skills-based professions, and increasing competition (Grayson, 2014). This new set of skills is also complementary to active citizenship competences, or in other words, one's capacity to accept responsibilities, participate in group decisions, resolve conflicts, exercise critical judgment, be resilient and adapt to change (Dabrowski and Wisniewski, 2011; European Commission, 2017b).

These new skills and competences are often referred to as '21st century skills and competences'. This concept reflects the needs of the emerging models of economic and social development[3] (Ananiadou and Claro, 2009). Learning objectives are no longer the exclusive domain of the skills associated with a certain academic discipline. Education is expected to 'develop individuals' abilities to deal with problems and complex demands, mobilising psychosocial resources, knowledge, skills and attitudes previously acquired in learning situations similar to the contexts, which they will come across in their daily, professional or academic lives" (Tiana et al., 2011). In light of these developments, the need arose to identify and define those competences, and to integrate them into the curricula of European education systems so they could effectively contribute to all-rounded learner's development (Busca Donet et al., 2017). The Defining and Selecting Competencies (DeSeCo) project[4] aimed to address this need and attempted to define what would come to be seen as important competencies for the future. DeSeCo identified nine key competences under three broad categorisations: (1) acting autonomously, (2) using tools interactively, and (3) functioning in socially heterogeneous groups (OECD, 2005).

Drawing on findings from the DeSeCo project, the European Reference Framework on Key Competences for Lifelong Learning (European Parliament and Council of the EU, 2006) identified eight key competences which combine the knowledge, skills and attitudes necessary for personal fulfilment, active citizenship, social inclusion and employment in the 21st century (see Figure 1 below). EU key competences

[3] Alongside social and economic changes, there are three main theoretical influences that have shaped the development of key competences as a policy objective. These include a social perspective on education originally attributed to John Dewey; constructivist learning theories which argue for the importance of active learning; and ideas about workplace competences (Grayson, 2014).

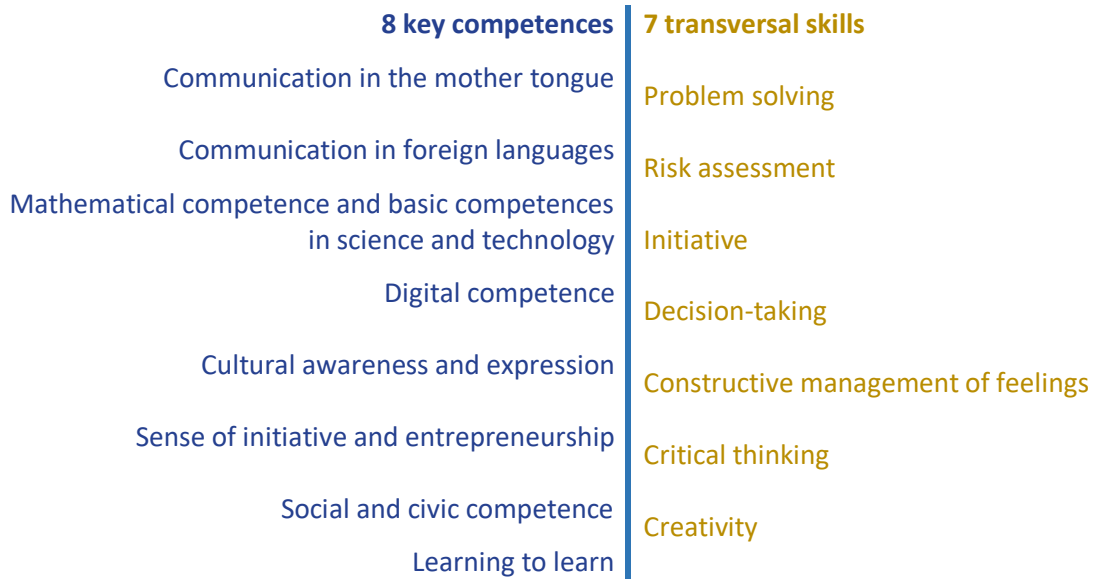
[4] Carried out by the Swiss Federal Statistical Office in collaboration with the OECD and the US Department of Education (OECD, 2005).

include both 'traditional' and primarily cognitive competences (such as mathematical competence) that are more quantifiable at national and international levels, alongside cross-curricular competences (such as digital competence, learning to learn, etc.). These competences overlap and intersect, and are further supported by transversal skills.

Key competences have become more prominent in European education systems in recent years, and most European countries have made significant progress towards incorporating them into national curricula frameworks, establishing standards for specific learning outcomes of students (OECD, 2013). However, there is a wide range of terminology used to refer to key competences in EU Member States, which sometimes reflects differences in emphasis and contexts (see further discussion in the section 3.1).

The key competence framework, with its emphasis on the application of knowledge in real world situations, represents a significant departure from 'traditional' content-based approaches, with subjects taught and assessed separately (Looney and Michel, 2014).

Figure 1. Key competence framework (2006).



Source: European Parliament and Council of the EU (2006).

To better reflect political, social, economic, ecological and technological developments since 2006, the European Commission has launched the review process of the Recommendation on Key Competences. This revision process aims to update the original key competences, and to help more people acquire the core set of skills necessary to work and live in the 21st century's knowledge-based societies; it has a special focus on promoting entrepreneurial and innovation-oriented mind-sets and skills. The New Skills agenda for Europe (European Commission, 2016a) aims to make better use of the available skills, and equip people with the new skills, that are needed in today's societies for 'employability and competitiveness'. Furthermore, the recent Commission Communication (2017a) on 'school development and excellent teaching' calls for the necessity to develop the resilience and ability to adapt to change in an increasingly mobile and digital society. At the same time, there is more emphasis among education stakeholders on the need to promote meaningful and equitable learning.

Recent statistics show that a large share of the EU population currently lacks a sufficient level of basic skills to be employable and take part in social life. Nearly 70 million European adults lack adequate levels of literacy or numeracy (European Commission, 2016a). A critical EU benchmark - that by 2020 the EU would count less than 15 % of 15-year-olds that are under-skilled in reading, mathematics and science - has not yet been reached. Discouragingly, the latest 2015 PISA data show that there has been a general increase in the share of low achievers in EU Member States across the three domains assessed by PISA (in science, reading and mathematics) (European Commission, 2016b). As stated in the 'New Skills Agenda', by 2020, 90 % of jobs will require at least some level of digital competence. Furthermore, in 2015 40 % of employers reported difficulties in finding candidates with the right skills, and many stressed the lack of transversal skills among job applicants (OECD, 2015a). In response to these trends, the revisions of the key competence framework will aim to further support Member States in developing education and training policies, with a focus on competence-based learning and on providing the basis for the development of assessment tools.

1.1.3. Challenges in assessing key competences and transversal skills

While the implementation of the key competence framework has been on the agenda of all Member States, with regard specifically to the assessment of 21st century key competences, studies conducted by Ananiadou and Claro (2009), Gordon et al. (2009), Pepper (2011), Halász and Michel (2011) reveal that this is still one of the weakest points in EU countries' efforts to integrate 21st century competences in school curricula. Gordon et al. (2009) referred to four different approaches to the assessment of key competences across 27 EU Member States: assessment of cross-curricular competences explicitly; assessment of cross-curricular competences implicitly; assessment of subject-specific competences; and assessment of knowledge rather than competence. While the first two approaches seem to be more relevant in assessing 21st century competences effectively, the latter two were still more common across most EU Member States (Voogt and Pareja Roblin, 2012).

Assessing key competences and transversal skills is challenging, since they refer to complex constructs that are not easily measurable. According to Eurydice (2012), national assessment tests in compulsory education settings in the EU do not cover all eight key competences. In most EU countries, national tests only covered traditional key competences (i.e. communication in the mother tongue, foreign language, science and mathematics), while competences such as learning to learn, digital competences, cultural awareness and entrepreneurship were not included in national assessment practices. For instance, even though more countries increasingly recognise entrepreneurship education as a cross-curricula objective, only a few show a more structured approach to comprehensively defining and assessing entrepreneurial learning outcomes (European Commission/EACEA/Eurydice, 2016). However, positive developments were observed in the assessment of civic competences as the number of Member States assessing these competences had risen from four in 2008, to 11 in 2012 and 14 in 2015 (European Commission/EACEA/Eurydice, 2012; OECD, 2015b). Assessments of students' social and emotional skills is generally administered in a formative format. In many countries, typical end-of-term students' assessments include evaluation of these skills; however, they still tend to be less transparent than academic achievement assessment (OECD, 2015b). The upcoming OECD "Longitudinal Study of Children's Social and Emotional Skills in Cities (LSEC)" aims to develop further recommendations and measurement tools for practitioners to better monitor and enhance social and emotional skills. The PISA 2018 innovative

domain "Global Competence" is another attempt to measure non-traditional skills that are of high importance in the modern world[5].

Although policies to support the assessment of cognitive competences have largely been implemented, they are often limited to the contexts provided by the subjects with which they are most closely associated, and also rarely include assessment of attitudes related to them (Pepper, 2011; Halász and Michel, 2011). Cross-curricular competences and transversal skills are harder to associate with individual subjects and to operationalise into specific learning outcomes (Ibid.)

Recent research also confirms that a broader range of non-cognitive competences, transversal skills and attitudes are usually not covered by 'traditional' assessment practices (Muskin, 2015), which as a result do not fully reflect the learning process. This could lead to assessing specific outcomes exclusively, rather than monitoring and providing feedback on the whole learning process (Bourke, 2015).

According to Pepper (2011), there are two major challenges for existing assessment practices: accessing information about, first, the components (knowledge, skills and attitudes) of each key competence and the way they are interlinked, and second, about the range of contexts in which they are applied. If only a few competences (or limited aspects of these competences) are assessed, such assessment practices might distort the curriculum, leading to the neglect of other competences. This in turn, may lead to a non-systematic, incidental development of skills and attitudes, which are more difficult to assess (European Commission, 2012a; Pepper, 2013).

Furthermore, Halász and Michel (2011) emphasise the need to change the thinking and practices of teachers in the classroom in terms of 'unpacking' the competences to be taught and assessed. Busca Donet et. al. (2017) highlight that other obstacles relate to 'academic traditions of basic education and teacher training, or the pressures of state educational institutions to achieve optimal results in international rankings' (p. 147).

1.2. Aims and research questions

This report reviews relevant European and international research to reveal how European education systems use classroom assessment tools to measure students' acquisition of key competences and transversal skills. The report highlights research and implementation gaps in assessment policies of key competences, and identifies promising initiatives and approaches that different schools are using to assess the full range of students' abilities and outcomes. This evidence can serve as a useful starting point for developing effective assessment and evaluation tools within the revised Key Competence Framework.

More specifically, the report aims to answer the following questions:

- What are the different roles and purposes of classroom assessment?
- How can classroom assessment better support '21st century learning and teaching', and what are the key conditions for aligning current assessment practices with the development of key competences?

[5] OECD Brochure, 'Global Competency for an inclusive world'. Available at please use direct link, not google search one!:
https://www.google.lt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=9&cad=rja&uact=8&ved=0ahUKEwjl58bToLvVAhXOJlAKHb-zBOIQFghlMAg&url=https%3A%2F%2Fwww.oecd.org%2Feducation%2FGlobal-competency-for-an-inclusive-world.pdf&usg=AFQjCNHoOQzF46V5s-RvB27oCDDcwy5wyQ

- What specific classroom assessment practices are effective tools to assess the full range of students' abilities and outcomes?
- What key recommendations can be made that can serve as important (first) steps to improve current policies on assessment?

1.3. Key concepts

Competences

This review follows the definition provided in the 2006 EU recommendation[6], where competences are defined as 'a combination of knowledge, skills and attitudes appropriate to the context', and more importantly the ability to apply them. Key competences are those which all individuals need for personal fulfilment and development, active citizenship, social inclusion and employment, and which constitute an integral part of an individual well-rounded competence-based education. Therefore, 'competence' is a broader concept than skill or competency[7] and it encompasses knowledge, competencies, skills, abilities, capacities, attitudes, values, attributes and qualities necessary for lifelong learning (Halász and Michel, 2011). It refers to the ability to successfully meet complex demands in varied contexts through the mobilisation of psychosocial resources, including knowledge and skills, motivation, attitudes, emotions, and other social and behavioural components. Key competences define teaching and learning as a process where not only skills and knowledge are acquired, but more importantly, where particular values, motivations, attitudes and dispositions that are deemed necessary for continuous, reflective and autonomous learning are to be nurtured (Takayama, 2013).

Non-traditional competences

The shift towards competence-based learning and changing realities in European societies led to the reconsideration of the traditional vision of competences as a fixed set of skills and abilities, emphasising their mobile and interconnected nature. The globalisation of the economy, the rapid advancement of information and communication technologies, and the increasing role of knowledge and associated social and cultural changes, ask for a new set of competences from learners. These competences, also often referred to as '21st century competences' or '21st century skills', are generally characterised as being (a) transversal (i.e. they are not directly linked to a specific field but are relevant across many fields); (b) multidimensional (i.e. they include knowledge, skills, and attitudes); and (c) associated with higher order skills and behaviours that represent the ability to cope with complex problems and unpredictable situations (Voogt and Pareja Roblin, 2012). These competences require new non-traditional assessment approaches to capture their dynamic and transversal nature[8].

The comparison of different competence frameworks made by Voogt and Pareja Roblin (2012) indicates that they all seem to agree on a common set of new non-traditional competences critical for today's world: collaboration, communication, ICT literacy, and social and/or cultural competences. Most frameworks also refer to creativity, critical thinking, productivity, resilience and problem-solving (Ibid). However, one should note that despite this general consistency, there are still no commonly agreed op-

[6] Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning (2006/962/EC). Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32006H0962
[7] Competency is usually referred to as general ability, being able to do something physically or intellectually (Tchibozo, 2011).
[8] Therefore, the authors use the concept 'non-traditional competences' as opposed to the traditional ones, for which there have always been comparatively reliable and valid instruments for their assessments.

erationalisation and categorisation procedures, which can lead to confusion and divergent interpretations of specific competences referred to as 'non-traditional competences'.

Formative assessment/assessment for learning

Formative assessment, also often referred to as Assessment for Learning (AfL), has been defined as 'activities undertaken by teachers — and by their students in assessing themselves— that provide information to be used as feedback to modify teaching and learning activities' (Black and Wiliam, 1998a). Therefore, formative assessment encompasses a variety of tools that provide feedback to teachers or students to help students learn more effectively.

Different learning theories may contain variations in the understanding of the concept of formative assessment. Cizek (2010) has synthesised current formulations into ten characteristics:

1. Requires students to take responsibility for their own learning;
2. Communicates clear, specific learning goals;
3. Focuses on goals that represent valuable educational outcomes with applicability beyond the learning context;
4. Identifies the student's current knowledge or skills and the necessary steps for reaching the desired goals;
5. Requires developments of plans for attaining desired goals;
6. Encourages students to self-monitor progress towards the learning goals;
7. Provides examples of learning goals including, when relevant, the specific grading criteria or rubrics that will be used to evaluate the student's work;
8. Provides frequent assessment, including peer and student self-assessment and assessment embedded within learning activities;
9. Includes feedback that is non-evaluative, specific, timely and related to learning goals and provides opportunities for the student to revise and improve work products and deepen understandings; and
10. Promotes metacognition and reflection by students on their work.

Source: Baird et al. (2014).

However, some researchers differentiate between assessment for learning and assessment as learning as two types of assessment conducted for formative purposes. 'Assessment as learning' reinforces and extends the role of formative assessment for learning and is 'a process through which pupil involvement in assessment features as an essential part of learning' (Dann, 2002, p.153). For further discussion of characteristics and the current debate on formative assessment, see section 2.1.2 and 2.1.3.

Summative assessment/assessment of learning

Summative assessments are usually defined as cumulative assessments that intend to capture what a student has learned, or the quality of learning, and judge performance against some standards (Dixson and Worrell, 2016). Summative assessments are generally 'high stakes' assessments and used to get a final judgment of how much learning has taken place — that is, of how much a student knows and has learned (Gardner, 2010). For a further discussion on summative assessments, see section 2.1.1, and on interrelation of different types of assessment, see section 2.1.4.

1.4. Methods and scope

This report primarily provides an overview of methods and approaches used to assess different types of key competences (defined in the 2006 EU framework) in secondary school classrooms. It explores how assessments can accommodate the full range of students' abilities, provide meaningful information on

learning outcomes, and support students and teachers in making use of ongoing feedback to individualise instruction and improve learning and teaching. Assessment is not a neutral activity, and throughout this review we also seek to depict how the literature has explored the relationships between assessment, learning and curriculum design.

The main source of information for this report is secondary data. The review is narrative in style, however, it also includes a more structured approach to the literature search on formative and summative assessments, as well as their links with specific key competences. The review draws on research that used a range of approaches including meta-analyses, quantitative and qualitative research, and analysing evidence from both small- and large-scale studies. It seeks to interpret different kinds of research, while giving due weight to findings with a particularly strong evidence base.

To identify relevant research for analysis, we applied both systematic and 'snowballing' search methods.

The systematic search of the literature was carried out in the following databases: *EBSCO Educational Databases; ERIC via EBSCO; JSTOR; SAGE Journals; ScienceDirect; Taylor & Francis Online;* and *Google Scholar.*

Apart from online databases, we also carried out searches in specific thematic journals, such as: *International Journal of Educational Research*; *Assessment in Education: Principles, Policies and Practices*; *European Journal of Education*; and *Journal of Curriculum Studies.*

Note: We acknowledge the limitations of the current searches, and acknowledge that there are other journals that could potentially provide relevant research evidence and be included in future reviews.

We conducted systematic searches of the literature on formative and summative assessment. Articles were retrieved according to the search terms 'formative assessment', 'assessment for learning', 'summative assessment', and 'assessment of learning'. Since a large number of meta-analyses and reviews were conducted on different types of school assessment practices and teacher assessment literacy to date (see e.g. Florez and Sammons, 2013; Baird et al., 2014; Gotch and French, 2014), we intended to complement these reviews with more recent research evidence on the topic. Therefore, our search was limited to articles written in 2014 and later.

This review was further complemented by structured searches of recent empirical evidence on the use of specific assessment practices for different types of key competences. We conducted these searches in the online databases and specific journals mentioned above, using a combination of terms referring to key competences and transversal skills, and four types of assessments identified during the narrative review (standardised assessment, performance-based assessment, peer- and self-assessment). We used the European Commission Staff Working Document (2012) on the 'Assessment of Key Competences in initial education and training' as a starting point for the search of examples of assessment practices, looking at research produced after 2012.

Since this review also provides a more practical overview of different examples of assessment practices that educators could use in the classroom, we included international comparative studies and practical guides in the analysis. Accordingly, we analysed relevant educational and learner data (OECD, TALIS, PISA, Eurydice) and work produced by European and international institutions and networks (e.g. European Commission, CoE, UNESCO, KeyCoNet, etc.), especially when it came to specific examples of assessment approaches. However, we acknowledge the limitations of the examples provided in terms of lack of empirical evidence on their effectiveness, where applicable.

2. THE ROLE OF CLASSROOM-LEVEL ASSESSMENT IN COMPETENCE-BASED EDUCATION

The way assessment and learning interact can either promote or hinder quality in education. This observation invites more research on assessment and learning to answer questions about how much children are actually learning at school, and how assessment can help enhance the learning process and ensure the acquisition of necessary key competences (e.g., Baird et al., 2014).

Discourse on the competences for the 21st century calls for accommodating the full range of learning outcomes into student assessment, including both cognitive and non-cognitive skills. To date, changes to curricula to incorporate key competences have not been fully reflected in changes to assessment practices (Pepper, 2013). Hill and Barber (2014) reflect on the need to adjust assessment as part of the ongoing process of re-thinking learning and teaching. This would build on the benefits of both standardised assessment and qualitative descriptions of various attributes that cannot be quantified, and go further to reflect a wider range of valued outcomes and allow for assessment of the full range of students' abilities.

This chapter summarises the existing debate on the different roles and visions of classroom assessment, and covers important principles of effective assessment practices.

2.1. Different functions of assessment

For decades, assessment has constituted one of the most controversial issues in education with respect to matters of theory, design, implementation, and educational policy. Many of the arguments surround what we assess, how we assess, and the ways in which information derived from assessments is used to shape educational practice (Lau, 2015; Pellegrino, 2016).

Hill and Barber (2014) emphasise that the field of educational assessment is currently divided and fragmented into differing and often competing philosophies, methods and approaches. The authors suggest that 'the resulting dichotomies have become the default basis for conceptualising and describing the field: quantitative versus qualitative; formative versus summative; norm-referenced versus criterion/standards-referenced; tests versus assessments; internal versus external; continuous versus terminal; measurement versus judgement; assessment of learning versus assessment for learning; etc.' (Ibid. p 25). Pellegrino (2016) adds that recent developments in cognitive and educational psychology have begun to reshape the vision of educational assessment. Among these is the realisation that 'assessment is fundamentally a process of reasoning from evidence that needs to be guided by theories, models, and data on the nature of knowledge representations and the development of competence in typical domains of classroom instruction' (Ibid). Other key understandings of assessment emphasise the multiple purposes assessment may serve and that the design of any assessment must be optimised for its intended purpose and use (Lau, 2015; Pellegrino, 2016).

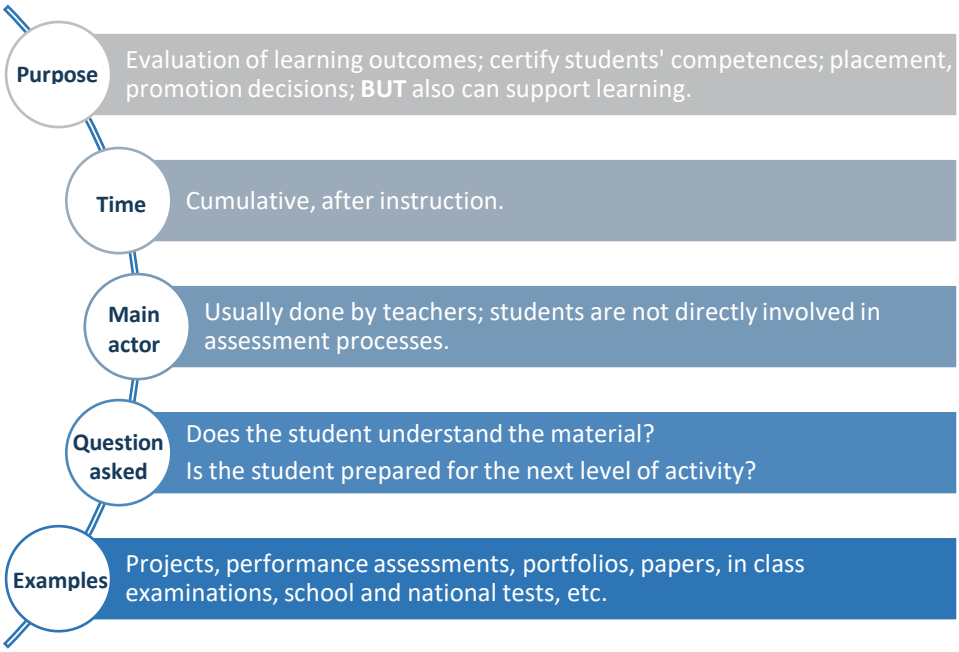
Assessment practices are also known to shape teaching and learning. That is, the focus of particular assessments at the national and school levels gives signals about what learning is important, and what aspects of learning merit and require more time and effort. Changes in assessment, therefore, can structure teaching priorities and methods, and in turn, impact what learners learn and how they learn. Assessing key competences, then, needs to both document learners' competences and simultaneously help develop them, by modifying teachers' practices and curricula focus (Pepper, 2013).

The following sections provide an overview of the three main visions of assessment discussed in the literature: assessment of learning, assessment *for* learning, and assessment *as* learning. It also includes a discussion of their relative advantages and limitations.

2.1.1. Assessment of learning

Assessment of learning, also referred to as summative assessment, involves reporting on learning, whether in classroom and school contexts or in wider national accreditation frameworks (Crossouard, 2011). Summative assessment methods rely on an extrinsic motivation for students, represented by marks, transcripts and diplomas. They are built on strategies to motivate students, provide information about student performance, serve to select or group students, and certify learning and award qualifications (Bennett, 2011; OECD, 2013, Dixson and Worrell, 2016). Summative assessment methods, in particular standardised testing, are also often considered to be more reliable than the alternatives, as they tend to be easier to interpret and are not influenced by the particular assessor or assessment (Pepper, 2013).

Figure 2. Characteristics of summative assessments



Source: authors' representation of Dixson and Worrell (2016).

However, the way summative assessments are used in the classroom can present both opportunities and challenges for assessing key competences and transversal skills. On the one hand, summative assessments are often believed to have positive effects on student learning and achievement, irrespective of students' prerequisites and backgrounds such as cognitive ability, socioeconomic status (SES) and gender (Thorsen, 2014; Thorsen and Cliffordson, 2012). International large-scale assessment tests such as the Programme of International Student Achievement (PISA) and the Trends in International Mathematics and Science Study (TIMSS) (it should be noted that neither are used as high stakes tests), as well as various national standardised tests, have shown the effectiveness of testing basic cognitive skills, and are considered reliable and valid. Furthermore, many stakeholders believe that more frequent testing

and grading has the potential to increase student performance by providing motivation to improve their grades and scores (Klapp, 2015).

The increased focus and promotion of summative assessments, especially in the early 21st century, has also been triggered by the traditional, dominant view of assessment underpinned by trait theory. This, simply put, views human capacity and intelligence as measurable traits (Taylor, 1994; Lau, 2015). As a result, this view prioritises standardisation, reliability and limited dimensionality; the emphasis is on ensuring that assessment is carried out and scored and interpreted in the same way for everyone. This focus on reliability has often resulted in sacrifices in the validity of assessment (Lau, 2015).

On the other hand, such a strict and limited view of summative assessment methods carries the risk that students will focus their learning, and teachers their teaching, on only those areas that are rewarded. This narrow approach to learning encourages surface learning, and can lead to decreasing student motivation to learn when the reward is no longer present (OECD, 2013; Lau, 2015). Therefore, if summative assessment does not explicitly incorporate key competences, it can limit curricula and learning (European Commission, 2012b). Furthermore, this may reduce enjoyment of learning and decrease students' focus on long-term goals (EPPI, 2002, Klapp, 2015). In addition, contrary to the frequent assumptions outlined above, Klapp (2015) found that grading may have a different impact on students depending on their background. The challenges associated with implementing summative assessments in the classroom are summarised in the table below.

Table 1. Challenges associated with implementing summative assessment in the classroom

Definition/focus	Limited view of summative assessment – as only assessing knowledge, undermining the potential of the summative assessment practices to support learning, especially when using innovative tools, can lead to the fragmented view of learning and curricula, teaching competence that can be easily quantified (Pepper, 2013; Lau, 2015).
Effectiveness for quality learning	The strong impact of summative assessment on teaching and learning has been widely reported. In many contexts, summative assessment dominates what students are oriented towards in their learning – this is typically described as the "backwash effect" of summative assessment (see e.g., Baartman et al., 2006). The marks, transcripts and diplomas that summarise student performance can be seen as rewards for student effort and achievement, which provide an extrinsic motivation for learning (Sjögren, 2009). However, recent research findings suggest that the exclusive use of extrinsic motivation may be problematic as it is often too closely related to the reward. If used in isolation, summative assessments with high stakes for students may in fact encourage surface learning approaches, reduce enjoyment of learning and decrease student focus on long-term goals (European Commission, 2012b; Pepper, 2013; Lau, 2015).
Measurement	Some research findings suggest that summative assessment methods, including standardised tests, can produce a bias in the results of students' performance, depending on the background of learners (Klapp, 2015). A test-takers' performance may not only be affected by the content and difficulty of the test, but also by their behaviour and emotional and motivational concerns related to test-taking (Stenlund et al., 2017). Furthermore, in the summative assessments used for selection and qualification purposes, it is important to introduce carefully controlled conditions to ensure fairness of the assessments. However, these conditions can be an obstacle to assessing students in real life contexts, which is often more appropriate due to the mobile and multi-dimensional nature of key competences (Pepper, 2013).
Teacher prepared-	The criterion-based student summative assessment often requires a better moderation

ness	process to ensure the consistency of student marking by teachers, a key area to guarantee fairness of student marking across schools in a given country. Therefore, the effective implementation of summative assessment also requires comprehensive teacher training (Crossouard, 2011; Pepper, 2013).

Source: compiled by authors based on the literature review.

The assessment and instruction should, therefore, focus on how to better support successful learning processes rather than achieving or receiving a grade. It is important that summative assessments are carefully controlled to ensure that their consequences are fair (European Commission, 2012b; OECD, 2013). Bennett (2011) further emphasises that summative assessment can fulfil its primary purpose of documenting what students know and can do, but, if carefully crafted, can also successfully meet the secondary purpose of support for learning.

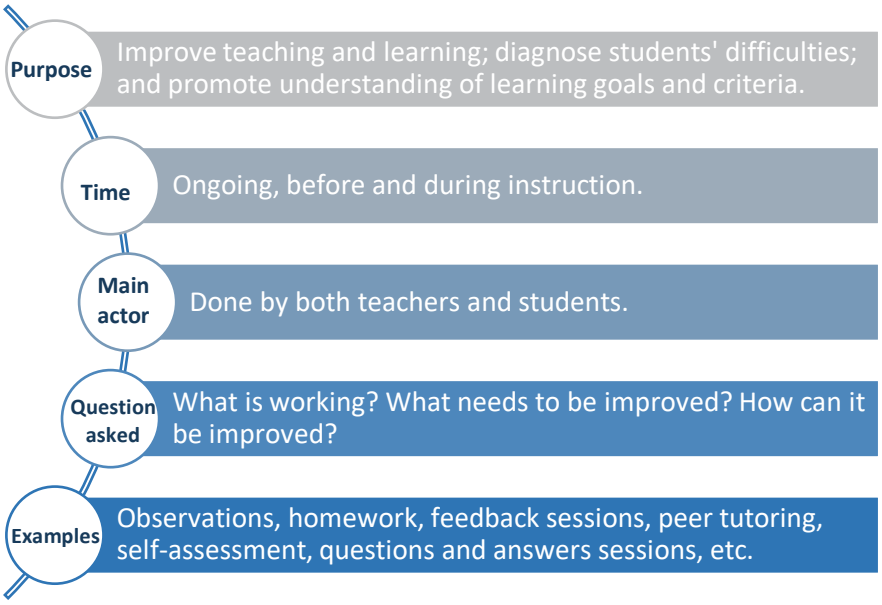
In sum, reporting learners' development of key competences can be a challenge to existing summative assessment practices, which often focus on documenting knowledge and skills limited to specific subject contexts, while non-traditional competences are often transversal and multi-dimensional in nature. Therefore, it is important that innovative summative methods are used to bring assessments into real-life contexts of students (Pepper, 2013). Chapter 4 looks in more detail at how certain summative assessment methods, such as standardised tests, attitudinal questionnaires and performance-based assessments, can be used to assess the development of key competences.

2.1.2. Assessment for learning

In contrast to summative assessment, assessment for learning is often described as providing direct support to students for improving their learning and teachers' practices. Black and Wiliam (1998a) underline that an assessment is considered to be formative when the assessment information is used to improve students' performance. This view places students in the centre of the assessment process and use of assessment results. Nevertheless, teachers have a key role in providing feedback, information about students' performances, and in particular in setting a goal or learning target (Brookhart, 2001; De Luca et al., 2016). The quality of the feedback is a key feature in any formative assessment process (Black and Wiliam, 1998a), yet the learner also becomes an important actor in the assessment process. Successful formative assessment consists of a sequence of two actions: the recognition by the learner of a gap between his or her current state and the desired goal, and, importantly, the action taken by him or her to close that gap (Black and Wiliam, 1998a). In this sequence, the teacher has a key role in interpreting the gap and communicating a message about it to the student, based on assessment information.

Formative assessment can allow students to become 'formative decision-makers' (Brookhart, 2011, p. 4). In the context of formative assessment methods, students may use descriptive information (at an adequate time in the learning process) to make productive decisions about their own learning (Ibid.). Evidence shows that assessment for learning can lead to significant achievement gains, in particular for lower achieving students, helping to reduce the inequity of student outcomes (Black and William, 1998a; Hattie and Timperley, 2007; Hattie, 2009). However, some research also warns that claims on the effectiveness of formative assessment should be considered with caution, due to limitations of the research methodology used by earlier studies (see e.g., Bennett, 2011; Baird et al. (2014). Nevertheless, and while limited in scope, recent empirical evidence has shown the positive impact of formative assessment methods (such as peer- and self-assessment) on teaching and on students' learning outcomes, in particular on the development of non-traditional competences and transversal skills (e.g. Baird et. al, 2014; Cornu et al., 2014; OECD, 2015b). At the same time, the success of formative assessment policies depends significantly on their effective implementation (Black and William, 1998a).

Figure 3. Characteristics of formative assessments



Source: authors' representation of Dixson and Worrell (2016).

Assessment for Learning (AfL) programmes[9] have also been growing internationally (Baird et al., 2014). Formative assessment is now seen as an integrated part of the teaching and learning process, rather than as a separate activity occurring after a phase of teaching. These methods include effective feedback; questioning techniques; comprehensive approaches to teaching and learning featuring formative assessment; and student self- and peer-assessment (Looney, 2011a). These are further discussed in Chapter 4. Integrated in the curriculum of a growing number of countries, AfL is seen as a tool to enhance, not only measure, the achievement of the curriculum, focusing on learning processes and individual progress (OECD, 2013). However, despite their recognised relevance in providing high-quality teaching and improving learning, formative assessment practices face a number of challenges that can impede their large-scale implementation, mainstreaming and sustainability (Hopfenbeck et al., 2015; Looney, 2011). The table below summarises the challenges associated with implementation of formative assessment methods in the classroom.

Table 2. Challenges in implementing formative assessment in the classroom

Defining assessment	The term 'formative assessment' is not used consistently in the literature. Some authors see all classroom assessment as formative and discuss summative assessments primarily in terms of external assessments. Some authors agree all classroom assessment can be formative, but only if students use the assessment information for formative purposes (Brookhart, 2001; Bennett, 2011). Furthermore, while the term 'formative assessment' has garnered much attention, the underlying theory is often forgotten, and is frequently reduced to a series of mechanical steps or tests that poorly resemble what formative assessment is ultimately about (Shepard, 2005).

[9] Traditionally, AfL has been closely associated with formative assessment because practices such as questioning and providing feedback help 'form' or 'shape' student learning. However, summative assessment strategies (e.g., portfolios) can also be effectively applied in AfL, when used for formative purposes, e.g., to identify a lack of understanding (e.g. in a particular area of the syllabus) and subsequently targets are set to rectify this (see e.g., Baird et al., 2014).

Effectiveness	The quality of formative assessment depends on the strategies teachers use to elicit evidence of student learning related to goals, with the appropriate level of detail to shape subsequent instruction (Heritage, 2010; Herman et al., 2010). However, evidence shows that in some contexts teachers still tend to develop only superficial questions to probe student learning, and provide only general feedback. Teachers may have difficulty in interpreting student responses or in formulating next steps for instruction (Herman et al., 2010; Florez and Sammons, 2013). Furthermore, one of the challenges for implementing formative assessment methods, where students are the main assessors, is that many students are not self-regulated enough to do it well, which in turn can lead to inaccuracy (Brown et al., 2015).
Measurement	Educational measurement usually involves four activities: designing opportunities to gather evidence, collecting evidence, interpreting it, and acting on interpretations. Bennett (2011) argues that the literature on formative assessment pays scant attention to the interpretation of observations, which represents entirely inferential process in this case. Formative inferences are often subject to uncertainty and subjectivity, related to perceptions associated with gender, race, ethnicity, disability, etc., and therefore can be biased depending on the teacher implementing it (Bennett, 2011).
Teacher preparedness	While many teachers agree that formative assessment methods are an important element of high quality teaching, there can be many structural barriers to integrating formative assessment into their teaching practice on a constant basis. This includes large classes, extensive curriculum requirements, and a lack of effective teacher education programmes to support teachers' capacity and professionalism in formation assessment (Looney, 2011, Florez and Sammons, 2013). Furthermore, many teachers feel that they lack sufficient preparation to effectively put into practice assessment methods that they have acquired at the ITE level, in particular regarding formative assessment (Stiggins, 2005).
Coherence with the overall system	The effectiveness of formative assessment is limited by the nature of the larger system in which it is embedded and, particularly, by the content, format, and design of the accountability systems and the role of summative assessments (Bennett and Gitomer, 2009; Baird et al., 2014).

Source: compiled by authors based on the literature review.

If these limitations are addressed, formative assessment has greater potential in supporting the development of key competences, especially the ones that are not easily measurable by traditional assessment practices. Teachers' feedback, and peer- and self-assessment, are better able to support 'deep-learning' (see Crooks, 1988), promoting active interaction and continuity of learning experiences (Dixson and Worrell, 2016), and therefore, in nurturing transversal skills such as critical thinking and learning to learn (Pepper, 2013; Sargent, 2014). Moreover, since formative assessment happens during the process of learning, it has more potential to assess dispositions that can only be demonstrated in action, and therefore, be set in a meaningful context (Hipkins, 2007) (see further discussion in Section 4.3).

2.1.3. Assessment as learning

Some researchers distinguish a third approach towards assessment: 'assessment as learning' (see for example Dann, 2002, Lee, 2013, Drake et al., 2014, and Hayward, 2015). It reinforces and extends the role of formative assessment for learning and is 'a process through which pupil involvement in assessment features as an essential part of learning' (Dann, 2002, p.153). The student is regarded as the critical connector between the assessment and learning process. Emphasising assessment as a process of

metacognition (Earl and Katz, 2006), assessment as learning encourages students to monitor and practice self-regulation over their thinking processes, and stresses the importance of fostering students' capacity over time to be their own assessors (Ontario report, 2010; Lee and Mak, 2014). Self and peer-assessment practices are often mentioned as examples of assessment as learning. These approaches help students take more responsibility for their learning and monitoring of future directions (Earl and Katz, 2006). Thanks to these functions, assessment as learning can be especially beneficial for developing and assessing transversal skills, such as initiative, decision-taking, critical thinking, etc. (Earl, 2013) (see further discussion on the use of specific methods in Section 4.3).

However, when designing classroom assessment practices, one needs to take into account the existing challenges for implementing formative assessment, as summarised in Table 2 above, and to ensure that there are clear assessment criteria and guidelines for teachers and students engaging in peer- and self-assessment.

2.1.4. Integrated approach towards assessment

Although there are certain limitations in implementing different types of assessment in isolation, using them in synergy and combination can allow the benefits of all three types to offset their individual challenges (see e.g., Earl and Katz, 2006). The reviewed literature suggests that integrating different approaches to assessment can allow for finding a balanced and consistent practice for assessing key competences (Lau, 2015). Traditionally, summative assessment is the predominant kind of assessment in most classroom activities (Lee and Mak, 2014, Drake et al., 2014). While assessment for learning has been used at various stages of the teaching and learning process, it tends to be informal and implicit. In many traditional classrooms, systematic assessment as learning is almost non-existent (Ibid.). Researchers emphasise that to prepare students to meet the demands of an information and knowledge-based economy, which requires students to work towards higher order thinking, autonomy and self-management, there is a need for an integrated focus on assessment. Such an integrated focus would include a range of formative and summative assessment approaches that complement each other, in order to provide the adequate level of challenge and support to each student (Earl and Katz, 2006; Earl, 2013).

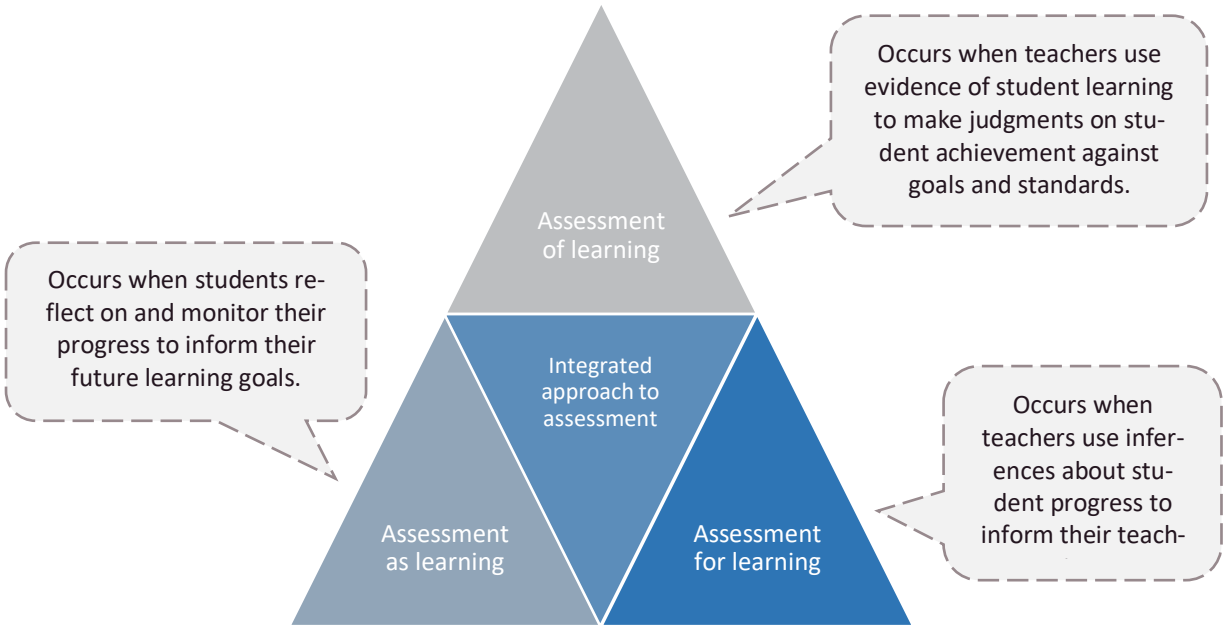
Some argue that the termination of certain summative assessment methods, such as marks and grade retention (and by relation, tracking), would allow students to focus on the learning process with less pressure and anxiety (GRDS, 2012). Terminating grading and grade retention systems can be seen as a way to improve learning through assessment: 'Assessment, which is indispensable to learning, does not classify, it validates whether learning is realised, or not' (GRDS, 2012, p. 30). However, it would be wrong to assume that there is a clear-cut 'good' and 'bad' divide between formative and summative assessment; summative assessment, when carefully designed, can also have positive impacts on student learning (Bennett, 2011; Lau, 2015). Indeed, the literature indicates that summative assessments in combination with other types of assessments can be effective in assessing certain key competences.

Numerous countries have aimed to integrate formative assessment methods with summative external approaches to build comprehensive and consistent assessment frameworks (Looney, 2011). Reforming summative assessment methods based on grading, by integrating them with formative methods, could also help to focus the learning process on the obligation to offer the same potential progress to all students, focusing on the reality of their capacities, and not on constructed deficits (Terrail, 2016). The integrated approach to assessment also allows for more flexibility when documenting the development of transversal and multidimensional key competences. These challenges highlight the need to embed formative assessment methods within school education systems' evaluation and assessment frameworks, including a range of complementary methods (Mottier Lopez, 2015). They also demonstrate the

importance of supporting investments in teacher initial teacher education (ITE) and continuous professional development (CPD) to strengthen the quality of assessment (Looney, 2011).

One example of integrating different types of assessment comes from the King's-Medway-Oxfordshire Formative assessment Project, in England, which has sought to address tensions between formative and summative assessment by making them symbiotic (William, et al., 2004). They have made assessment continuous, integral to pedagogic discourse and classroom interactions, and actively involved both teachers and learners (see Figure below) (Crossouard, 2011). Inspired by the English experience, Scotland has introduced the comprehensive assessment framework 'Assessment is for Learning'[10], promoting active learning and the re-distribution of learning power.

Figure 4. Integrated vision of assessment



Source: authors' representation of Crossouard (2011).

A wider range of assessment approaches can provide a more comprehensive account of learning, in particular for children with additional learning needs. Since learners with high needs benefit from a range of different teaching approaches, diverse assessment approaches has the potential to better inform and summarise their learning, as well as capture the multi-dimensionality of the learning process, as concluded by (Bourke and Mentis (2014) based on teachers' perceptions analysis.

Hayward (2015) further argues that assessment principles to improve learning are additionally relevant for entire national systems, and linking assessments to specific functions with the prepositions 'of', 'for' and 'as', without considering their integrated nature, may divert attention away from the key construct 'assessment is learning'.

[10] See further: http://www.gov.scot/Publications/2005/09/20105413/54156

2.2. Validity, reliability and equity of assessments

In addition to instructional and practical considerations, the overall technical quality of assessment practices is crucial. Technical quality refers to factors such as whether the assessment measures what its developers claim it measures, and whether it provides consistent and meaningful results across students, tasks, and versions (Soland et al., 2013). This is covered by three criteria: validity, reliability and equity.

Validity

In education assessment, validity is a central concept as it provides an overarching criterion for the quality of assessment strategies (Pepper, 2013). It is, therefore, the foremost technical consideration for any assessment, including the assessment of key competences. Validity is often described as 'a judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment' (Messick, 1989, p. 13).

Much of the literature on the validity of educational assessments has tended to focus on assessments designed for large-scale and often high-stakes purposes in the context of educational policy. Only in the last decade have there been more attempts to discuss assessments intended to function much closer to classroom teaching and learning, including the design and validation of such assessments (Pellegrino et al., 2016). Irrespective of the level at which an assessment is intended to function, or of the framing of the assessment activities in terms of theories and models of learning and knowing, issues of validity remain paramount. Pellegrino et al. (2016) argue that three components of cognitive[11], instructional[12], and inferential[13] validity need to be considered, regardless of the assessment's intended purpose and/or level of functioning relative to ongoing processes of teaching and learning. These elements of validity could be analysed using multiple data sources, for example, expert analyses, student cognitive protocol studies, teacher surveys and logs, etc. The authors call for the application of clear standards of validity to a variety of assessment tools and materials that are frequently used in the classroom, to better understand whether they support or undermine effective practices of teaching and learning (Ibid).

Reliability

Validity as a criterion for evaluating assessments often encompass other important but narrower criteria, such as reliability and equity (see, for example, Morris, 2011). Reliability is '…often defined as, and measured by, the extent to which the assessment, if repeated, would give the same result' (Harlen, 2007, p. 18). Although validity and reliability are often seen as being in tension with one other, reliability is in fact one aspect of the broad concept of validity. For example, a test can be made more reliable by limiting its question types and response formats, making it more straightforward and easy to interpret. However, such a test would provide a narrow picture of the key competences needed for lifelong learning, undermining its overall validity. Alternatively, day-to-day teachers' feedback and peer-assessment

[11] This component addresses the extent to which an assessment taps important forms of domain knowledge and skill in ways that are not confounded with other aspects of cognition such as language or working memory load (the construct)(Pellegrino et al., 2016).
[12] This component addresses the extent to which an assessment is aligned with curriculum and instruction, including students' opportunities to learn, as well as how it supports teaching practice by providing valuable and timely instruction related information (Ibid.).
[13] This component is concerned with the extent to which an assessment reliably and accurately yields model-based information about student performance, especially for diagnostic purposes (Ibid.).

can provide a broader picture of learners' competences and skills. However, if the formative assessment is subject to biased judgment (e.g., teachers or peers' preferential attitudes), this may compromise the reliability of assessment, as the interpretation and weighting of observations would differ (Pepper, 2013).

In practice, though, assessments can find a balance between reliability and overall validity according to the assessment purpose. Accordingly, assessments designed for summative purposes would emphasise reliability, assessing a limited number of performances and range of the curriculum. Assessment designed for formative purposes would emphasise overall validity, assessing more performances in a wider range of contexts (Pepper, 2013). This literature review explores some of the potential for assessment to move beyond this dichotomy (see Chapter 4 for further discussion of different assessment methods).

Equity

Some authors argue that equity should be the third crucial feature of assessment (see e.g., Binkeley et al., 2010; Kirova and Hennig, 2013). While ensuring equity, assessment often involves striking a balance between validity and reliability, and a variety of assessment approaches, in terms of design and functions (Looney and Michel, 2014). It is important that assessments allow all students to demonstrate what they know and can do without being unfairly disadvantaged by individual characteristics that are irrelevant to what is being assessed (Binkley et al., 2010). Equity emphasises the social nature of assessment and highlights the need to consider differences that, while not the focus of an assessment, could nevertheless influence the assessment. For example, assessments can be developed or modified to ensure that when a learners' disability is not relevant, it is not assessed. Furthermore, the growing diversity of European classrooms call for the development of more creative and flexible ways of assessment that can take this diversity and the different ways diverse students learn, into account. While there is research to show that learners with diverse needs are often recognised to be taking multiple and different pathways in their learning, less research focuses on the diversity of assessment practices used to measure, support and facilitate these students' learning (Bourke and Mentis, 2014). Bourke and Mentis (2014) highlight the importance of using an integrated assessment framework to accommodate the needs of diverse learners. Kirova and Hennig (2013) also emphasise the need for assessment practices to be linked to a socio-cultural theory of learning, and acknowledge that there are diverse ways of knowing and validating this variety. Assessment, if not done with equity in mind, privileges and validates certain types of learning and evidence of learning over others, can hinder the validation of multiple means of knowledge demonstration, and can reinforce the feeling of alienation and lack of sense of belonging within students (Montenegro and Jankowski, 2017).

In addition to these three criteria, the literature also highlights *transparency*, referring to the extent to which all participants – teachers/trainers, learners, assessors, parents, administrators and end-users - know and understand what is required in the assessment. Additionally, *freedom from bias* and *usability* refers to how policy makers, school leaders, teachers, parents and students make sense of and respond to the assessment results (OECD, 2013).

Key messages

- **The field of educational assessment is currently divided and fragmented into differing and often competing philosophies, methods and approaches**: for example, quantitative versus qualitative, formative versus summative, measurement versus judgement, etc.
- Research has demonstrated that there is no single way or universal approach towards assessment of key competences. Different types of assessments, be it for summative or formative purposes, have limitations when implementing them in particular contexts. Therefore, **an integrated approach is needed** to capitalise on the benefits and opportunities of both, in order to ensure meaningful and inclusive learning for all students.
- Irrespective of the different functions of assessment, the literature emphasises that the primary role of assessment is to improve learning and ensure that a set of key competences and transversal skills is developed in a comprehensive way. This principle should be relevant for entire national education systems. Therefore, linking assessments to specific functions with prepositions 'of', 'for' and 'as', without considering their integrated nature, may divert attention away from the key construct that 'assessment is learning'.
- When designing assessment practices, teachers and education policy makers need to consider their **validity, reliability and equity**. To ensure equity, assessment framework needs to strike a balance between reliability and validity and incorporate variety of assessment approaches, in terms of design and functions.

3. ALIGNING ASSESSMENT PRACTICES WITH KEY COMPETENCE DEVELOPMENT: MAIN ISSUES TO CONSIDER

This chapter looks at the importance of aligning assessment practices with other important elements of the education process. More specifically, it provides an overview of supporting conditions for the successful assessment of key competences that need to be in place to ensure coherent assessment practices for '21st century learning'. Some of these conditions include:

- the importance of a clear definition of student competences in terms of learning outcomes, and their reflection in the school curriculum;
- teacher education, including ITE, the induction stage, and CPD, providing teachers with a common understanding of key competences and assessment guidance throughout their careers;
- collaboration mechanisms in the form of teacher learning communities (TLCs), with the aim to improve assessment practices to align them better with key competences; and
- the use of (summative and formative) assessment results for in-school and outside of school purposes.

3.1. Defining learning outcomes in the curricula

Classroom assessment is crucial for the development and use of key competences in school education, increasingly defined as the ability to apply specific learning outcomes adequately in a defined context (Cedefop, 2014a; Pepper, 2011; 2013). Cedefop (2014a, pp. 164-165) provides two interrelated definitions of learning outcomes:

- as 'statements of what a learner knows, understands and is able to do on completion of a learning process, which are defined in terms of knowledge, skills and competence'; and
- as 'sets of knowledge, skills and/or competences an individual has acquired and/or is able to demonstrate after completion of a learning process, either formal, non-formal or informal'.

These two definitions are interlinked in the relationship between *intended* and actually *achieved* learning outcomes. *Intended* learning outcomes are desired targets, while *achieved* learning outcomes can be identified at the end of the learning process, notably through assessment (Cedefop, 2016).

However, there is a lack of clarity in the intended learning outcomes for teaching key competences in school curricula, which in turn challenges their effective assessment. Although eight key competences were defined in the EU's 'Recommendation on Key Competences for Lifelong Learning' in 2006, they are not described in terms of specific learning outcomes that could guide daily teaching and learning (Gordon et al., 2009).

Classroom assessment provides information about learners' progress towards learning outcomes, and accordingly should help in adapting teaching and learning more effectively (European Commission, 2012). This underlines the importance of operationalising key competences in terms of concrete learning outcomes, in order to provide a good basis for consistent assessment practices, striking a necessary balance between the level of detail of learning outcomes defined in policy documents and the scope for interpretation left to teachers and learners in practice.

The concept of competences is increasingly used across EU countries to cover both cognitive and non-cognitive aspects of learning (OECD, 2013). Sets of key competences are used to describe expected learning outcomes in teaching and learning settings, notably through concepts such as basic competences, core skills, key skills, or '*socle commun*' (Halász and Michel, 2011; Pepper, 2011). However, there are numerous diverse definitions of what constitutes key competences in different Member States, es-

pecially when it comes to their non-cognitive components and transversal skills. This diversity of definitions challenges the implementation of recommendations and the modernisation of assessment practices across all sectors of education (UNESCO, 2015).

The distinction between 'traditional' key competences (communication in the mother tongue, communication in foreign languages, mathematical competence, and basic competences in science and technology) and 'non-traditional' competences (digital competence, cultural awareness and expression, sense of initiative and entrepreneurship, social and civic competence, and learning to learn) also needs to be further developed (Gordon et al., 2009; Pepper, 2011). Current assessment policies across Europe tend to support and evaluate learners' competences in 'traditional' key competences in a limited range of subject areas, while there is not yet enough support for the development of non-traditional key competences and transversal skills in a broader range of contexts, across curriculum (Pepper, 2011).

In parallel, there is a growing emphasis on learning outcomes in education policy and as the basis of framework documents such as curricula, programmes, standards or syllabi (Cedefop, 2009; 2016; Proitz, 2015) (see examples from EU countries in Box 1 below). Although they remain mostly structured in terms of subject matters or areas, learning outcomes increasingly cut across boundaries and refer to broader key competences (European Commission, 2012). National curricular documents, education standards, or legislation, can provide good references for defining learning outcomes. In other cases, learning outcomes are described in more detail at the school level for use in classrooms (Cedefop, 2016).

Box 1. Examples of EU countries using learning outcomes to define and assess key competences

In the **Czech Republic**, the Framework Education Programme (FEP) identifies a range of key competences similar to the EU Framework, each of which is expressed as a set of specific goals which define the level of competence that learners should develop by the end of compulsory schooling.

In **Flanders (Belgium)**, the shift towards competence-based education has aimed to increasingly focus on learning outcomes. Since 2013, all learning outcomes are formulated based on competences derived from the European key competences framework.

In **Ireland**, five "key skills" were developed based on the European key competences framework (information processing, critical and creative thinking, communicating, working with others, and being personally effective), and were been broken down into main elements and intended learning outcomes. The competences are embedded in the assessment system, which has been reviewed at all education levels based on the new framework.

In **Romania**, since 2011, learning outcomes are assessed by final exams based on the eight key competences defined in the EU framework, aiming to increase coherence with the National Qualification Framework (NQF).

In **Scotland**, the last reform of the school curriculum (*Curriculum for excellence*) was accompanied by guidance about associated pedagogical and assessment approaches. The curriculum includes a wide range of learning outcomes linked to teaching practice, including cognitive components and personal and affective dimensions, not all of which are amenable to traditional assessment methods.

Source: Cedefop (2016); Livingston and Hutchinson (2016); Pepper (2011).

Beyond skills or competences, schools have a key role to play for developing young people's attitudes, character traits and dispositions that will support them in confronting emerging societal challenges, such as resilience, adaptability, entrepreneurship, or sensitivity to cultural and personal differences. Cultivating such outcomes can be a more complicated process than developing skills and understanding, because it means engaging students in situations where these qualities matter and can be experienced and reflected upon (Hill and Barber, 2014). Moreover, in the EU the assessment of attitudes that support the development of knowledge and skills is still not widely developed (Pepper, 2011).

This new understanding of the concepts of key competences and learning outcomes put additional demand on assessment practices (Gordon et al., 2009). Defining learning outcomes as a basis for planning

and assessing learning implies making them more specific in terms of 'sub-competences, and in turn, statements of learning outcomes' (Pepper, 2011, p. 340). This questions the way in which assessment can make learning meaningful and provide feedback, which can in turn inform decisions about the curriculum and provide evidence of the outcomes of learning and teaching. According to Pepper (2011), the challenge facing EU education systems is to '[specify] key competences in sufficient detail to plan and assess learning – but not with so much prescription that the process of developing competences is reduced to a series of procedural tasks that are completed without full appreciation of underlying concepts' (Ibid., p. 341). Assessment practices that are insufficiently aligned with key competence-based curricula or that cannot measure complex performances may undermine educative innovations aimed at learner-centred approaches and competence-based learning. In addition, the lack of agreement on how to assess 'softer' transversal skills, such as creativity and initiative, may also mean that teachers pay less attention to them (Grayson, 2014).

3.2. Teacher education

TALIS 2013 results reveal that 11.6 % of teachers in participating OECD countries still report a high level of need for professional development in the area of student evaluation and assessment practice (OECD, 2014)[14]. Research shows that teachers feel that they lack sufficient preparation to effectively put into practice assessment methods that they have acquired in ITE, struggle to interpret assessment policies and implement assessment practice aligned with contemporary assessment methods (De Luca et al., 2016), in particular formative assessment (Stiggins, 2005). Several empirical studies from the US[15] show that ITE does not provide future teachers with an adequate understanding and appreciation of assessment, and that student teachers do not follow many assessment practices recommended in ITE in their teaching practice (see e.g.: Begeny and Martens, 2006; Campbell and Evans, 2000; Doolittle, 2002). Campbell and Evans (2000) suggest that ITE programmes should be better designed to understand student teachers' 'attitudes, beliefs, behaviours and environmental pressures as related to the practice and use of classroom assessment' in order to reduce the discrepancy between ITE courses and their application by future teachers (p. 354). Schneider and Bodensohn (2017) underline that despite the complexity of facets and tasks which assessment encompasses in practice, 'formal learning opportunities specifically designed to develop future teachers'[16] assessment competence in the form of specific lectures or courses within teacher training programmes [in Europe] are often scarce or even non-existent' (p. 128).

In their systematic review of assessment literacy[17] measures in the US and their adequacy for use in teacher evaluation, Gotch and Fench (2014) highlight the need for further research to connect teacher assessment literacy to student outcomes. Research has underlined the gap between theory and practice in teacher education and its subsequent lack of effectiveness on teacher learning and practice (Korthagen, 2010), in particular when it comes to the application of assessment competences (Brevik et al., 2016). Moreover, CPD opportunities can be considered as 'disconnected and irrelevant to the real problems of classroom practice', including assessment (Lieberman and Pointer Mace, 2008).

The continuum of teacher education (spanning ITE, induction and CPD) has a crucial role in preparing teachers to apply and use complex contemporary assessment practices to enhance teachers' assess-

[14] While 15.7 % of surveyed teachers reported the same need in 2008 (OECD, 2013).
[15] The literature review undertaken for this study has shown that there is a general lack of research evidence from Europe on the way that future and practising teachers are prepared to assess students' key competences. The majority of the articles found came from non-European Anglo-Saxon countries such as the US and Canada.
[16] Future teachers refer to students enrolled into initial teacher education programmes.
[17] De Luca et al. (2016) define assessment literacy as 'the ability to construct reliable assessments and then administer and score these assessments to facilitate valid instructional decisions anchored to state or provincial educational standards' (p. 251-252).

ment competences and promote student learning (De Luca and Klinger, 2010; Gotch and French, 2014; Livingston and Hutchinson, 2016; Smith, 2011). Formal training and informal professional development throughout teacher education are crucial for developing a shared understanding among teachers, in particular on how to assess learning outcomes relating to key competences (Pepper, 2013). Integrating assessment practices into ITE courses as well as into practice in schools can make ITE programmes more coherent and better connect theory and practice (Brevik et al., 2016).

Recent empirical evidence shows that specific courses on assessment in ITE can be effective in enhancing future teachers' confidence in their own assessment capacity and understanding (De Luca and Klinger, 2010; Schneider and Bodensohn, 2017). Induction programmes and CPD courses supporting teachers in integrating complex and innovative assessment practices can also add value to the extent to which teachers use diverse classroom assessment practices, as noted by Koloi-Keaikitse (2016) in the context of Botswana. In their study of assessment literacy development in one ITE programme in Canada, De Luca and Klinger (2010) found that direct assessment instruction is beneficial to student teachers' perceived readiness to assess students in the classroom (including their capacity to report student achievement, modify assessments, develop constructed-response items, distinguish between assessment and evaluation, accommodate assessment for second-language speakers, or use alternative assessment methods).

Despite evidence of improved student learning and achievements through the use of formative assessment (Black and Wiliam, 1998a; Wiliam, 2007; Wiliam et al., 2004), AfL is only slowly implemented in schools, notably due to a lack of sufficient teacher competences (Smith, 2011). Nevertheless, De Luca et al. (2016) note that, since 2000, assessment literacy standards are increasingly integrating the concept of AfL and assessment education. The authors support the development of 'responsive and targeted' teacher education, addressing teachers' learning experiences and preferences for assessment education while using data on their strengths and weaknesses in assessment to 'curtail the persistent low assessment literacy rates that pervade amongst teachers' (Ibid., p. 268). Brevik et al. (2016) underline that future teachers should be given opportunities to provide feedback formatively to effectively practice AfL: 'If student teachers are expected to provide quality feedback at an early stage of students' work and during classroom activities, it is crucial that they encounter such opportunities in their teacher education as well' (Ibid., p. 17). Benett (2011) stresses that assessment components should be 'internally coherent' in aligning formative and summative assessments, as well as a form of 'external coherence'. This demonstrates the need for ITE providers to 'give teachers the fundamental skills they need to support and use assessment effectively' (Ibid., p. 19).

De Luca and Yan Lam (2014) also note that only few ITE programmes in the US manage to prepare student teachers to integrate the link between assessment and student diversity, and hence that the majority of student teachers 'articulate relatively superficial connections between assessment and diversity' (Ibid., p. 18). Furthermore, assessment often interlinks with the culturally-embedded understandings of learning across stakeholders (Raveaud, 2004). In a small-scale study looking at the use of formative assessment in two schools in Scotland, Crossouard (2011) found disjuncture between assessing the complexities of specific tasks and the 'assessment vocabularies' used by teachers.

Teacher educators have a key role in ensuring that student teachers, as well as practising teachers, develop the necessary assessment competences. However, as for many student teachers and practising teachers, teacher educators also tend to lack the required assessment literacy (De Luca and Klinger, 2010). Many teachers may find it challenging to adapt to the necessary changes posed by the implementation of contemporary assessment practices, due to limited training and experience. This challenge could be tackled by increasing collaboration within and beyond schools to foster innovative pedagogies, as well as by offering training and competence development programmes to teacher educators (Looney and Michel, 2014).

The increasing use of school curricula based on key competences underlines the need to ensure the mirroring development of teacher competences and learning outcomes for teachers in teacher education for the assessment of key competences. However, in Europe, the Key Competences Framework is still used with an uneven level of consistency across Member States and across education and training sectors (European Commission/EACEA/Eurydice, 2012). This shows the need to provide more support to teachers, as well as to teacher educators and school leaders in order to make the best out of the key competence approach. Pepper (2013, p. 24) provides examples of teacher assessment competences formulated in terms of learning outcomes, including:

- formulating learning outcomes for students;
- using a range of techniques for formative and summative assessment;
- facilitating peer and self-assessment;
- using assessment information effectively and responsibly; and
- attitudes that support theses assessment practices.

In their study on student teachers' perceptions of the importance of assessment in teacher education in Germany, Schneider and Bodensohn (2017) found that student teachers were serious about the assessment competence standards that they are expected to integrate, and willing to comply with them in their teaching practice. Interestingly, the authors show that student teachers demonstrated scepticism regarding the use of large-scale assessments, while they tend to feel that 'their key tasks are in-classroom assessment activities' (Ibid., p. 141).

Box 2. Key competences in initial teacher education in Europe.

In **Austria**, the introduction of standardised assessment of 'competence standards' for primary and lower secondary education (see also section 4.1 below), led to the adaptation of supporting structures such as CPD, and a special budget was allocated to universities providing ITE to prepare teachers to use the standards (European Commission, 2012b).

In **Germany**, the standards from the Standing Conference of the Ministers of Education and Cultural Affairs of the Lander (*Kultusministerkonferenz*, KMK) for teacher education do not explicitly refer to the assessment of key competences. However, the KMK standards relative to teacher assessment competences require that teacher graduates 'know the principles of formative assessment (…) and are capable of harmonising (pupils') learning prerequisites and (the curriculum's) desired learning outcomes', that they 'know different approaches in educational assessment (…) and apply them adequately in constructing tasks', and that they 'give feedback and substantiate evaluation and assessment outcomes in addressee's language while providing perspectives for future learning' (Schneider and Bodensohn, 2017, p. 131).

In **Spain**, the curriculum of primary education students and their future teachers is broken down by competences. Accordingly, the qualification requirements for primary education teachers to verify official ITE degrees state that 'future teachers must be able to promote the acquisition of key competences by primary students' (De-Juanas Oliva et al., 2016, p. 124). In a study on primary education teachers' assessments of the importance of teacher competences necessary for developing pupil's key competences in Spain, De-Juanas Oliva et al. (2016) found that the teaching competence related to assessment is considered necessary by a sample of 286 Spanish teachers to develop all key competences that should be assessed, in particular those related to curricular areas.

Source: compiled by the authors.
Note: These examples were chosen by the authors as illustrations.

The practical implementation of a competence-based approach to classroom assessment requires behavioural changes. Competence-based learning is likely to be resisted as long as teacher educators, teachers, parents, and students only emphasise the importance of subject-specific knowledge without giving sufficient attention to 'softer' and transversal skills (Halász and Michel, 2011).

3.3. Collaboration and teacher learning communities

Caena (2014) emphasises that a current feature of educational policy consists of designing and imposing sets of teaching standards and lists of competences, underlining the standardisation of the teaching profession. However, frameworks and standards tend to lead to a technical approach of teaching that overlooks the contextual factors and personal influence in teacher knowledge (European Commission, 2011). Korthagen (2010) claims that teacher education policy should not overlook the 'bottom-up, idiosyncratic, nature of professional learning' and focus on the personal needs and individual development of future teachers (Ibid., p. 417). Collaborative learning environments can be effective tools for all stages of the teacher education continuum (European Commission, 2015). Professional networks and collaboration platforms can provide a useful mechanism for educational stakeholders such as teacher educators, teachers, researchers and policy-makers to reflect upon and enhance the effectiveness of their assessment practices.

Professional networks act as teacher learning communities (TLCs), providing targeted and sustained professional development that gradually but fundamentally changes assessment practices (Grayson, 2014; Pepper, 2013). The increasing implementation of formative assessment methods can be supported by the creation of TLCs (Benett, 2011). TLCs can support teachers to take up formative practices of assessment and have beneficial impacts on teachers, their students and their schools (Harrison, 2005; Wiliam, 2006). Lieberman and Pointer Mace (2008) note that recent evidence demonstrates the effectiveness of TLCs on teacher learning in addition to student learning.

Wiliam (2006; 2007) and Wiliam and Leahy (2014) identify five key principles to establishing and sustaining TLCs to embed formative assessment practices in schools and support student learning:

- Gradualism ('small steps'): as teachers tend to be slow and to resist change to their classroom practice, changes in assessment practices aimed at introducing formative methods should be gradually integrated into teachers' existing routines.
- Flexibility: teachers should be allowed to 'make adjustments' to recommended assessment techniques, in order to take differences from school to school and class to class into account in the local context.
- Choice: teachers should be allowed to choose the areas of their practice which will be the most productive and beneficial for them to prioritise and develop, within the framework of suggested strategies of formative assessment professional development.
- Accountability: despite the freedom granted to teachers to choose to change or adapt their assessment techniques, teachers should remain accountable to the TLC for the changes implemented, and should be able to justify that these changes can improve student learning.
- Support: structures that provide support to teachers should be created, while keeping them accountable for developing their practice of formative assessment: the effectiveness of the learning community can be described as 'supportive accountability' (Wiliam, 2007, p. 199).

Teachers can be supported to change the way they approach assessment by feeling that they are part of a professional community where they discuss concrete issues related to assessment practice and where they can explore possible practical solutions to implement in their classrooms (Livingston and Hutchinson, 2016). The use of portfolios of students' work can, for example, be an effective formative tool to support student learning, and to share teachers' experiences of assessing key competences as a documentary basis for comparison in the context of TLCs (Mottier Lopez, 2006; Pepper, 2013; Smith, 2011).

Learning communities can help to 'encourage reflection, interchange, and support for improving classroom practice in a way that is flexible enough to account for differences among teachers' (Benett, 2011, pp. 9-10). School leaders and staff, school networks, municipalities, regional as well as national authori-

ties can all be influential to support learning communities. School leaders have a role in supporting collaborative practices and promote assessment-related professional learning for all staff (Livingston and Hutchinson, 2016; Smith, 2011). Based on a discussion on the implementation of AfL in Norway, Smith (2011) emphasises that significant resources need to be allocated to CPD purposes to implement AfL so that it reaches the whole teaching staff, and potentially makes an impact on students' learning.

Local, regional or national authorities are responsible for allocating resources, space and time devoted for CPD processes and learning communities to share knowledge with peers on best practices to expect sustainable change (Lieberman and Pointer Mace, 2008; Smith, 2011). Similar results have been found in Sweden, where TLCs have been used as a vehicle for teachers' CPD and for implementing AfL practices (Jonsson et al., 2015). The implementation of AfL on a large scale depends on key conditions for the establishment of TLCs such as the availability of a supportive structure of meetings, networks and opportunities for sharing ideas. In the Swedish case, Jonsson et al. (2015) note that the implementation of AfL through TLCs has led to an increase in pedagogical discussions at the schools, an increase in transparency regarding expectations from students, an increase in AfL practices in the classrooms, as well as an increase in workload for teachers.

Improving students' learning through improved assessment practices has been a high priority in Norway for a few years (Smith, 2011). Hopfenbeck et al. (2015) observed that local TLCs helped the implementation of AfL in the country within and across schools. Bottom-up implementation processes based on trust, dialogue and high level of teacher agency can be key factors to support the successful large-scale implementation of innovative assessment practices in schools (Hopfenbeck et al., 2015; Smith, 2011). According to a recent European Commission's *Guide on policies to improve Initial Teacher Education*, 'the best collaborative learning environments are those that are tailored to the local context and are accepted and monitored by local partners'. (European Commission, 2015, p. 56). However, the creation of professional learning communities cannot be considered as the only answer to the objective of improving teaching and learning, but only constitute the process through which change could happen: 'research shows that it is what teachers do in the classroom that really matters – not having teachers meet in workshops to talk about how to assess student work or what students' scores on tests mean for the curriculum' (Wiliam, 2007, p. 200).

3.4. The use of assessment results by teachers

This section is concerned with how assessment results are used for both summative and formative purposes in different contexts, to store information, provide feedback to students, and make decisions about their educational trajectory. The effective use of assessment results depends on teachers' assessment literacy and ability to appropriately integrate assessment data and learning in schooling (OECD, 2013). School education can benefit from being based on valid, reliable data and effective use of assessment results (Stiggins, 2005).

Summative assessment results can be used to take decisions regarding students' educational trajectories, within and outside schools, mainly as an instrument of selection to the next level of the education system (Klapp, 2015; OECD, 2013). Internally, summative results are used for keeping records and giving reports on progress to other teachers, parents and pupils (Assessment Reform Group, 2006).

For decisions beyond school, in OECD countries summative assessments are typically used for certification purposes after successful completion of compulsory education (OECD, 2013). This certificate is generally used as a minimum requirement for admission to higher education, or as a selection criterion by employers. They can also be used for selection and meeting statutory requirements (Assessment Reform Group, 2006).

Assessment results can also be used for formative purposes, in different cycles (Wiliam, 2006):

- Long-cycle (across marking periods, semesters or years): from four weeks to one year or more;
- Medium-cycle (within and between teaching units): from one to four weeks; and
- Short-cycle (within and between lessons): from 5 seconds to two days.

Most OECD countries use and promote formative assessment in school education, although little information is available about how teachers apply and document formative assessment results in their daily practice (OECD, 2013). Formative assessment information can be used in different ways and at different times. Long- and medium-term formative methods such as regular formative pupil reports, Individual Development Plans (IDPs) or Individual Student Plans (ISPs), are used to document and monitor student learning progress, notably towards subject-specific and transversal competences (Ibid.). Short-cycle formative assessment methods such as daily interactions between and among students and teachers are reported to bring 'the most direct and measurable impact on student achievement' (Ibid., p. 212). Regular development talks between school leaders/teachers and students and their guardians can help to identify areas for further improvement and develop broad teaching strategies to address specific needs (Ibid.).

Norway, while having an increasing tendency to use data from national tests, has become the only country in the world with legislation guaranteeing students' right to have AfL in their schools (Hopfenbeck et al., 2015). The Norwegian Directorate for Education and Training (DET) has developed a programme aimed to change assessment practices based on AfL, so that students and apprentices:

- understand what to learn and what is expected of them,
- obtain feedback that provides information on the quality of their work or performance,
- are given advice on how to improve, and
- are involved in their own learning process and in self-assessment (Ibid.).

Key messages

- **Defining key competences in terms of concrete learning outcomes** can provide a good basis to adapt learning and develop consistent assessment practices, striking a necessary balance between the level of detail of learning outcomes defined in policy documents and the scope for interpretation left to teachers and learners in practice.
- Research underlines the role of teacher education (ITE, induction and CPD) to provide teachers with a common understanding of key competences and relevant assessment practices to assess students' learning. **Integrating assessment practices at the ITE level and during practice in schools** can make teacher education programmes more coherent, better connect theory and practice and increase teachers' assessment literacy rates.
- **Collaborative learning environments can be effective tools to support teachers at all stages of the teacher education continuum in enhancing their assessment literacy**. By helping to encourage reflection, exchange and support for improving classroom practice, teacher learning communities (TLCs) can support the implementation of formative assessment methods, and have beneficial impacts on teachers, students and schools.
- School leaders and staff, school networks, municipalities, regional as well as national authorities, can all be influential to support and monitor TLCs to reflect upon and enhance the effectiveness of assessment practices.
- **Assessment results can be used for both summative and formative purposes** in different contexts, to store information, provide feedback to students and make decisions about their educational trajectory, as well as improve the education system as a whole.

4. MEANINGFUL CLASSROOM-LEVEL ASSESSMENT PRACTICES: EXAMPLES

The following sections look in more detail at selected classroom assessment practices aimed at assessing students' progress and learning needs, taking into account a variety of learning processes and outcomes. The authors overview practices that take a holistic approach, put students at the centre, manage local needs, and focus on improving classroom practices and teaching quality. The authors explore summative and formative assessment practices and cite available evidence on the effectiveness of specific assessment methods in measuring and developing different types of competences.

Overall, the literature suggests that new tools and criteria to help teachers assess competences – for both summative and formative purposes—are needed. Policies to support formative assessment reinforce key competence approaches (see e.g. Grayson, 2014). Classroom-based summative assessments help evaluate the performance of a student at a certain time and his or her learning outcomes in comparison with other peers, whereas formative assessments emphasise in-depth questioning and extended dialogues, self and peer-assessment, as well as feedback and guidance on improvement. As mentioned above, there is an argument for the necessity of a balanced, coherent and continuous approach that would integrate and capitalise on the different functions of assessment (Crossouard, 2011; Looney, 2011).

Nevertheless, the evidence suggests that standardised tests, teacher assessment and/or portfolio assessment (see sections 4.1 and 4.2) are the most common instruments used for assessing key competences (Gordon et al., 2009). Standardised tests are mainly used for the assessment of competences related to core subjects (particularly mother tongue, mathematics and science), whereas teacher assessment and portfolio assessment are more commonly used to assess cross-curricular competences. Some countries also started to experiment with the use of self- and peer-assessment for assessing transversal competences (Voogt et al., 2012).

Table 3 presents an overview of how particular assessment practices serve summative and/or formative purposes and can be suitable for measuring different types of competences and skills. The research review shows that standardised assessment methods, such as standardised tests and multiple-choice assessments, are well suited for summative purposes (Black, 1998; Pepper, 2013). In some cases, standardised assessments can support formative purposes as well (Pepper, 2013). For instance, surveys of competences or attitudinal questionnaires can help teachers adjust their teaching approach (Pepper, 2013). Assessment formats, such as performance-based assessment, teacher, peer- and self-assessments, serve formative purposes as well, especially by boosting reflective activities and empowering pupils to assess their own performance (European Commission, 2012; OECD, 2013; Pellegrino and Hilton, 2012; Pepper, 2013). Furthermore, Table 3 presents a summary of ICT-based practices that have the potential to develop certain key competences and transversal skills. The research review showed that ICT tools can serve well for 'optimising' classroom assessment (Redecker, 2013; Shute and Rahimi, 2017). A detailed overview of ICT-based assessment tools is presented in section 4.4.

Overall, the research review indicates that it is more challenging to assess non-cognitive and meta-cognitive competences (such as learning to learn or cultural awareness) and transversal skills, which are often taught across subjects. Nevertheless, the following sections provide evidence on practices that have already taken a step in this direction.

Table 3. The potential of different assessment practices to support key competences and transversal skills

	Standardised tests	Multiple-choice assessments	Surveys of competences, attitudinal questionnaires	Performance-based assessments	Teacher, peer and self-assessments	Computer-based assessment	Computer adaptive tests	ePortfolios	Game-based assessment	Simulations	Intelligent Tutors	Learning Analytics
Key competences												
Communication in the mother tongue	S	S	S	SF		S	S	SF	F		SF	SF
Communication in foreign languages	S	S	S	SF		S	S	SF			SF	SF
Mathematical competence and basic competences in science and technology	S	S	S	SF		S	S		SF	SF	SF	SF
Digital competence	S	S	S	SF		S	S	F	SF	SF		SF
Learning to learn			SF	F	F		F	F				SF
Social and civic competence	S			SF	F	S	F		SF			
Cultural awareness and expression				F				SF	SF			
Sense of initiative and entrepreneurship				F	F				SF			
Transversal skills												
Problem solving	S	S		F	F	S	S	F	F	F	SF	SF
Risk assessment					F							
Initiative			S	F	F			F				
Decision-taking				F	F			F	F			
Constructive management of feelings					F							
Critical thinking				F	F	S	S	F		F	F	
Creativity	S	S	S	F	F				F			

Source: compiled by the authors based on the literature review. The overview of the potential ICT-based tools to support key competences is adapted from the Redecker (2013) report.

Note: F – Formative assessment; S – Summative Assessment.

4.1. Standardised assessments

Standardised assessments can be designed and marked inside or outside individual schools (at national or system level), in order to ensure the consistency and comparability of questions, conditions for administering, scoring procedures, and interpretations (OECD, 2013). This type of assessment mainly aims at measuring students' learning outcomes in terms of overall curriculum goals and standards (Black, 1998). In other words, standardised assessments are primarily designed for summative purposes, in line with a data-based decision making (DBDM) approach in education (Van der Kleij et al., 2015).

Although they represent a predominant method of classroom assessment, standardised assessments face challenges in evaluating a full range of key competences and their complexity. Standardised assessments are often criticised for their inability to assess students in real-life contexts, because they require carefully controlled conditions (European Commission, 2012; Pepper, 2013). Nevertheless, this review shows successful examples of standardised assessments that measure a variety of skills from both EU and non-EU countries that can also be used for formative purposes. We further discuss advantages and disadvantages of the particular standardised assessment practices: standardised tests, multiple-choice assessments, and attitudinal questionnaires.

4.1.1.Standardised tests

Standardised tests are used for evaluating and monitoring purposes at individual (student and teacher), school and system level. The practice of standardised testing for monitoring the progress of individuals and groups has long influenced school education policy in several countries across Europe (Silfver et al., 2016) and elsewhere (Moss et al., 2010; Van der Kleij et al., 2015). Standardised tests can indeed be helpful to summarise students' learning at the end of a learning activity for both summative and formative purposes, or to compare learning outcomes across a population of students.

In Europe, standardised national tests 'focus on the basic skills, especially the teaching of the mother tongue (or the language of instruction) and mathematics, and to a much lesser extent science and foreign languages' (European Commission/EACEA/Eurydice, 2012, p. 10). Discussing the relevance of performance-based assessment approaches for gifted children, VanTassel-Baska (2013) notes that standardised tests are able to examine students' content mastery in key areas, such as in grammar and language mechanics. Moreover, evidence shows that in order to complete standardised tests, students need to employ a broad range of skills. On the other hand, tests can also help to address 'transversal' key competences such as digital competence, social and civic competences, as well as transversal skills such as creativity and problem-solving (Pepper, 2013). In their review of the literature on assessment instruments to measure primary and secondary students' ICT literacy (or digital competence), Siddiq et al. (2016) underline that this key competence can be assessed through a broad range of instruments, including standardised tests, multiple-choice and short answer tests, as well as performance-based and self-assessments.

Standardised tests have been criticised for not taking into sufficient consideration such things as classroom practice, student background, needs and circumstances (Shepherd and Hannafin, 2013; Van der Kleij et al., 2015). Moreover, other authors note that the excessive reliance on high-stakes standardised testing exclusively can narrow the curriculum, classroom content and teachers' pedagogies (Putwain et al., 2012; Silfver et al., 2016; Turner, 2014). Nevertheless, others stress that although standardised tests can create a sense of urgency about effectively preparing students for tests and therefore place them

under pressure, standardised testing can also lead to positive outcomes and attitudes for students including excitement, resilience and mastery (Putwain et al., 2012).

In order to assess a wide range of key competences and adequately reflect the educational context, Pepper (2013) suggests that standardised tests include the following items:

- structure and content that reproduce real-life contexts authentically;
- multiple steps requiring a chain of reasoning and a range of competences; and
- a range of formats allowing responses that require different competences.

Importantly, Bennett (2015) emphasises the need for the design, format and content of tests to better reflect the key competences they aim to assess in order to have a positive impact on teaching and learning: 'For a test to effectively exemplify key competencies, those competencies have to have been codified in a form that can serve as the basis for assessment design', so that the design of the assessment 'can reflect the key competencies and their structure, calling the attention of teachers and learners to those attributes.' (Bennett, 2015, p. 379). According to Wilson et al. (2015), the evolution of the understanding of 21st century skills such as ICT literacy to be a 'developmental progression in student skills and thinking' has prompted changes in assessment, notably toward complex technology-enhanced assessments (Ibid., p. 79).

A Eurydice report stresses that an increasing number of countries in Europe employ standardised tests to better assess key competences, such as civic and social competences (European Commission/EACEA/Eurydice, 2012) (see examples in Box 3 below). For instance, to assess different aspects of learners' social and civic competences, standardised tests in Spain and Poland include multiple response formats. In Germany, tests for measuring pupils' competences in biology, chemistry or physics include interpretive tasks, which aim to assess communication skills as well. The strength of such formats is that they better correspond to real-life situations, when the learners employ a variety of skills to complete a task. However, the evidence suggests that there is a lack of robust instruments implemented at national or system level to assess other key competences such as learning to learn, sense of initiative and entrepreneurship, cultural awareness and expression (European Commission/EACEA/Eurydice, 2012; KeyCoNet, 2014).

Box 3. Examples of standardised tests assessing a variety of key competences[18]

In **Germany**, in biology, chemistry or physics, tests include interpretative tasks, which help assess communication competence apart from subject skills. This corresponds to real-life situations, when individuals need to use a broad variety of competences at the same time.

In **Poland**, the requirements for social and civic competences are defined in the core curriculum and provide the basis for external examinations. The test includes several contrasting items, such as: listing three political rights of the EU citizen, recognising which sentences reflect opinion, and writing an essay to express their views on social and political involvement.

In Spain, standardised tests assess social and civic competences in primary and lower secondary education. They are focused on: the individual, society, social organisation, societies in the past, current democratic societies, and human rights. To assess different competences, the test encompasses various formats such as: true-false, matching, sentence completion, and short response items. Many of the test items include multiple choices and reflect an emphasis on reliability. Some of the tasks also require more reflection, description and explanation.

In **Norway**, students at different grades in primary and secondary education participate in the Pupil Survey, conducted by Norwegian Directorate for Education and Training that includes assessment of students' social and

[18] These constitute examples chosen by the research team for their illustrative quality.

emotional well-being at school. The results from this survey may be used to analyse and improve the learning environment at schools. The questionnaire for students includes items such as 'So you enjoy school?', 'Do you have any classmates to be with during recess?', 'Are you interested in learning at school?', etc.

Source: compiled by the authors based on European Commission (2012b); KeyCoNet (2014); Pepper (2011); Norwegian Directorate for Education and Training.

Although these examples focus on the assessment of key competences by national standardised assessments, they can also have an impact on classroom-level assessment. In Sweden, all pupils in grade three take the national test in mathematics, which consists of open-ended questions where pupils answer by wording, symbols or pictures (Silfver et al., 2016). However, students only receive their grades for the first time in the sixth grade (Ibid.). Olovson (2015) showed that changes in in the country's assessment practices led to more frequent and detailed communication of learning goals and knowledge requirements, greater emphasis on students' understanding, more focus on discipline in schoolwork, and increased documentation of students' subject knowledge. The author noted that teachers experienced challenges in assessing complex competences. Although the change in grading aimed at increasing students' performance, further research is still needed to evaluate the effectiveness of the reforms on students' learning outcomes.

Although attempts to assess a wider range of key competences do exist across Europe, our review of the literature shows that still in many countries standardised tests tend to be limited to 'traditional' key competences such as communication in the mother tongue, foreign languages, mathematics, science and technology, and have not yet been used broadly to assess other key competences such as learning to learn, cultural awareness and expression, and sense of initiative and entrepreneurship.

4.1.2. Multiple-choice assessments

Multiple-choice assessment is one of the standardised forms of assessment commonly used in classrooms and in the context of national standardised tests, where students have the possibility to select correct answers from predefined lists. Multiple-choice assessments may be both paper-based and computer-assisted. The literature suggests several advantages and disadvantages of multiple-choice assessment. Computer-assisted assessments especially aid to speed-up and reduce the costs of the assessment. Since assessments are machine-scored, they provide reliable data on student performance and are less expensive to administer (Looney, 2011b; Nicol, 2007).

Multiple-choice assessments can be used for the assessment of key competences, such as digital competence. In their review of available assessment instruments of ICT literacy, Siddiq et al. (2016) show that the majority of the instruments they analysed were constructed as multiple-choice tests, or combined multiple-choice and dynamic formats which require students to interact with the tasks. The authors found that several of the five competence areas that constitute overall 'digital competence' (according to the methodology of the DIGCOMP framework on digital competence in Europe; see: Ferrari, 2013), including *Information*, *Communication*, and *Content-creation*, and to a lesser extent *Safety*, *Problem solving* and *Technical skills,* were assessed by these tests (Siddiq et al., 2016). Multiple-choice assessments can also serve to assess other key competences. In the Netherlands, many schools enter their students for the tests occurring at the end of basic education which assess their attainment in language and mathematics, as well as in 'study skills and world orientation' through multiple-choice questions (Pepper, 2011).

On the other hand, multiple-choice assessments have been criticised by some authors for their tendency to promote a recall of factual information, and for not encouraging high-level cognitive processes which could better relate to key competences. The OECD (2013, pp. 175-176) stresses that 'well-

designed multiple-choice items can assess higher-order knowledge, but they cannot access broader skills and competencies such as the ability to develop an argument or communicate effectively'. However, Barlow and Marlot (2012) argue that multiple-choice tests can effectively provide more than a summative check of knowledge and skills. By analysing examples of assessment of mathematics knowledge, the authors demonstrated that multiple-choice items can be used to guide instruction, communicate mathematical expectations, and develop student dispositions. The careful crafting of multiple-choice questions can push middle grades students beyond rote memorization and promote a more challenging classroom environment (Ibid). Some assessments also use multiple-choice items to measure critical thinking or collaboration skills (see example in the Box below).

Box 4. Examples of multiple-choice assessments to measure critical thinking.

Raven's Progressive Matrices[19] shows pictures of patterns with a five-sided piece missing. Test takers are then given options for the piece that best completes the picture. Unlike most traditional IQ tests, Raven's Progressive Matrices is nonverbal, which means it can be used for students of different ages and from varying language backgrounds.

The California Critical Thinking Skills Test[20] (CCTST) is another example of MCA. Although most of the questions are multiple choice, they ask students to examine reading passages, charts, pictures, paintings, and the like and then draw inferences from them. For each student, the assessment returns scale scores for analysis, evaluation, inference, deduction, induction, and overall reasoning.

Situational judgement tests, using multiple choice logic, have also been increasingly used to measure interpersonal skills. These tests present students with scenario to test their mind-set or motivation and ask them to respond to that situation (e.g., Mission Skills Assessment). These approaches are more effective in placing students in a real-life environment, especially when the right answers are not always obvious.

Source: Soland et al. (2013); Kyllonen (2012).

In the area of ICT literacy, however, Siddiq et al. (2016) note that standardised multiple-choice assessments have been criticised 'for presenting students with structured choices which hardly reflect performance in real life' (Ibid., 2016, p. 62). Multiple-choice assessment can preclude the possibility of measuring interpretative skills, such as the capacity to develop and structure arguments. Siddiq et al. (2016) underline that some key aspects of digital competences, such as problem solving or safety and technical skills in ICT, are not equally covered by most tests of students' ICT literacy, including a majority of multiple-choice assessments. Soland et al. (2013) similarly argue that there are competences that cannot be assessed simply using multiple choice, such as complex problem solving skills or the ability to compose an essay. Other types of assessments can help address such limitations.

4.1.3.Attitudinal questionnaires

Research shows that students' attitudes towards learning can help explain their learning outcomes (Alexander and Winne, 2006; Stobart, 2008). Surveys are one type of method to measure students' attitudes. Major international surveys such as PISA, TIMSS and PIRLS integrate questions on students' attitudes to learning in their questionnaires. In contrast to the learning to learn competence defined in the European key competence framework that identifies learning outcomes in their own right, these surveys instead seek to explain students' individual performance (European Commission, 2012a; Pepper,

[19] See: http://www.raventest.net/raven-test.html
[20] See: http://www.insightassessment.com/About-Us/California-Critical-Thinking-Skills-Test-Family

2013). The assessment of attitudes as learning outcomes remains problematic for several reasons (European Commission, 2012b, p. 25):

- attitudes vary over time and between real-life contexts, but questionnaires provide only snapshots in 'test' contexts; and
- these snapshots are indirect measures, relying on the accuracy of learners' self-reports; and
- these self-reports may be influenced by the 'social desirability' of particular responses to questionnaire items, particularly if learners' perceive their answers as having consequences for themselves'.

Another important feature of attitudinal questionnaires is their specific role in assessing the accuracy of learners' self-assessments or self-efficacy. Research suggests that cross-subject key competences such as learning to learn can be assessed by combining several methods. Particularly, combining learners' self-assessment judgements with standardised tests can provide important information on students' competence to learn (European Commission, 2012a).

Pepper (2013) suggests that attitudinal questionnaires tend not to capture various social contexts and complex emotional experiences, and therefore lack information related to learning outcomes associated with specific contexts. His review of the literature on self-regulated learning suggests that classroom observation and dialogue, as well as methods based on self-assessment of students' learning outcomes, are more relevant than questionnaires for the assessment of attitudes. OECD (2015b) also suggest that assessments of social and emotional skills are often based on teachers' observations and judgment of students' everyday behaviour in different contexts. On the other hand, a combination of various assessment methods such as self-assessment via questionnaires, and assessments via tests for formative or summative purposes, might be better able to assess social and emotional competences such as learning to learn (Ibid.). In their review of available school-wide assessment methods for social/emotional well-being of students, Haggerty et al. (2011) recommended the Devereux Student Strengths Assessment (DESSA) as a valid and reliable teacher rating assessment tool. The DESSA is a behavioural rating scale assessing social-emotional competences from kindergarten through eighth grade, assessing students' self-awareness, social-awareness, self-management, goal-directed behaviour, relationship skills, personal responsibility, decision making, and optimistic thinking (Ibid.).

Box 5. Examples of attitudinal questionnaires at local level.

In **Belgium (Flemish Community)**, various institutions such as educational centres, academic institutes and NGOs have developed tools available for teachers to assess non-cognitive aspects of learning, such as attitudes or well-being. The most widely used tool for assessment of non-cognitive performance is the SAM-scale (Scale for Attitude Measurement – ***Schaal voor AttitudeMeting***), which measures students' attitudes, has been developed to assess, guide, stimulate and orientate students. It helps teachers determine to what extent pupils score high or low for certain attitudes, such as flexibility, diligence and responsibility.

An EU-funded project in **Lithuania** identified four elements within the learning to learn competence for learners in the 11-12 and 13-14 age groups: (1) attitude towards learning and willingness to learn; (2) setting objectives and planning activities; (3) organisation and management of activities; and (4) reflection on learning activities and outcomes, and self-assessment. The questionnaires for students and teachers are an integral part of classroom assessment. These elements are described in more detail for each of the age groups at four levels of progression towards the development of the learning to learn competence.

Source: compiled by the authors based on European Commission (2012b) and OECD (2013).

The box above presents some national/regional level examples of integration of attitudinal questionnaires in classroom assessment practices that have been implemented in the EU.

Key messages

- **Although they represent a predominant method of classroom assessment, standardised assessments face challenges in evaluating key competences** and transversal skills in a comprehensive manner, and tend to be criticised by some scholars for their incapacity **to assess students performing in real-life contexts.** Nevertheless, our review shows that examples of well-designed standardised assessments can be implemented effectively to measure some of the non-traditional competences, especially if combined with other assessment methods.
- **Multiple-choice assessments can be used for the assessment of certain key competences, such as digital competence**, **languages**, **mathematics, and science.** There are also examples of multiple-choice tests that can measure critical thinking and collaboration skills.
- **The assessment of attitudes as learning outcomes remains problematic**. Attitudinal questionnaires tend not to capture various social contexts and complex emotional experiences, and therefore tend to lack information related to learning outcomes associated with specific contexts. However, **valid and reliable teacher rating attitudinal assessment tools can help to assess diverse social-emotional competences.**
- A combination of different assessment methods, such as self-assessment, teachers' observations and assessments via tests for formative or summative purposes, could better assess transversal competences such as learning to learn.

4.2. Non-standardised, performance-based assessment

There has been increasing attention given to research on innovative forms of standardised assessments (OECD, 2013; Sturgis, 2014). Performance-based assessment is one of the examples used to assess a wider range of competences in a structured way. This type of assessment can serve as a method to measure students' performance as well as to support broader learning (OECD, 2013). Performance-based assessment includes assessment through specific tasks such as essays, oral presentations, portfolios, '3P assessments'[21], computer simulations, group work, reflective diaries, projects, role plays, and interviews (European Commission, 2012a; Hao and Johnson, 2013; OECD, 2013).

The literature suggests that performance-based assessments are valuable for both summative and formative purposes. They are seen as more effective than standardised tests in capturing more complex performances and processes (Hao and Johnson, 2013; Looney, 2011b, Pepper, 2013). Performance-based assessments allow the assessment of pupils over a period of time and use a wide range of tasks. Furthermore, the reliability of teachers' judgements can be comparable to those of standardised tests (Pepper, 2013). Because of the active nature of tools used for this approach, the learner can be placed in a problem situation and has the possibility of demonstrating his or her competences required to solve the problem (Tchibozo, 2011). The method also promotes in-depth learning and assesses competences needed for 'real-life situations' (Hao and Johnson, 2013).

Furthermore, performance-based assessments tend to be more favourable for ensuring equity in education. The method can help select students' learning trajectory and control their performance over a period of time (Sturgis, 2014). Spatial reasoning tasks (tasks involving location and movement of objects) used in performance-based assessments are also considered to be effective at identifying capabilities of children from a socioeconomically disadvantaged, migrant and/or minority background (VanTassel-Baska, 2014).

[21] '3P assessments' refer to assessments using product, portfolio and performance. These types of assessments help assessing 'higher order' thinking skills (OECD, 2013).

4.2.1.Portfolio assessments

Portfolio assessment is receiving more and more interest as an effective method for assessing learners' progress (Pepper, 2013). Portfolio assessment is a systematic and longitudinal collection of student work that shows his or her learning process, progress and performances (Frejd, 2013; Pepper, 2013). Portfolios can include students' texts (Burner, 2014), modelling tasks, diary notes, and written data of achievements (Frejd, 2013). The collected data also help teachers monitor learners' performance in real-life contexts or tasks that represent real-life situations (Pepper, 2013).

In the context of second/foreign language classrooms, Burner (2014) shows that portfolio assessment has potential benefits due to its process-oriented, authentic, integrated, interactive and learner-centred characteristics, and can lead to increased motivation and learner autonomy. Most of the portfolio assessment methods are now being digitalised. In modern classrooms, e-portfolios receive special interest since they can help develop digital competences, social competences, learning to learn, and problem-solving competences (Pepper, 2013, Redecker, 2013). E-portfolios have been used for a decade in some of the EU Member States as well (see section 4.4).

Research evidence shows that e-portfolio-based assessment[22] can promote self-regulated learning (Pepper, 2013; Yastibas and Yastibas, 2015; Lam, 2017) and that writing portfolios can encourage students to work more continuously with their texts and therefore develop their writing skills (Burner, 2014; Lam, 2017). A student-centred approach and continuous feedback are considered as key success factors for developing writing skills. Feedback-rich learning environments help pupils to be more independent and reflective in their writing (Lam, 2017). McLaren (2012) shows that the use of e-portfolios for providing feedback to pupils brings positive results in supporting learning, teaching and assessment, and that it also brings useful evidence for diagnostic and formative assessment purposes. The literature review implemented by Frejd (2013, p. 432) revealed that portfolios have a high potential to be developed as 'a valid and reliable assessment mode for [mathematical] modelling, but there are too few research studies focusing on these modes'.

Portfolios can also serve as a tool for diversifying tasks in multilingual classrooms. The table below presents some practical examples of how the European Language Portfolios are adapted in multilingual classrooms. The European Language Portfolio is an accompanying tool to the Common European Framework of Reference for Languages[23]. It is designed to communicate to learners, teachers, schools, and other stakeholders about respect for linguistic and cultural diversity; mutual understanding beyond national, institutional and social boundaries; the promotion of plurilingual and intercultural education; and the development of the autonomy of the individual citizen (Herzog-Punzenberger et al., 2017). The European Language Portfolio (ELP) consists of three tools: a language passport, a language biography, and a dossier (Ibid.). Box 6 below presents examples of recognition of language skills through the ELP.

Box 6. Recognition of language skills through the Language Portfolio approach

In the **Netherlands**, pupils use the European Language Portfolio (ELP) to report their language learning activities undertaken outside the classroom (e.g., the use of a different language than the schooling language with family or friends), as well as self-assess their competences. The ELP allows the recognition of pupils' first language skills acquired outside of formal education. This is highly appreciated by learners and it also enables teachers to better understand their classrooms' diversities.

In **Austria, Carinthia**, the trilingual language portfolio 'Kajpataj' was commissioned by the Federal Ministry of

[22] More information about e-portfolios is provided in section 4.4.
[23] See: http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf.

Education and regional education authority. Pupils can document their language progress in German, Slovene and Italian over a period of eight years. This portfolio was created based on the constitutionally guaranteed right to minority language instruction.

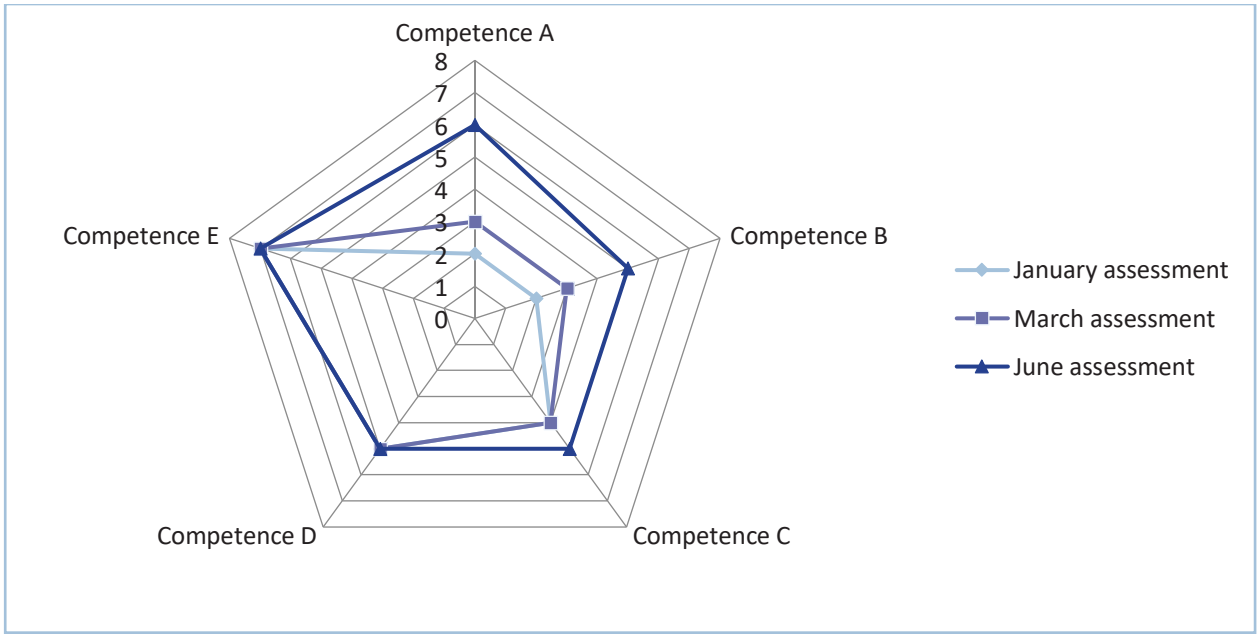
Source: Herzog-Punzenberger et al. (2017)

Although the portfolio assessment method can be beneficial for formative purposes (VanTassel-Baska, 2014), it can be more challenging to apply this method for summative purposes. Moreover, Tchibozo (2011) argues that portfolio assessments may not be effective in assessing specific competences if clear judgment criteria of the performance are not developed. Other authors note that in order to make performance-based assessment effective, teachers need to apply a variety of methods, including oral communication and writing tasks (Hao and Johnson, 2013; VanTassel-Baska, 2014).

4.2.2.Holistic scoring rubrics

Some authors (Panadero and Jonsson, 2013; Tchibozo, 2011; VanTassel-Baska, 2014) advocate for holistic scoring rubrics to assess students' competences. Holistic scoring rubrics are comprehensive descriptive scales that enable scoring competences in problem situations and allow mapping learning outcomes at the end of the curricula (Tchibozo, 2011). Holistic scoring rubrics are more commonly used for summative purposes. For instance, specific competences can be assessed on a scale from 1 to 8 (Tchibozo, 2011) or according to degrees of '*none*, *little*, *some*, *adequate*, *good*, and *excellent* demonstration of required ability, skills, or presentation' (VanTassel-Baska, 2014, p. 45). Teachers can visualise the progress of the pupils by a radar graph (see Figure 5 below).

Figure 5. Development and final configuration of a learner's competences profile using the holistic scoring rubrics



Source: Tchibozo (2011).

By employing holistic scoring rubrics, teachers can assess the competences employed for a particular problem-solving task. Each pupil's competence is assessed following an elaborated instruction on what elements the specific competence should encompass. The method is unique in the way that the evaluator can evaluate a pupil in more complex situations (Tchibozo, 2011). Schools also tend to use report cards for assessing learning skills and work habits (OECD, 2015b). For example, in Canada these skills in-

clude six categories – responsibility, organisation, independent work, collaboration, initiative and self-regulation – for each of which the teacher gives ones of the four scores: 'excellent', 'good', 'satisfactory' and 'needs improvement', which are recorded in the report cards at the end of each grade (Ibid).

Besides the summative approach of the method, it can also serve for formative purposes. Based on the results of a research review on the effectiveness of holistic rubrics scores, Panadero and Jonsson (2013) describe several benefits that the use of rubrics can bring to mediate pupils' learning, including increasing transparency, reducing anxiety, aiding the feedback process, improving student self-efficacy, and supporting student self-regulation. The authors found stronger evidence for the effectiveness of the rubrics when pupils could assess their own progress. However, at the school level it is important to provide extensive guidance how to self-assess skills, and apply the method continuously.

4.2.3.Assessment in project-based learning

To foster a variety of students' skills, teachers often employ project-based learning. Kokotsaki et al. (2016, p. 267) define project based learning as 'an active student-centred form of instruction which is characterised by students' autonomy, constructive investigations, goal-setting, collaboration, communication and reflection within real-world practices'. Their systematic review shows that project-based learning can effectively enhance skills such as critical thinking, problem solving, communication, collaboration, and self-management (Ibid.).

Project-based learning assessment refers to a broad approach which measures a performance by encompassing rubrics, reflection, self- and peer-assessment methods (Kokotsaki et al., 2016). E-portfolios can also be integrated into project-based learning assessment; however, their application requires significant time for teachers to assess students' skills and provide constructive feedback (Spector et al., 2016).

Applying assessment of project-based learning can be challenging because teachers often lack the guidance and toolkits to implement effective assessment. To help teachers, organisations in the US such as the Buck Institute for Education (BIE)[24], the National Education Association (NEA)[25], and Exemplars[26], organise workshops, webinars, and prepare toolkits for teachers. For instance, the Buck Institute for Education (BIE) stores practical webinars and blog entries for teachers on how to assess their students' projects. The organisation also provides examples of assessment maps and rubrics. Exemplars provide toolkits for teachers to assess standardised project-based learning in science, technology, engineering and mathematics (STEM). For instance, the organisation prepared a toolkit for science teachers with examples of portfolios, holistic scoring rubrics, and strategies for cooperative query[27].

At the European level, project-based teaching is not a new method, however, assessment practices for project-based learning are rather fragmented and are based on single practices. Table 4 below showcases some examples of toolkits that are used by European teachers. Some intelligent tutoring systems that support project-based learning assessment are also presented in section 4.4.

[24] Buck Institute for Education (BIE). https://www.bie.org/about/what_pbl.
[25] The National Education Association (NEA) (USA). http://www.nea.org/tools/lessons/57730.htm.
[26] Exemplars. See: http://www.exemplars.com/.
[27] Exemplars Science Teacher's Toolkit. See: http://www.exemplars.com/assets/files/toolkit.pdf.

Table 4. Examples of toolkits to support project-based learning

Tool	Description
Guidance how to use holistic scoring rubrics	A global partnership 'New Pedagogies for Deep Learning'[28] that develops deep learning practices involves 1000 schools in 10 countries, including Finland and the Netherlands. Among other toolkits, they provide guidance on how to use holistic scoring rubrics to assess six competences: collaboration, creativity, critical thinking, citizenship, character and communication.
An online course 'Introducing Project-Based Learning in your Classroom'[29]	European online platform for school education 'School Education Gateway' invites European teachers to join a course on project-based learning in the classroom. When enrolling in the course, teachers can find a wide range of resources such as classroom videos, lesson plans, interviews, presentations, digital tools, and classroom materials, for peer-review and self-assessment activities.

Source: School Education Gateway; A global partnership 'New Pedagogies for Deep Learning'.

Besides the strong potential of performance-based assessments to be applied for both summative and formative purposes, the literature shows several limitations of this method. Looney (2011b) argues that it is expensive to administer, and that it might lack reliability, especially when the scores are awarded by human assessors. This limitation can be addressed by clear guidelines and through teacher training. Involvement of external examiners could also mitigate the remaining challenges. For instance, in Germany and Austria, external examiners can review and moderate teachers' assessment judgements to increase the comparability of results between students, teachers and schools (European Commission, 2012a).

Overall, performance-based assessment is seen as an effective method to measure wide-ranging competences and higher-order skills. The main strength of the method is that it focuses on a learner-centred approach and on continuous monitoring of learners' progress. Furthermore, employing a variety of techniques, such as portfolios, holistic scoring rubrics, and feedback, enables the assessment of broader learner competences. However, in order to use this method effectively, clear definitions and a scale for competence development need to be included in the assessment practice, and the country-specific context should be taken into account (Hao and Johnson, 2013).

Key messages

- **Performance-based assessment has the high potential to measure and foster wide-ranging competences and higher-order skills**, since it encompasses different assessment techniques and integrates feedback mechanism.
- The key strengths of performance-based assessment are: **the focus on the learners' personalised needs, clear definitions of the goals, and timely feedback**. Furthermore, individualised approaches help to reduce the achievement gaps between students.
- **Portfolios (e-portfolios)** is beneficial due to its process-oriented, authentic, integrated, interactive and learner-centred characteristics, and can lead to increased motivation and learner autonomy.
- **Clearly defined rubrics and project-based learning assessment techniques are considered as effective tools**, especially if applied together with concise and timely feedback. Research

[28] A global partnership 'New Pedagogies for Deep Learning': http://npdl.global/.
[29] School Education Gateway: Introducing Project-Based Learning in your Classroom: http://academy.schooleducationgateway.eu/web/developing-project-based-learning-in-your-classroom.

demonstrates that **project-based learning can effectively enhance skills such as critical thinking, problem solving, communication, collaboration, and self-management.** Although a variety of toolkits supporting performance-based assessment exists, the **European market is quite fragmented** and a more centralised approach would be helpful for practitioners.

4.3. Peer and self-assessment

Due to their collaborative nature, peer and self-assessments have a strong potential to enable students' deep learning[30] (Black and Wiliam, 1998a; Pepper, 2013). To make learning interactive, teachers can use various methods, from written work to observation and discussion (Black and Wiliam, 1998a). A variety of assessment methods might enable teachers to better understand the capabilities of the student and in this way set individualised milestones.

Successful teacher, peer and self-assessments are useful in building many key competences, such as initiative and entrepreneurship, learning to learn and social competence, as well as transversal skills such as critical thinking, creativity, initiative, problem-solving, risk assessment, decision-taking, and constructive management of feelings (European Commission, 2012a). Social and emotional competences are also better assessed through multiple forms of assessment, including student self-evaluation, teacher evaluation, and parent evaluation where appropriate, and making use of various modes such as self-reports, checklists, and direct behaviour assessment (Weissberg et al., 2015). In particular, peer assessment has a strong potential to boost transversal skills and it enables giving feedback to larger number of students (Topping, 2009).

4.3.1.Self-assessment

Black and Wiliam (1998a) underline that self-assessment is essential for successful formative assessment. To improve the learning process and outcomes, learners need to understand the main purposes of their learning and reflect on the feedback provided by teachers, and potentially by peers. Student self-assessment is a central component of current conceptions of classroom assessment, particularly formative assessment (Brown et al., 2015), and a number of studies have demonstrated a positive association between self-assessment, learning and achievement (see Brown and Harris, 2013). Some of the self-assessment practices that can be used by students include self-estimation of performance, self-correction, self-rating, rubric based judgements, etc. (Ibid). Some researchers argue that student self-assessment should no longer be treated as an assessment, but instead as an essential competence for self-regulation (Brown and Harris, 2014).

Building a clear understanding of the learning goals enables students to self-assess their performance. According to Pepper (2013), teachers and learners should share a common understanding of learning outcomes relating to key competences. To internalise this understanding, both learners and teachers need to identify and understand three elements: (1) the expected learning outcomes; (2) their present position in relation to those outcomes; and (3) how they could close this gap. However, pupils often lack the understanding of the targets of their learning and often are passive learners (Baird, 2014). This limits the effectiveness of learning, because understanding the main purposes of the learning is seen to be crucial for formative assessment (Black and Wiliam, 1998a).

[30] Deep learning refers to collaborative learning through reflection in action and on action (Fullan and Langworthy, 2014). The goals of deep learning is to provide students with competences and dispositions necessary to become 'creative, connected, and collaborative life-long problem solvers and healthy, holistic human beings' (Ibid, p. 2).

One of the challenges for self-assessment to be implemented effectively is that many students are not self-regulated enough to do it, which in turn can lead to inaccuracy (Brown et al., 2015). Inaccurate self-assessment might also be attributed to the social environment of classrooms, where the pressure to enhance or even protect one's own self-worth can result in overestimation of one's ability, or inaccurate self-reporting of grades or test scores (Ibid). The challenges related to grading and trust or respect can be managed by implementing self-assessment in a context likely to promote accuracy – or at least not promote inaccuracy – meaning that self-assessments should not count towards grades and should be private. Furthermore, research emphasises the need for thorough training for both teachers and students before they can engage in self-assessment as a taught and learned competence (Brown and Harris, 2014). Damgaard and Nielsen (2017) suggest that using nudging policies, such as coaching and mentoring, have the potential to improve students' self-control abilities and decision-making process when it comes to their self-assessment.

Box 7. Students' self-assessment examples

Finland is seeing a powerful push for student self-assessment, so that students may understand their progress and help design their own activities. This is most pronounced in the tenth through twelfth grades (upper secondary school), when students engage in self-directed, self-paced learning. During these years, students build their own personalised learning schedules (comprised of 70-minute classes and six-week units), and they may complete the courses at a pace appropriate to their abilities and unique circumstances. Most complete the prescribed courses in three years, though some students progress more rapidly or slowly.

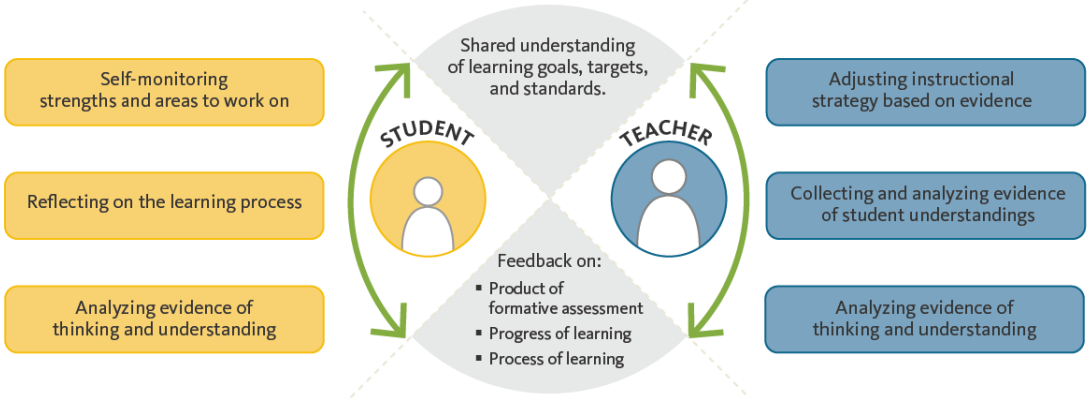
In **New Zealand**, the science teams in some schools introduced scaling sheets for display in the classroom, where students could post their names in relation to how they self-assess specific skills. As the unit proceeded, and as they produced evidence of their achievements, they could move their name up on the scale. This is an example of a good 'starter' activity, as it increases the use of self-assessment within a familiar unit of work. Empowering students to discuss their learning progress is a strength, and it could be even further enhanced were students to assist in the construction of the scale.

In **Ireland**, self-assessment is used within the curriculum subject of Social, Personal and Health Education as a means to enhance pupils' self-awareness of their social and emotional skills. At the secondary level, self-assessment is complemented with peer assessments based on fixed criteria.

Source: Bristow and Patrick (2014), Hipkins (2007), OECD (2015b).

Providing relevant feedback is crucial for enabling students to self-assess their performance and continue strengthening their skills. There is growing evidence that learners acquire better skills if they are aware of the nature of the mistakes they make, rather than only knowing the mistakes themselves (Roscoe and Chi, 2007; Shute, 2008). The empirical study on self-evaluation strategies in math conducted by Ramdass and Zimmerman (2008), demonstrated that accurate self-reflection leads to enhanced students' academic achievement in math and more importantly, improved self-regulatory skills, provided students' self-efficacy judgments are carefully monitored by teachers. Pellegrino and Hilton (2012) notice that the feedback students receive is often neither timely nor informative, and due to this, students keep practicing tasks incorrectly. To mitigate these concerns, teachers are encouraged to give explanatory feedback to their pupils. Feedback could take the form of a dialogue, which 'should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas' (Black and Wiliam, 1998a, p. 8). The authors also note that the means of class tests or exercises are important to track progress, but they should be frequent and short rather than infrequent and long. This provides an opportunity for teachers to give feedback in the middle of the learning process. Providing timely and clear feedback is highly important to help students learn. This represents an even more important element for those students that tend to fall behind their peers and lack the self-esteem to boost their skills (Black and Wiliam, 1998a).

Figure 6. Formative assessment feedback process



Source: Measured progress Toolkit (http://www.measuredprogress.org/)

4.3.2.Peer assessment

Peer assessment is another form that is receiving more and more attention in the literature. It is described as an educational arrangement for learners where students can assess a product or performance of other peers and in this way, stimulate reflection, discussion and collaboration (Strijbos and Sluijsmans, 2010; Topping, 2009). Students can assess their peers' products, such as writing, oral presentations, portfolios, test performance, or other skilled behaviours (Topping, 2009). Although peer assessment can be applied for summative purposes (Kaufman and Schum, 2010; Kearney et al., 2015), there is a lack of research evidence on this approach and its potential and limitations.

Peer assessment is considered as a powerful tool to engage students in active learning, reflection (Dziedzic et al., 2008; McMahon, 2010) and raise their motivation and academic standards (Black and Wiliam, 1998a; McMahon, 2010). Kearney et al. (2015) reveal that 'students in the early stage of their tertiary education were able to judge their own work as well as peers' work with reasonable accuracy' (p.12). Also, thanks to peer assessments, students may change their perceptions of their capabilities, improve their work, and develop skills necessary for autonomous learning (McMahon, 2010). Peer assessment strategies in the classroom can help learners develop critical thinking, meta-cognitive skills and deep-thinking (Hou et al., 2007; Sitthiworachart and Joy, 2008). These aspects are especially important for developing transversal skills and the competence to learn. Peer assessment constitutes an effective way to improve the effectiveness and quality of learning, which can be as useful as gains from teacher assessments (Topping, 2009). Both the receiver and the giver of the feedback receive mutual benefit, since they can improve their self-assessment skills and consequently their learning to learn skills. Peer-reviewing can also serve as a means for strengthening collaborative aspects. For instance, it can help build a stronger learning community (Søndergaard, 2009) and strengthen group work in the classroom (Topping, 2009). On the one hand, peer assessments are often considered as a means to save teacher's time. On the other hand, to smoothly implement this method, teachers need to invest a substantial amount of time for organising, training and monitoring (Topping, 2009). Bonn (2014) showed that peer assessment practices can be effectively applied in primary school writing. The author discovered that pupils are able to give appropriate feedback if it meets following conditions: if it is task-involving and useful; if there is sufficient time given for them to act on it and discuss it with their peers; and if they are asked to reflect on how it has been used to improve the quality of the assessed work (see Box below).

Box 8. Peer assessment in entrepreneurship education

In **Austria**, You^th Start initiative comprising several activities and projects in the area of entrepreneurship education. The You^th Start framework of reference for entrepreneurship competence[31] consists of statements of what learners can do and is used as a planning and design tool addressed principally to educators and school governance in secondary and vocational education institutions.

The You^th Start framework tackles a wide range of entrepreneurship competences including knowledge, skills and attitudes. Teaching methods associated with You^th Start are based on hands-on learning, competitive learning - applied through business idea and business plan competitions – as well as debates, buddy systems, project and group work, role plays, etc. Peer-assessment practices, in the form of coaching, mentoring and buddy systems (peer-review) are often used to assess such components as 'skills' and 'attitudes'. The case study conducted by Komarova et al. (2015) concluded that peer-assessment strategies are effective methods in promoting students' learning and collaboration skills. Peer and external reviews, including presentations and events, have a strong validation character, and are thus important for both assessment and teaching.

However, the authors also emphasise that it is important to use a range of assessment methods to fully cover each entrepreneurship competence component (Ibid).

Source: Komarkova et al. (2015).

Although the findings presented above show the strong potential of self and peer assessment, research focused on the school level reveals that the practical implementation of these methods could be challenging (Boon, 2014; Borghouts et al., 2016; Van den Berg et al., 2016). Van den Berg et al. (2016) explored formative assessment practices for primary mathematics education. Teachers integrated into their study practice three types of assessment: a short-term assessment feedback loop, an intermediate assessment feedback loop, and a long-term assessment loop. Each assessment feedback loop contained four elements: goal setting for instruction, assessment, instructional feedback, and evaluation. Although the results of the study showed that students increased their mathematical performance, there is a risk that at a larger scale many teachers will probably have difficulties with classroom management skills and the time that is required for tasks (van den Berg et al., 2016).

The literature also suggests that peer assessment might be influenced by social processes, such as social loafing (failing to participate), free rider effects (having the work of others accepted as one's own), diffusion of responsibility, and interaction disabilities (Salomon and Globerson, 1989; Tchibozo, 2011). This might result in a weakening of the reliability and validity of peer assessments: 'peer assessments can be partly determined by friendship bonds, enmity, or other power processes, the popularity of individuals, perception of criticism as socially uncomfortable, or even collusion to submit average scores, leading to lack of differentiation' (Topping, 2009). A teacher can mitigate these challenges by careful planning. Well-organised assessment practices help learners acquire important social and communication skills, such as negotiation and diplomacy, and teamwork skills (Topping, 2009). Cho and MacArhtur (2010) also noticed that peer assessments are more effective when students receive feedback from multiple peers. This especially helps improve their writing quality. Furthermore, clear guidance and training for students on the essence of peer-assessment improves students' ability to regulate and make specific decisions (Gan and Hattie, 2014).

Several EU Member States have made a step towards promoting peer- and self- assessment practices. For instance, some schools in the Czech Republic participated in a pilot project, during which they developed self-assessment booklets for each pupil with specified goals and details in progress of their goals. Pupils in Finland use individual study plans which help them reflect and monitor their learning.

[31] YouthStart Framework of Reference for Entrepreneurship competences. (2014, Version 15). Impulszentrum für Entrepreneurship Education (eesi) des bmbf and Initiative for Teaching Entrepreneurship (ifte), Vienna 2014.

Schools in Denmark, Finland, Norway, and Sweden are currently 'focused on continuous assessment based on dialogue, rather than summative tests and exams, complemented by self- and peer assessments. The ultimate objective is to nurture self-regulated learners who can set their own learning goals and work towards them independently' (Adams Becker et al., 2017, p. 6).

Key messages

- Due to their collaborative nature, **peer- and self-assessments have a strong potential in enabling students' deep learning and self-regulation** and are important elements of integrated assessment framework to be used in the classroom.
- Research suggests that **peer- and self-assessments are particularly useful in engaging students' in active learning and reflection and developing non-traditional competences**, such as initiative and entrepreneurship, learning to learn and social competence, as well as **transversal skills** such as critical thinking, creativity, problem-solving, risk assessment, decision-taking, and constructive management of feelings.
- However, for these assessments methods to be effective, **it is important to provide comprehensive training to teachers** as observers and supporters of the assessment process, **and students** as the main assessors. It needs to be recognised that being an accurate assessor is a competence in itself.
- **Interactive assessment methods are also more flexible in adjusting to diverse learning needs**, and therefore should be promoted as an integral element of classroom assessment practices.

4.4. The use of technology in classroom assessment

The use of ICT can be beneficial to generate relevant information for large-scale performance-based assessments, as well as to facilitate classroom-based formative assessment (Halász, 2016; Looney, 2011). ICT-based methods can also help improve equity in assessment by being sensitive to the needs of particular groups such as second language learners or students with special educational needs (OECD, 2013). New technologies have indeed opened opportunities to provide assessment formats which can help to comprehensively capture complex key competences (Johnson et al., 2014; Redecker, 2013).

The increasing use of ICT in teaching and learning implies new models of assessment. For instance, the OECD (2016) reports that technology significantly facilitates the use of formative assessment because it allows immediate feedback for both learners and teachers, and ensures the learner's participation. In turn, this can enhance better targeted teaching and engaged learning. Technology can help measure complex skills such as reasoning or problem solving through measures such as essays, blogs or virtual learning environments (Ramirez-Corona et al., 2013). Computer feedback can also help continuously develop writing skills (Graham, 2015). Binkley et al. (2012) describe two key strategies for the use of ICT in assessment: 1) to deliver traditional assessment more effectively and faster; and 2) a 'transformative' strategy that aims to use ICT to change the way competences are assessed and find effective solutions for assessing transversal skills, which were difficult to assess with traditional methods. In this section we will concentrate on the latter strategy.

The Joint Research Centre (JRC) Institute for Prospective Technological Studies (IPTS) has spent significant effort to assess the potential of innovation in education and the skills needed to respond to 21st century challenges. In a recent report (Brecko et al., 2014), the JRC recognises the need to mainstream ICT-enabled learning innovations and as part of it, the need to reform assessment practices. The authors propose to reform assessment strategies and examination systems in the following ways: (a) at national level, to assess both factual knowledge and key competences; (b) at the classroom level, to shift owner-

ship of assessment from teachers to learners; and (c) to promote formative assessment forms, where they are considered to be an integral part of the learning process (Brecko et al., 2014). In line with these directions, in this section we present several technology-based assessment formats that have a strong potential in assessing a wider range of skills.

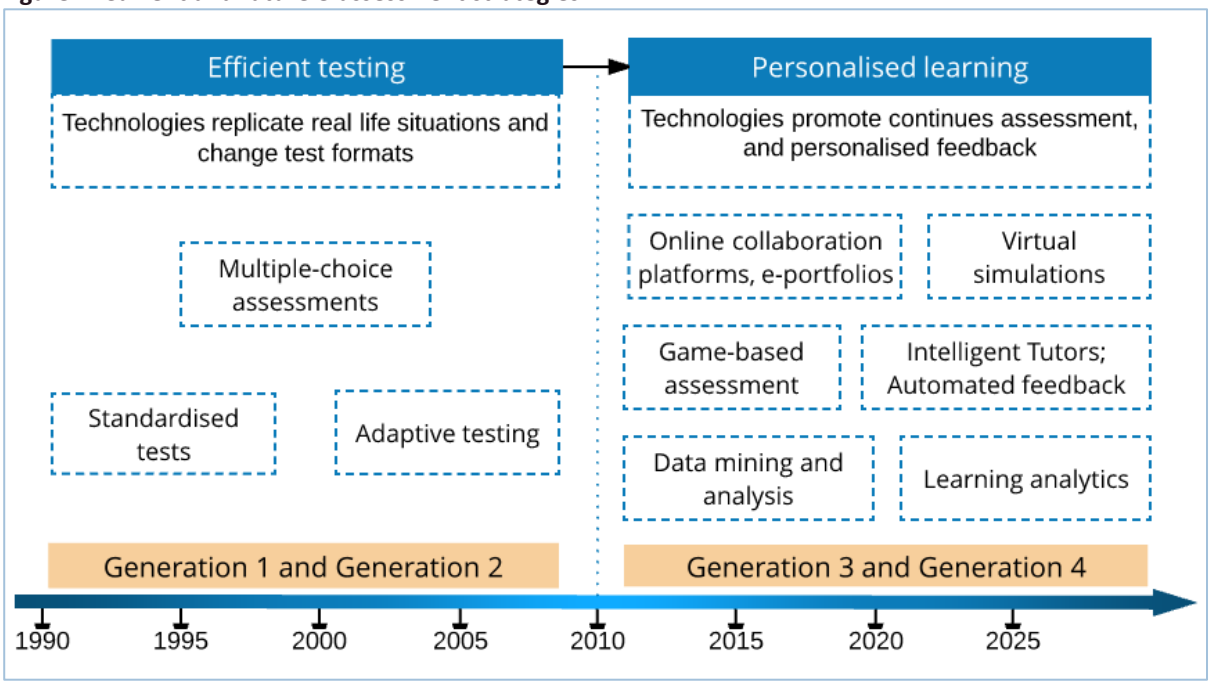
Computer-based assessment, often referred as e-assessment, offers solutions to enhance the potential of Assessment for Learning (AfL). Looney (2011b, p. 196) notes that 'computer-based performance assessments may potentially assess more complex performances through simulation, interactivity, collaboration and constructed response formats. Increasingly sophisticated ICT programmes that score 'open-ended performances' may address concerns regarding reliability of human-scored assessments, and validity of multiple-choice assessments that do not effectively measure higher-order skills.'

Bunderson et al. (1989) defined four generations of e-assessments:

- **Generation 1:** Computerised testing, which helps administer standardised tests and in this way reduces teachers' workload;
- **Generation 2:** Computerised adaptive testing, which helps adjust the difficulty of tests to learners' abilities;
- **Generation 3:** Continuous measurement, which allows test performance multiple times and in this way integrates assessment practices into pedagogic activities; and
- **Generation 4:** Intelligent measurements (automated scoring), which produces inferences. Intelligent measurements encompass several functions such as: scoring complex responses, generating interpretations based on individual profiles of profiles, and providing advice during continuous measurement to optimise learner and assessor progress.

Although this categorisation was developed as a prediction, educational researchers, including the IPTS, continue to use this four-generation assessment categorisation. Figure 7 below presents a timeline of the changing strategies of e-assessment and the concrete methods assigned to assessment generations.

Figure 7. Current and future e-assessment strategies



Source: IPTS based on Bennett, 2010; Bunderson, et al., 1989; Martin, 2008.

The first two generations have been widely integrated in classroom assessments. However, 'the main challenge now lies in making the transition to the latter two, the era of Embedded Assessment, which is based on the notion of 'Learning Analytics', i.e. the interpretation of data about students' proficiency in order to assess academic progress, predict future performance, and tailor education to individual students' (Redecker and Johannessen, 2013, p. 81). Based on the IPTS reconceptualisation, the second and third generation assessment marks a cultural shift from efficient testing (computer-based assessment) to personalised learning (embedded assessment). Embedded assessment helps to merge formative and summative assessment within the learning process because it allows for the continuous monitoring and guidance of learners via the digital environment (Redecker and Johannessen, 2013).

Assessment methods such as data mining, learning analytics and behavioural tracking encompass third-generation assessment practices. Fourth-generation assessments are associated with automated feedback and intelligent tutors. Although the third and fourth generation assessments are seen to have the potential to capture a wider range of key competences and transversal skills, these assessment practices are still under development (Bennet, 2015; Redecker and Johannessen, 2013). We further discuss recent developments in e-assessment practices and how they respond to the needs to develop key competences and transversal skills.

According to the literature review by Shute and Rahimi (2017), computer-based assessment tools for instructional support in the classrooms were used from the early 1960s to the 1990s and, when technology advanced, computer-based assessments started measuring more complex competences like problem solving skills. These basic tools for instructional support can be referred to as 'generation one e-assessments', and those supporting more complex competences as 'generation two e-assessments'.

First generation e-assessments primarily serve to optimise standardised tests and multiple-choice assessments. ICT-based multiple-choice assessments that adapt to the learners' level of knowledge are called **computer adaptive tests (CAT)**. This approach is more advanced than the first generation assessments, since it enables a teacher to better understand learners' knowledge and skills. Students who answer correctly are directed to a more difficult question, and those who answer incorrectly receive easier questions. In this way CAT is considered to be a method which provides more precise scores of student performance than typical assessments (Looney, 2011b). CAT is especially handy for assessing language skills in which the computer elects and presents test items to examinees according to the estimated level of an examinees' language ability (Dunkel, 1999).

Computer adaptive tests are applied in mathematics as well. Wu et al. (2017) showed that adaptive tests are sensitive enough to provide information about an individual's ability, change processes and mediation strategies. In their research, they controlled three groups of 5th grade students, who received: (1) the adaptive dynamic assessment with individualised instruction; (2) individualised instruction without adaptive dynamic assessment; and (3) traditional classroom remedial instruction. Those students who were assessed using adaptive dynamic assessment performed better than other pupils, and those who received the individualised instruction (integrated in the ICT test) performed even better. As do all methods, CATs have their limitations. One of the main limitations of CAT is the demand for a high number of test questions, which increases development costs (Looney, 2011b). Another limitation is the lack of comparison of results between students, since tests are individualised and adapted to each students' responses (Ibid.).

The following examples are already able to be linked to the third assessment generation, since the measurements enable testing performance multiple times and focus on personalised learning. At classroom level, this is in line with the AfL approach, which encompasses two main purposes: (a) to provide

appropriate and timely feedback for students; and (b) to provide personalised learning (Shute and Rahimi, 2017). The third generation e-assessments could include the following methods: e-portfolios, web-based peer assessment, computer-based quizzes and simple games, multimedia science experiments, virtual worlds and games, computer simulations and virtual laboratories, and data mining and analysis.

4.4.1.E-portfolios

Technologies help to effectively use portfolio assessments, which were introduced in section 4.2. E-portfolios are already well integrated into some European schools for both summative and formative assessments. In 2011, Eurydice reported that Austria, Belgium, Portugal, Romania, the UK and Turkey had already been implementing this method in school education, and that e-portfolios had been piloted in Bulgaria, Germany, France and Iceland (Eurydice, 2011a). Redecker (2013, p. 83) notes that e-portfolios are 'powerful tools for improving communication in the mother tongue, communication in foreign languages, and cultural awareness and expression competences'. They can also be used for collaborative purposes because teachers and peers can easily produce their feedback (Pepper, 2013).

E-portfolios can be effective when employing other methods as well. For instance, Chen and Chen (2009) developed a mobile formative assessment tool that helps teachers analyse e-portfolios. This tool is based on data-mining principles (Redecker, 2013). However, to fully employ such tools, teachers need additional support. Currently, teachers often lack an understanding that e-portfolios can encourage online collaboration, self- and peer-assessments, and help acquire learning to learn skills. Thus, the opportunities of e-portfolios are not fully employed (Redecker, 2013).

Apart from e-Porfolios, other innovative tools are attracting interest for peer and self-assessments. Some web-based tools for peer-reviewing have already been developed, however, they are more widely employed at university level (e.g. SparkPlus[32] in Australia or PeerWise[33] in New Zealand). SparkPlus enables students to rate their own or their peers' contributions to team tasks and improve their judgement through benchmarking (Redecker, 2013), while PeerWise supports students in creating, sharing, evaluating and discussing assessment questions.

4.4.2.Tools for online feedback

Although web-based peer-reviewing tools are usually applied at the level of higher education, recently there have been some developments at the level of secondary education as well. For instance, Wang (2011) developed a peer-driven assessment module of the web-based assessment and test analysis system (PDA-WATA). It aims to encourage self-regulated learning and improve e-learning effectiveness by proposing five strategies: (a) adding answer notes, (b) stating confidence, (c) reading peer answer notes, (d) recommending peer answer notes; and (e) querying peers' recommendation on personal answer notes. The module had a positive impact on improving self-regulatory skills and learning skills (Shute and Rahimi, 2017). Another example is the ASISSTments tool, which provides formative feedback to students, teachers, school administration and parents on an online platform. One study found that regular use of ASISSTments increased performance on summative assessments (Shute and Rahimi, 2017). Table 5 presents some practical examples of toolkits that support teacher, self- and peer-assessment practices in the classroom. These toolkits provide students real-time formative feedback on

[32] See: https://uts.sparkplus.com.au/login.php.
[33] See: https://peerwise.cs.auckland.ac.nz/.

their writing. Such e-assessment practices can be highly beneficial for students' learning outcomes, but to be effective, they must be well integrated into the pedagogy (Agelii Genlott and Grönlund, 2016).

Table 5. Examples of toolkits to support online teacher and peer assessment

Tool	Description
FOLIOS (FR)	FOLIOS is a tool for continuously tracking pupils' academic and extracurricular skills. The objective of FOLIOS is to follow pupils throughout their academic career and recognise their skills. FOLIOS supports classroom activities by allowing teachers to share documents, information, and resources with pupils and provide guidance to students. Each pupil holds an individual account (e-portfolio), where he/she can track his/her own performance. This makes pupils more involved in their learning process and encourages reflection. Additionally, FOLIOS allows exchange information with parents. FOLIOS was launched in France in September 2015. After a pilot phase, the measure is currently being implemented[34].
ASSISTments (US)	ASSISTments is a free web-based tool for US teachers to assign mathematics homework and track students' performance. The tool allows teachers to give timely feedback to their pupils and at the same time receive structured information about students' work. Teachers can either select tasks/problems from the content library or create their own content. Research on the effectiveness of ASSISTments revealed that the tool significantly increased students' scores on an end-of-year standardised mathematics assessment compared to students who used traditional homework practices. Students with low prior mathematics achievement benefited the most (Roschelle et al., 2016). Thus, ASSISTments has a high potential to decrease achievement gaps.[35]
Write to Learn (SE)	Write to Learn is a peer assessment tool used in Sweden that allows children from 1st grade to use several ICT tools to write texts and subsequently discuss and refine them together with classmates and teachers using digital real-time formative feedback and assessment. The Write to Learn method was launched in 2011 with the aim to increase communication skills in literacy and mathematics, as well as close the gender gap among pupils (Agelii Genlott and Grönlund, 2016).

Source: School Education Gateway; Agelii Genlott and Grönlund (2016).

As discussed, feedback is an essential component for assessment of learning practices. Maier et al. (2016) investigated the feedback effectiveness given through a Moodle environment in ten biology classrooms in Germany. The researchers used the close-format for multiple-tier concept tests that included both content and reasoning tasks. In their investigation, they considered three treatment groups: students who received instruction-based feedback, students who received a dichotomous verification feedback, and those who received no feedback. Verification feedback only confirms if a learner responded correctly to a question, while elaborated feedback includes explanatory information (Shute and Rahimi, 2017). Maier et al. (2016) found that the elaborated feedback was too long and detailed and thus had no significant impact on formative learning. In reaction, Shute and Rahimi (2017) suggested avoiding cognitive overload when providing feedback in manageable units. They also reviewed stud-

[34] See: http://www.schooleducationgateway.eu/en/pub/resources/toolkitsforschools/detail.cfm?n=481.
[35] More information: https://www.assistments.org/.

ies on the effectiveness of an accelerated math system 'designed to improve math achievement through individualised drill and practice and to help teachers provide appropriate feedback and monitor their students' progress' (Shute and Rahimi, 2017, p. 8). The research findings showed that longer use of such assessment practices is more beneficial than shorter use.

Although employing various ICT tools in assessment practices has high potential, they should be used critically by assessing their actual effectiveness on enhancing skills. For instance, Cayton-Hodges et al. (2015) explored how the use of tablets in mathematics classrooms can help improve math skills and some transversal skills such as problem-solving or self-reflection. Although the authors discovered a number of apps built for math assessment, most of them poorly engaged pupils in meaningful mathematical practices. Only a small number of apps provided a rich mathematical dimension, and tasks in the apps more often than not did not require complex problem-solving and reflection skills.

4.4.3.Learning analytics

Learning analytics as an assessment form has made first steps and garnered substantial interest from researchers over the recent decade (Johnson et al., 2011; Shute and Rahimi, 2017). According to Johnson et al. (2011, p. 28), 'learning analytics refers to the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues'. Although authors emphasised the method's potential to capture deeper learning skills, they also listed several challenges, such as: difficulties with collecting data from different sources; concerns about student privacy; and the challenges of acquiring advanced technical skills.

Researchers and companies that develop data-driven learning platforms need to address these challenges so that the technologies could enter European schools and effectively support students' learning (Johnson et al., 2014). More and more schools are piloting learning analytics tools, especially in northern European schools. In a recent report on technologies for 'Nordic schools'[36], Adams Becker et al. (2017) forecasted that this method could potentially become mainstream in a two to three year period. Some examples of learning analytics (data-driven learning) tools are provided in the table below.

Table 6. Examples of learning analytics tools

Tool	Description
'Itslearning' platform	Itslearning is a cloud-based learning platform used worldwide. Itslearning helps teachers and students achieve personalisation, individualisation and differentiation in teaching and learning. It tailors learning objectives, content, method, pace and environment to each student's unique learning needs. The platform allows for making learning more creative and interactive. Teachers can use a mix of face and online instruction, as well as employ or create a variety learning styles such as educational videos, learning games and applications. The tool also allows pupils to create their own videos, animations, webpages, music, etc. Eight EU Member States (Denmark, Finland, France, Germany, the Netherlands, Poland, Sweden and the United Kingdom) are currently using this platform in classroom activities[37].

[36] Nordic schools in the report encompass Denmark, Finland, Norway, and Sweden.
[37] More information: http://www.itslearning.eu/teaching-strategies.

Tool	Description
SAM learning	SAM learning is a platform that showcases integration of analytics in learning environments. The platform allows teachers to monitor students' progress and to identify and assist students who have difficulties with particular tasks or learning materials. The Fisher Family Trust's report[38] revealed that students who used SAM learning for at least ten hours improved their scores by one GCSE grade (Johnson et al., 2014, p. 20). To drive school improvement, SAM learning provides such learning analytics tools as 4Matrix, SISRA and FFT[39]. These are learning analytics tools that make it easier to review, set targets in learning, and analyse data on school or pupils' performance.
Learning Analytics Community Exchange (LACE)	Learning Analytics Community Exchange (LACE) was an EU-funded collaborative project that aimed to support learning community working on learning analytics and data mining from schools, workplace and universities to share best practices.[40] Currently, the project website stores relevant information, studies and materials on how to apply learning analytics in different levels of education. The project outcomes encompass evidence hub, evaluation framework for learning analytics, relevant studies, and useful comments from the community.
Assessment Strategy at Martinlaakson Lukio	Grapling with a large number of students in their classrooms, the Finnish upper secondary school Martinlaakson Lukio provides an autonomous math course for pupils. The formative assessment tools are embedded in a learning management system. They allow pupils to control the pace of their learning. Data-driven learning allows pupils to develop self-evaluation skills, and teachers are able to allocate more time for pupils' individual needs (Johnson et al., 2014, p. 20).
AMOEBA	AMOEBA is a learning analytics tool built to help secondary school teachers manage collaboration among their students in a programming environment called 'I can PROgram' (IPRO[41]). The tool analyses pupils' programming behaviours in the IPRO environment. After identifying the learning traits of pupils, teachers can group students according to their level. During group work in the classroom, pupils receive assignments to solve programming problems (Shute and Rahimi, 2017, p. 12).

Source: Adapted from Johnson et al., 2014; Shute and Rahimi, 2017

4.4.4.Intelligent (virtual) tutors

There is a consensus among educational researchers that feedback is essential for improving learning outcomes. In order for feedback to be effective, it needs to be timely and clear (Black and Wiliam, 1998a; Maier et al., 2016; Sansone et al., 2016; Shute and Rahimi, 2017). Effective feedback should also be rather short (Shute and Rahimi, 2017). Intelligent tutors support the function to provide automated feedback. These programmes provide information if the answer is correct, and can explain why it was incorrect (Nunan, 2010). The software that employs machine learning can detect patterns in written work, speech, and other actions, which can be individually adapted to students' learning styles and needs (Adams Becker et al., 2017).

[38] Available at: https://www.samlearning.com/about-us/.
[39] More information: https://www.samlearning.com/.
[40] More information: http://www.laceproject.eu/lace/.
[41] IPRO is a visual programming environment based on 'drag and drop' functionality to create virtual soccer player robots to play with other virtual robots.

The table below provides several examples of intelligent tutoring systems. The examples illustrate that intelligent tutors can assess traditional key competences, as well as being useful for assessing transversal skills such as problem solving or critical thinking.

Table 7. Examples of intelligent tutoring systems

Tool	Description
AutoTutor (US)	'AutoTutor is an intelligent tutoring system that holds conversations with the human in natural language. AutoTutor has produced learning gains across multiple domains (e.g., computer literacy, physics, critical thinking). Three main research areas are central to AutoTutor: human-inspired tutoring strategies, pedagogical agents, and technology that support natural language tutoring'.[42] The methodologies are still being developed in the US. Despite the existing potential of the AutoTutor system, there is no evidence of its application in classrooms.
Kikora (NO)	Kikora is a next generation learning tool for mathematics instruction developed in Norway. Kikora enables students to learn and teachers to teach math more efficiently. Kikora provides immediate feedback on students' performance. The software also allows for tracking progress, and based on a pupil's performance, proposes critical learning areas that are most in need of improvement. In this way, teachers receive an elaborated overview and save time. Kikora stores a databank with approximately 3000 exercises for students in lower and upper secondary education. The tasks are strictly in accord with the requirements of the mathematics curriculum in Norwegian education.(Redecker, 2013, p. 33).[43]
Smartbooks at Frederiksvaerk school (DK)	The Frederiksvaerk school in Denmark uses SmartBooks. These are digital tutors that use artificial intelligence to tailor individual learning paths. The individual pathways are proposed based on reading comprehension after implementing an assignment (Adams Becker et al., 2017).
Mr Albert (SE)	In Sweden, more than 25 schools have tested Mr Albert – an artificial intelligence math tutoring system that provides personalised lessons to students (Adams Becker et al., 2017)[44].

Source: Adapted from Adams Becker et al., 2017; Redecker, 2013.

4.4.5.Game-based assessment

Game-based assessment is another assessment method that could be linked to both third and fourth generation assessments. There is growing interest in game-based assessment since it 'continues to demonstrate its effectiveness for learning for students of all ages', and 'the greatest potential of games for learning lies in their ability to foster collaboration, problem-solving, and procedural thinking' (Johnson et al., 2011, p. 7). Game-based learning can also foster complex-thinking and creativity skills (Akcaoglu, 2016). Game-based assessment tools integrated into the virtual environment are being developed. Game-based assessments can have a format of simple quizzes, but can also integrate more advanced methods. For instance, to measure middle-school students' problem-solving skills, Shute et al. (2016) embedded stealth assessment in a game called 'Use Your Brainz'. The researchers validated their

[42] More information: http://www.autotutor.org.
[43] More information: http://www.kikora.com/.
[44] More information: https://www.hejalbert.se/.

findings by collecting two external problem-solving measures: students who played the game for three hours were asked to complete Raven's Progressive Matrices and MicroDYN (Shute and Rahimi, 2017). These measures showed that their game-based assessments were valid. However, a larger validation study would be needed to support the findings and help develop tools further for teachers.

Another example is 'Kahoot!' – a web-based game developed in Norway, which is used for formative assessment in the classrooms. This is an interactive way of quizzing, where teacher can facilitate the game in a similar way as in TV shows. 'Kahoot!' is well suited for formative assessment, since teachers can easily assess acquired knowledge. Research conducted by Wang et al. (2016) revealed that 'Kahoot!' significantly improves learning motivation, engagement, enjoyment, and concentration for the gamified approach. However, the authors found no significant evidence of learning improvement.

Game-based assessment can also be useful for improving pupils' knowledge and developing digital skills (Admiraal, 2015; Wouters and van Oostendorp, 2013). Admiraal (2015) conducted a study in the Netherlands on teaching with a digital role-play game to increase students' reflective internet skills, which are a component of digital skills (Ferrari, 2013). Seventh grade students played a six-level game, and in each of the levels they played different roles, such as a manager, a marketing director, or a journalist. Pupils' skills were assessed over a period of time by self-assessment questionnaires. Participating students increased their reflective internet skills. The game had an even more significant effect for boys.

Table 8. Examples of game-based assessments

Tool	Description
Use Your Brainz (US)	'Use Your Brainz' is an educational game developed by GlassLab Games, Electronic Arts, and PopCap Games in the United States. 'Use Your Brainz' was transformed from the popular game 'Plants vs Zombies' to one that enables the assessment of problem-solving skills in the classroom. The main challenge of the game is to determine which plants could be used and where to place them to defeat all zombies. Log files integrated in the game helps researchers track students' progress.
Kahoot! (NO)	'Kahoot!' is a free game-based learning platform launched in 2003 in Norway. It is a student response (quizzing) system that focuses more on engaging and motivating students through attractive graphical user-interfaces and audio, and by gamifying student response experience. Students in the classroom participate in the game in the format of a TV show, where pupils can compete with each other. Teachers can use the game as a formative assessment tool, as it helps quickly assess if pupils acquired the knowledge.
SplitsZ! (NL)	'SplitsZ!' is a digital role play game used in the Netherlands that aims to develop students' reflective skills. The game is split into several stages, and pupils can play different roles at these stages. However, the game works rather like a learning tool, and learning outcomes were assessed only by questionnaires.

Source: Compiled by the authors based on Admiraal (2015), Shute et al. (2016) and Wang et al. (2016).

Based on the given examples, game-based learning can be used for measuring learning outcomes by integrating stealth assessment in the game; or it can serve as formative assessment tool to help teacher evaluate the acquisition of particular themes and proceed in teaching activities. At the current stage, games are employed more often as learning tools, but not assessment tools.

4.4.6.Augmented reality assessment

There is also increased interest in augmented reality, virtual worlds and immersive games to assess 21st century skills (Adams Becker et al., 2017; Redecker, 2013). Augmented reality (AR) complements physical reality by virtual realities. The method has the potential to supplement information delivered via computers, mobile devices, video, or a printed book (Johnson et al., 2011). Johnson et al. (2011) and Redecker (2013) provided several examples of how augmented reality can be applied across different disciplines: in chemistry, students can use handheld devices to explore a physical space to uncover clues and receive data related to a simulated environmental disaster detailed in a game-based scenario using AR simulations; in geography, students can study an augmented globe in a textbook; in history, images can be viewed when visiting particular locations. According to Adams Becker et al. (2017), AR tools could already be adopted in Nordic schools by 2019 or 2020. However, most of the available examples lack information on opportunities to use augmented reality for assessing skills.

Meaningful integrations of ICT tools into assessment practices are often linked to their interaction with physical objects, various environments, peers and digital information. Interactive activities can be monitored and automatically assessed by using mobile devices (Santos et al., 2015). However, the question of how to use ICT for assessing and forming skills remains. Santos et al.. (2015) argue that the application of mobile devices can be meaningful in forming skills and competences only when the assessment is supported by scaffolding[45] mechanisms and when assessment is done before, during and after the activity. Abrams and Gerber (2013) also note that assessment practices using ICT are effective when a learner is involved as an assessor as well.

A systematic literature review on assessment instruments of primary and secondary school students' ICT literacy conducted by Siddiq et al. (2016) showed that even digital competences lack sufficient assessment. The majority of the tests were designed to assess components of ICT literacy such as searching, retrieving, and evaluating digital information, and technical skills. These components cover only a part of digital competence, which is defined as consisting of five components: information and data literacy; communication and collaboration; digital content creation; safety; and problem solving (Ferrari, 2013).

Although researchers have forecast the emerging potential of augmented reality assessment and learning analytics, these methods are still in a trial period. It will still take several years for the first schools to adopt these methods and assess their effectiveness.

Key messages

- **The increasing use of ICT in teaching and learning implies new models of assessment.** The technology significantly facilitates the use of summative, e.g., by enhancing their reliability and time-effectiveness, and formative assessments, by allowing for immediate feedback and active students' participation.
- **Standardised e-assessments are already well-applied to deliver traditional assessment more effectively and quickly**. These assessments evaluate well traditional competences such as mathematics, communication in mother tongue and foreign language competences. Computer-adaptive tests are slightly more advanced, since they can also adjust tasks according to the level of a student.

[45] Scaffolding 'refers to a variety of instructional techniques used to move students progressively toward stronger understanding and, ultimately, greater independence in the learning process'. (The Great School Partnership website. Available at: http://edglossary.org/scaffolding/).

- More innovative technology-based assessments **shift focus from efficient testing to personalised learning** and provide opportunities to assess different competences, including the learning to learn competence, and transversal skills such as problem-solving, critical thinking and creativity skills. This way, the research proved that e-portfolios are powerful tools for improving language competences, as well as cultural awareness and expression; game-based assessment demonstrates its effectiveness in fostering collaboration, problem-solving and procedural thinking, as well as creativity skills. The application of learning analytics shows great potential to monitor and identify pupils' key strengths and the skills that need to be most improved. Intelligent tutoring systems bring the possibility to provide personalised and timely feedback.
- On the one hand, new technologies opened opportunities to capture complex competences and skills; on the other hand, specific tools often lack profound testing and more systematic application. Therefore, they need to be **used critically by assessing their actual effectiveness on enhancing skills. Using technologies for classroom assessment should be accompanied by effective feedback, and if possible, scaffolding mechanisms**. Assessment practices using ICT are effective when a learner is involved as an assessor as well.
- Apart from standardised e-assessments, **the pool of e-assessment toolkits in Europe is rather fragmented and their effectiveness often needs further research**. Practitioners would benefit from more systematic application of technology-based tools into the curricula across EU Member States.

5. LESSONS FOR POLICY AND PRACTICE

In recent decades, education systems have shifted from traditional content-based education towards a more comprehensive and cross-cultural competence-based orientation of learning. In many countries, curricula are more and more defined not only in terms of knowledge in different academic subjects, but also in terms of attitudes, skills, behaviours and values. The benefit of competence-based learning lies in its emphasis on the constructivist vision of the learner developing his or her own competences and skills, being able to mobilise and critically reflect on his or her knowledge.

Social and economic transformations of contemporary knowledge societies demand new ways of thinking and learning. Knowledge alone is not sufficient; a broader range of skills and abilities is needed to navigate a moving landscape characterised by the growing importance of ICT, the decline of functional skills-based professions and increased competition. These skills and competences are no longer associated with a certain academic discipline, but are instead transversal and multi-dimensional in nature. The European key competence framework identified eight key competences, further supported by seven transversal skills necessary for personal fulfilment, active citizenship and all-rounded development in the 21st century.

While implementation of the key competence framework has been on the agenda of all Member States, evidence shows that policies and practices for the assessment of these competences have yet to be fully implemented in EU countries' efforts to integrate competence-based education into school curricula.

Assessing key competences and transversal skills is a challenging task, as they refer to complex constructs that are not easily measurable. Although assessment policies supporting the assessment of traditional key competences such as maths, languages and science have largely been implemented, they are often limited to the contexts provided by the subject matters with which they are most closely associated, and rarely assess related attitudes. Cross-curricular competences and transversal skills are harder to associate with individual subjects and to be reflected in specific learning outcomes. An innovative approach to assessment practices is needed to grasp the complexity and multiple roles of modern learning.

Nevertheless, inspiring policies and initiatives do exist at local levels and are currently being developed. They can lay the foundation for the comprehensive implementation of competence-based education in European countries. This report has selectively reviewed current practices and approaches focused on the assessment of different types of key competences. As a result of this review, the following conclusions and recommendations can be drawn.

5.1. The importance of a comprehensive vision of assessment at the policy and practice level

- **The field of educational assessment is currently divided and fragmented** into differing and often competing theories, methods and approaches: quantitative versus qualitative, formative versus summative, norm-referenced versus criterion/standards-referenced, tests versus feedback, internal versus external, measurement versus judgement, etc. However, at the same time more education stakeholders realise that assessment is a process, which aims not only to document learning, but actually inform and improve it, and therefore it needs to be guided by theories, models, and data. In this light, different functions of assessment need to be seen in synergy with each other and built into comprehensive assessment frameworks. Integrating different approaches to assessment can allow finding of a balanced and consistent practice for assessing key competences.

Recommendations

- Policy-makers and schools should promote and employ an **integrated approach towards classroom assessment**, capitalising on the benefits and opportunities of summative, formative and diagnostic assessment to grasp the multi-dimensional nature of key competences and transversal skills. The promotion of only one particular type of assessment may limit the implementation of competence-based curricula and the comprehensive assessment of non-traditional competences and transversal skills.
- Re-forming summative assessment methods based on grading, by integrating them with formative methods, could also help focus the learning process on the obligation to offer the same potential progress to all students, emphasising pupils' capacities, rather than constructed deficits.
- Assessment practices need to **both document the learners' competences and at the same time help develop them,** by modifying teachers' practices and curricula focus.

› Assessment practices are important in **shaping teaching and learning**. Therefore, particular assessment policy at national and school level can indicate what type of learning is important, and to what aspects of learning more time and effort need to be devoted. Changes in assessment can therefore structure teaching priorities and methods, and in turn, impact what and how learners learn.

Recommendations

- Educational assessment policy should have a clear understanding that **assessment also implies learning**, taking into consideration its interconnectedness with curriculum development and teacher education.
- **Reaching a shared understanding of standards and expectations** through discussion and dialogue is a key element for teaching and ensuring a robust and credible assessment system.

› Apart from conceptual and instructional considerations, all assessment practices used in the classroom need to adhere to overall technical requirements, such as **being valid and reliable**. Assessments and the qualifications they underpin are increasingly important because they enable individuals to access further life opportunities - especially in employment and continuing levels of education. There is growing awareness of the importance of assessment and certification processes being equitable, fair and inclusive. Hence, in addition to validity and reliability, **transparency, equity and freedom from bias** are key requirements for effective assessment.

Recommendations

- When designing the integrated assessment framework, teacher and policy-makers should **find a balance between reliability and overall validity of assessment, taking into account the assessment purpose.** In this regard, employing both summative methods, which have stronger reliability, in combination with formative methods that strengthen the overall validity of the assessment approach is important.
- Assessments should allow all students to demonstrate what they know and can do without being unfairly disadvantaged by individual characteristics that are irrelevant to what is being assessed.

5.2. The need to develop an effective implementation framework for the assessment of key competences in classrooms

- **Defining key competences in terms of detailed and concrete learning outcomes can provide a good basis for consistent assessment practices**. In the process of implementation of key competence framework many national education policies tended to focus on 'traditional' key competences in a limited range of subject areas rather than throughout the curriculum. Nevertheless, learning outcomes defined in curricular documents, education standards or legislation, increasingly cut across subject boundaries and refer to key competences. Making learning outcomes specific can help adapt learning and assessment practices.

Recommendations

- **Key competences should be well defined in relevant policy documents and/or school curricula**, including a wide range of clear, concrete and detailed learning outcomes linked to teaching practice, taking into account cognitive and non-cognitive components.
- Assessment methods used at classroom level should be **appropriately adapted to key competences**, notably in terms of sub-competences and learning outcomes.

- Research underlines the role of teacher education (ITE, induction and CPD) in providing teachers with a common understanding of key competences and relevant assessment practices to assess student learning. **Integrating assessment practices at the ITE level and during practice** in schools is needed to make teacher education programmes more coherent, better connect theory and practice, and help increase teachers' assessment literacy. **Collaborative learning environments can be effective tools to support teachers** at all stages of the teacher education continuum. By helping to encourage reflection, exchange and support for improving classroom practice, teacher learning communities (TLCs) can support the implementation of formative assessment methods, and have beneficial impacts on teachers, students and schools. School leaders and staff, school networks, municipalities, and regional as well as national authorities can all be influential in supporting and monitoring TLCs to reflect upon and enhance the effectiveness of assessment practices.

Recommendations

- **Teacher education systems should better support teachers** at all stage of their careers **to effectively put into practice classroom assessment methods** that help to assess key competences and improve student learning.
- ITE courses, induction and CPD programmes should focus on a **broad range of complementary assessment methods for formative and summative purposes** to enhance teachers' assessment literacy.
- Teacher educators should be offered initial training and CPD programmes to develop their assessment competences and better prepare (student) teachers accordingly.
- **Collaborative learning environments such as teacher learning communities (TLCs) should be promoted at all stages of the teacher education continuum** to support teacher professional development and assessment competences.
- **Support structures tailored to local contexts should be created** to effectively support teachers in implementing assessment methods for formative and summative purposes. Local, regional and/or national authorities should allocate resources, space and time devoted for CPD and TLCs to support and sustain knowledge sharing in the area of classroom assessment.

- Bottom-up implementation processes based on trust, dialogue and high level of teacher agency can be key factors to support the successful large-scale implementation of innovative assessment practices in schools.

- **The effective use of assessment results depends on teachers' assessment literacy and their ability to appropriately integrate assessment data in their teaching**. Assessment results are used for both summative and formative purposes to store information, provide feedback to students and make decisions about their educational trajectory. Formative assessment information can be used through long- and medium-term methods to document and monitor student learning progress, notably towards subject-specific and transversal competences. However, short-cycle formative assessment methods such as daily interactions between and among students and teachers are reported to bring the most direct and measurable impact on student achievement.

Recommendations

- **Streaming and grade repetition decisions based on summative assessment results should be avoided** as they tend to reduce student learning opportunities and fail to raise student educational outcomes.
- The use of assessment results for formative purposes in short, medium and long cycle can be helpful to document students' learning progress and help to improve their academic achievement.

5.3. The relevance of employing a combination of assessment approaches for documenting and supporting different types of key competences

- Research demonstrates that **there is no single method that would fully measure key competences and transversal skills, nor serve as a best practice for student assessment**. To assess a variety of skills, several methods and types of assessment need to be used. The effectiveness of a method depends on its purposes and design, as well as on schools' and teachers' capacity to use it.
  At the same time, there is no universal combination of methods that would serve as a recipe. Teachers may be rather flexible in their choice of methods, so long as they serve multiple purposes and follow the principles of validity, reliability and equity.

Recommendations

- At the policy level, there is a need for **a clear vision and strategy for assessment** where different approaches developed nationally and locally serve clearly defined purposes, and where the format of assessment is aligned to that purpose.
- There is a need to develop clear goals and reference points to guide student assessment at the classroom level.
- Schools should be provided with **comprehensive guidance** concerning valued learning outcomes in national curricula and standards.
- Education policy needs to ensure a good balance between formative and summative assessment, utilising the benefits of both, without over-relying on one method.

- Although they represent a predominant method of classroom assessment, standardised assessments can face challenges in evaluating key competences and transversal skills, if narrowly focused solely on the content of the subject being assessed. Our review demonstrates that standardised

tests, requiring multiple reasoning techniques, can capture a broader range of skills, such as digital competence, social and civic competences, as well as transversal skills such as creativity and problem-solving, which are necessary to complete complex tests. Other formats to measure non-traditional competences include tests with open-ended questions and surveys measuring social and emotional well-being at school.

Well-designed multiple-choice assessments can assess higher-order knowledge and provide **more than a summative check of knowledge and skills.** Some assessments also use multiple-choice items to measure critical thinking. On the other hand, literature demonstrates that **there are competences that cannot be assessed simply using multiple choice, such as complex problem-solving skills or the ability to compose an essay.** Other types of assessments can help address such limitations.

The standardised assessment of attitudes as learning outcomes remains challenging. Attitudinal questionnaires tend not to capture various social contexts and complex emotional experiences, and therefore lack information related to learning outcomes associated to specific contexts. However, valid and reliable teacher rating attitudinal assessment tools can help to assess diverse social-emotional competences.

A combination of different assessment methods, such as self-assessment via questionnaires, and assessments via tests for formative or summative purposes, could better help to assess social and emotional competences such as learning to learn.

Recommendations

- The assessment framework developed in schools should allow teachers to **draw on multiple sources of evidence** in order to form their best judgments on students' levels.
- Education policy needs to ensure consistency of assessment and grading across schools. When designing standardised assessment techniques, it is important to take into account the structure and content that reproduce real-life contexts, integrating multiple steps and formats, requiring a chain of reasoning and competences.
- Integrating ICT tools into standardised assessment practices can help not only make the process of assessment more efficient, but also transform the design and the capacity of the method, integrating both summative and formative functions.
- Schools' capacity to use **the potential of ICT for developing sophisticated assessment instruments should be better supported**.

› Performance-based assessment can serve both summative and formative purposes. By employing a variety of methods, such as rubrics, project-based assessment and portfolios, this approach can help assess a larger variety of competences and skills. Research demonstrates that rubrics ensure a clear scale to measure a specific learner's competence, while **project-based learning can effectively enhance skills such as critical thinking, problem solving, communication, collaboration, and self-management.** Performance-based assessment is effective in responding to an individual learner's needs, because it can evaluate a learner's progress from his or her starting position.
Research evidence also shows that assessment methods can only help develop certain skills when followed by feedback. The length of and time when feedback is provided play an important role in its effectiveness.

Recommendations

- **The competences that need to be assessed should be clearly defined**, and clear goals should support pupils advancing in a particular competence. Portfolios and rubrics can be helpful for forming goals and monitoring student progress. The systematic development of these methods should be supported.
- Teachers should be supported in order to change practices in the classroom to **'unpack' the competences to be taught and assessed**.
- Performance-based assessments need to include clearly defined judgement criteria to effectively assess learners' competences.
- **Although a variety of toolkits supporting performance-based assessment exists, the** European market is quite fragmented **and a more centralised approach would be helpful for practitioners.**

- **Successful teacher, peer and self-assessments are useful in building many key competences**, such as initiative and entrepreneurship, learning to learn and social competence, **as well as transversal skills,** such as critical thinking, creativity, problem-solving, risk assessment, decision-taking, and constructive management of feelings. Peer and self-assessment is considered to be a powerful tool to engage students in active learning, reflection and raise their motivation and academic standards. However, for these methods to be effective, there is a need for careful planning and accurate teachers' feedback to enhance students' self-regulated skills necessary for these types of assessments.

Recommendations

- **Integrating peer-and self-assessment as both assessment and learning process** in the classroom activities, help in developing learners' self-regulation, critical thinking and initiative.
- However, for these assessments methods to be effective, **it is important to provide comprehensive training to teachers** as observers and supporters of the assessment process, **and students** as the main assessors. It needs to be recognised that being an accurate assessor is a competence in itself.
- Criteria to judge performance in relation to national goals and learning outcomes should be better clarified and illustrated when using interactive formative assessment methods.
-

- Although technology-based assessment is receiving more and more interest from researchers, practitioners and policy makers, **only first- and second-generation e-assessments**, such as standardised tests, multiple-choice assessments and adaptive tests, **are widely applied at classroom level**. Among third-generation assessment methods, online collaboration platforms and e-portfolios have been introduced by some schools.
  Use of ICT in assessment allows to deliver traditional assessment more effectively and faster and at the same time offer opportunities to change the way competences are assessed, finding effective solutions for assessing non-traditional competences. This way, the research proved that e-portfolios are powerful tools for improving language competences, as well as cultural awareness and expression; game-based assessment demonstrates its effectiveness in fostering collaboration, problem-solving and procedural thinking, as well as creativity skills. The application of learning analytics shows great potential to monitor and identify pupils' key strengths and the skills that need to

be most improved. Intelligent tutoring systems bring the possibility to provide personalised and timely feedback.

Recommendations

- Although technologies play an important role in contemporary classrooms, they should be carefully implemented. **Using technologies for formative purposes should be accompanied by effective feedback, and if possible, scaffolding mechanisms**. Before choosing technologies, clear goals need to be set.
- **Policy makers should promote the introduction of innovative tools in the classroom and integrate specific training to meaningfully use these tools into teacher education programmes.**
- Apart from standardised e-assessments, **the pool of e-assessment toolkits in Europe is rather fragmented and their effectiveness often needs further research**. Practitioners would benefit from a more systematic application of technology-based tools into the curricula across EU Member States.

REFERENCES

Abrams, S. S.; Gerber, H. R. (2013). 'Achieving through the Feedback Loop: Videogames, Authentic Assessment, and Meaningful Learning'. *The English Journal*, *103*(1), 95–103.

Adams Becker, S.; Cummins, M.; Freeman, A.; Rose, K. (2017). *2017 NMC Technology Outlook for Nordic Schools: A Horizon Project Regional Report*. Austin, Texas: The New Media Consortium.

Admiraal, W. (2015). 'A Role-Play Game to Facilitate the Development of Students' Reflective Internet Skills'. *Journal of Educational Technology & Society*, *18*(3), 301–308.

Agelii Genlott, A.; Grönlund, Å. (2016). 'Closing the gaps: Improving literacy and mathematics by ict-enhanced collaboration', *Computers and education*, ISSN 0360-1315, E-ISSN 1873-782X

Airasian, P. W. (1994). *Classroom assessment* (2nd edition). New York, McGraw-Hill.

Akcaoglu, M. (2016). 'Design and Implementation of the Game-Design and Learning Program'. *TechTrends*, *60*(2), 114–123. https://doi.org/10.1007/s11528-016-0022-y

Alexander, P. A.; Winne, P. H. (2006). *Handbook of Educational Psychology*. New Jersey: Lawrence Erlbaum.

Ananiadou, K.; Claro, M. (2009). '21st Century Skills and Competences for New Millennium Learners in OECD Countries', *OECD Education Working Papers*, No. 41, OECD Publishing, Paris.

Assessment Reform Group (2006). 'The role of teachers in the assessment of learning'. Supported by the Nuffield Foundation. Available at: http://www.nuffieldfoundation.org/sites/default/files/assessment_booklet.pdf.

Baartman, L. K. J.; Bastiaens, T. J.; Kirschner, P. A.; Van der Vleuten, C.P.M. (2006). 'The wheel of competency assessment: Presenting quality criteria for competency assessment programs', *Studies in Educational Evaluation, Vol. 32*, pp. 153-170.

Baird, J. (2014). 'Assessment and attitude', *Assessment in Education: Principles, Policy & Practice, 21:2,* pp. 129-132, DOI: 10.1080/0969594X.2014.904076.

Baird, J.; Hopfenbeck, T. N.; Newton, P. E.; Steen-Utheim, A. T. (2014). *State of the Field Review: Assessment and Learning*, Oxford University Centre for Educational Assessment Report OUCEA/14/2.

Barlow, A.; Marolt, A. (2012). Effective Use of Multiple-Choice Items in the Mathematics Classroom, Middle School Journal, 43:3, 50-55.

Begeny, J. C.; Martens, B. K. (2006). 'Assessing Pre-Service Teachers' Training in Empirically-Validated Behavioral Instruction Practices', *School Psychology Quarterly*, Vol. 21, No. 3, pp. 262-285.

Bell, S. (2010). 'Project-based learning for the 21st century: Skills for the future'. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*, 39–43.

Bennett, R. E. (2010). 'Technology for Large-Scale Assessment'. In P. Peterson, E. Baker & B. McGaw (Eds.), *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 48-55), Oxford: Elsevier.

Bennett, R. E. (2011). 'Formative assessment: a critical review', *Assessment in Education: Principles, Policy and Practice*, Vol. 18., No. 1, pp. 5-25.

Bennett, R. E. (2015). 'The Changing Nature of Educational Assessment', *Review of Research in Education*, Vol. 39, pp. 370-407.

Bennett, R. E.; Deane, P.; Rijn, P. W. van. (2016). 'From Cognitive-Domain Theory to Assessment Practice'. *Educational Psychologist*, *51*(1), 82–107. https://doi.org/10.1080/00461520.2016.1141683

Bennett, R. E.; Gitomer, D. H. (2009). 'Transforming K-12 assessment: Integrating accountability testing, formative assessment, and professional support'. In C. Wyatt-Smith and J. Cumming (eds.), *Educational assessment in the 21st century*, New York: Springer, pp. 43–61.

Binkley, M.; Erstad, O.; Herman, J.; Raizen, S.; Ripley, M.; Rumble, M. (2010*). Draft White Paper 1: Defining 21st Century Skills*, http://atc21s.org/wpcontent/uploads/2011/11/1-Defining-21st-Century-Skills.pdf.

Binkley, M.; Erstad, O.; Herman, J.; Raizen, S.; Ripley, M.; Miller-Ricci, M.; Rumble, M. (2012). 'Defining Twenty-First Century Skills'. In B. McGaw & E. Care (eds.), *Assessment and Teaching of 21st Century Skills,* New York, NY: Springer, pp. 17-.66.

Black, P. (1998). *Testing: friend or foe? Theory and practice of assessment and testing.* London, Falmer Press.

Black, P.; Wiliam, D. (1998a). 'Assessment and Classroom Learning', *Assessment in Education: Principles, Policy & Practice*, 5 (1), 7–74.

Black, P.; Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. In King's College London School of Education (Ed.). London: GL Assessment.

Boilard, S. D. (2011). 'Perspectives: Prior learning assessment challenges the status quo', *Change*, *43*, pp. 56–59.

Bourke, R. (2015). 'Liberating the learner through self-assessment', *Cambridge Journal of Education*, Vol. 46, No. 1, pp. 97-111.

Bourke, R.; Mentis, M. (2014). 'An assessment framework for inclusive education: integrating assessment approaches', *Assessment in Education: Principles, Policy and Practice*, Vol. 21, No. 4, pp. 384-397.

Brecko, B. N.; Kampylis, P.; Punie, Y. (2014). *Mainstreaming ICT-enabled Innovation in Education and Training in Europe: Policy actions for sustainability, scalability and impact at system level*. JRC Scientific and Policy Reports. Seville: JRC-IPTS, doi:10.2788/52088.

Brevik, L. M.; Blikstad-Balas, M.; Engelien, K. L. (2017). ,Integrating assessment for learning in the teacher education programme at the University of Oslo'. *Assessment in Education: Principles, Policy & Practice*, *24*(2), 164–184. https://doi.org/10.1080/0969594X.2016.1239611

Bristow, S.; Patrick, S. (2014). *An International Study in Competency Education: Postcards from Abroad*, International Association for K–12 Online Learning.

Brookhart, S. M. (2011). 'Educational Assessment Knowledge and Skills for Teachers', *Educational Measurement: Issues and Practice*, Vol. 30, No. 1, pp. 3-12.

Brookhart, S. M. (2001). 'Successful Students' Formative and Summative Uses of Assessment Information', *Assessment in Education: Principles, Policy & Practice, 8:2*, pp. 153-169.

Brown, G.; Andrade, H. L.; Chen, F. (2015), Accuracy in student self-assessment: directions and cautions for research, *Assessment in Education: Principles, Policy & Practice*, DOI: 10.1080/0969594X.2014.996523

Brown, G. T. L.; Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), The Sage handbook of research on classroom assessment (pp. 367–393). Thousand Oaks, CA: Sage

Brown, G. T. L.; Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*.

Budginaitė, I.; Siarova, H.; Sternadel, D.; Mackonytė, G.; Spurga, S. (2016). 'Policies and practices for more equality and inclusion in and through education: Evidence and policy guidance from European research projects funded under FP6 and FP7', NESET II report, Luxembourg: Publications Office of the European Union.

Bunderson, V. C.; Inouye, D. K.; Olsen, J. B. (1989). 'The four generations of computerized educational measurement'. In R. L. Linn (Ed.), *Educational measurement (3rd ed., pp. 367-407)*. New York: Macmillan.

Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. *Studies in Educational Evaluation*. https://doi.org/10.1016/j.stueduc.2014.03.002

Buscà Donet, F.; Ambròs Pallares, A.; Burset Burillo, S. (2017). 'Bibliometric characteristics of articles on key competences indexed in ERIC from 1990 to 2013', *European Journal of Teacher Education, 40:2*, pp. 144-156.

Caena, F. (2014). 'Teacher Competence Frameworks in Europe: policy-as-discourse and policy-as-practice', *European Journal of Education*, Vol. 49, No. 3, pp. 311-331.

Camacho, D. J.; Legare, J. M. (2016). 'Shifting gears in the classroom—movement toward personalized learning and competency-based education', *The Journal of Competency-based Education*, 1, pp. 151–156.

Campbell, C.; Evans, J. A. (2000). 'Investigation of Preservice Teachers' Classroom Assessment Practices During Student Teaching', *The Journal of Educational Research*, Vol. 93, No. 6, pp. 350-355.

Cayton-Hodges, G. A.; Feng, G.; Pan, X. (2015). 'Tablet-Based Math Assessment: What Can We Learn from Math Apps?', *Journal of Educational Technology & Society*, *18*(2), pp. 3–20.

Cedefop (2009). *The shift to learning outcomes: policies and practices in Europe*. Luxembourg: Publications Office, Cedefop reference series; No. 72.

Cedefop (2014). *Terminology of European education and training policy. A selection of 130 key terms.* Second edition, Luxembourg: Publications Office of the European Union.

Cedefop (2016). *Application of learning outcomes approaches across Europe: a comparative study.* Luxembourg: Publications Office of the European Union. Cedefop reference series; No. 105.

Chen, C. M.; Chen, M. C. (2009). 'Mobile formative assessment tool based on data mining techniques for supporting web-based learning'. *Computers and Education, 52(1),* pp. 256-273.

Cho, K.; MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20(4)*, pp. 328-338.

Cizek, G. (2010). An Introduction to Formative Assessment: History, Characteristics, and Challenges. In: H. L. Andrade and G. J. Cizek (eds.) Handbook of Formative Assessment. New York, Routledge.

Cornu, O.; Decaudin, F.; Doignon, P.; Levrault, V. (2014). 'L'évaluation sans notes dans un lycée expérimental', Democratisation-scolaire.fr, GRDS. Available at : https://www.democratisation-scolaire.fr/spip.php?article197.

Crossouard, B. (2011). 'Using formative assessment to support complex learning in conditions of social adversity', *Assessment in Education Principles Policy and Practice Policy & Practice (1):59-72*.

Crooks, T. J. (1988). 'The impact of classroom evaluation practices on students', *Review of Educational Research, 58, 4*, pp. 438-81.

Dabrowski, M.; Wisniewski, J. (2011). 'Translating key competences into the school curriculum: lessons from the Polish experience', *European Journal of Education*, *46, 3*.

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

De-Juanas Oliva, A.; Martin del Pozo, R.; Pesquero Franco, E. (2016). 'Teaching competences necessary for developing key comepetences of primary education students in Spain: teacher assessments', *Teacher Development, Vol. 20, No. 1*, pp. 123-145.

De Luca, C.; Klinger, D. A. (2010). 'Assessment literacy development: Identifying gaps in teacher candidates' learning'. *Assessment in Education: Principles, Policy & Practice*, 17 (4), pp. 419-438.

De Luca, C.; Lam, C. Y. (2014). 'Preparing Teachers for Assessment within Diverse Classrooms: An Analysis of Teacher Candidates' Conceptualizations', *Teacher Education Quarterly*, Vol. 41, No. 3, pp. 3-24.

De Luca, C.; LaPointe-McEwan, D.; Luhanga, U. (2016). 'Teacher assessment literacy: a review of international standards and measures', *Educational Assessment, Evaluation and Accountability*, Vol. 28, No. 3, pp. 251-272.

Dixson, D. D.; Worrell, F. C. (2016). 'Formative and Summative Assessment in the Classroom'. *Theory Into Practice*, *55*(2), pp. 153–159. https://doi.org/10.1080/00405841.2016.1148989

Doolittle, A. E. (2002). 'Classroom Assessment: What Teachers Need to Know' (Second Edition), W. James Popham, Boston: Allyn & Bacon, 1999 (Reviewed by A. E. Doolittle), *Journal of Educational Measurement*, Vol. 39, No. 1, pp. 85-90.

Drake, S. M.; Reid, J. L.; Kolohon, W. (2014*). Interweaving curriculum and classroom assessment: Engaging the 21st century learner*. Canada: Oxford University Press.

Dunkel, P. A. (1999). 'Considerations in Developing or Using Second/Foreign Language Proficiency Computer-Adaptive Tests', *Language Learning & Technology, Vol. 2, No. 2,* pp. 77-93

Dziedzic, M.; Janissek, P. R.; Bender, A. P. (2008). *Assessment by peers - An effective learning technique*, In: 38th Annual Frontiers in Education Conference, Saratoga Springs, NY. Proceedings of the 38th Annual Frontiers in Education Conference. Piscataway, NJ: IEEE, 2008. v. 1. pp. 978-978.

Earl, L. M.; Katz, S. (2006). *Rethinking classroom assessment with purpose in mind*. Retrieved from http://www.edu.gov.mb.ca/k12/assess/wncp/full_doc.pdf.

Earl, L. M. (2013). *Assessment as Learning: Using classroom assessment to maximize student learning*. Corwin 2013.

EPPI (2002). *A systematic review of the Impact of Summative assessment and tests on students' motivation for learning*, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.

European Commission (2011). *Literature review. Teachers' core competences: requirements and development*. Prepared by Francesca Caena for the ET2020 Thematic Working Group 'Professional Development of Teachers'. Brussels: DG EAC. Available at: http://ec.europa.eu/education/policy/strategic-framework/doc/teacher-competences_en.pdf.

European Commission (2012a). Staff Working document 'Assessment of Key Competences in initial education and training: Policy Guidance', SWD (2012) 371 final. European Commission, 'Shaping career-long

perspectives on teaching. A guide on policies to improve Initial Teacher Education'. ET2020 Working Group on Schools Policy (2014-15), DG Education and Culture. Luxembourg: Publications Office of the European Union.

European Commission (2012b). Education and Training 2020 Work Programme. Thematic Working Group 'Assessment of Key Competences', Literature review, Glossary and examples. Available at: http://ec.europa.eu/dgs/education_culture/repository/education/policy/school/doc/keyreview_en.pdf.

European Commission (2015). Shaping career-long perspectives on teaching: A guide on policies to improve Initial Teacher Education. Report prepared by ET2020 Working Group on Schools Policy (2014/15).

European Commission (2016a). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *A New Skills Agenda for Europe. Working together to strengthen human capital, employability and competitiveness*. COM (2016) 381 final.

European Commission (2016b). *PISA 2015: EU performance and initial conclusions regarding education policies in Europe.* Available at: https://ec.europa.eu/education/sites/education/files/pisa-2015-eu-policy-note_en.pdf.

European Commission (2017a). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, School development and excellent teaching for a great start in life, COM/2017/0248 final.

European Commission (2017b). Reflection paper on the social dimension of Europe, COM/2017/0206 final.

European Commission/EACEA/Eurydice (2012). *Developing Key Competences at School in Europe: Challenges and Opportunities for Policy*. Eurydice Report. Luxembourg: Publications Office of the European Union.

European Commission/EACEA/Eurydice (2016). Entrepreneurship Education at School in Europe. Eurydice Report. Luxembourg: Publications Office of the European Union.

European Parliament and Council of the EU, Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. Official Journal of the European Union, L 394/10, 30 December 2006.

Eurydice (2011). *Key Data on Learning and Innovation through ICT at School in Europe 2011*. Education, Audiovisual and Culture Executive Agency.

Ferrari, A. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*, European Commission, Joint Research Centre Institute for Prospective Technological Studies. Luxembourg: Publications Office of the European Union.

Florez, M. T.; Sammons, P. (2013). Assessment for learning: effects and impact, CfBT Education Trust.

Frejd, P. (2013). 'Modes of modelling assessment - a literature review'. *Educational Studies in Mathematics*, *84*(3), 413–438.

Gan, M. S.; Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science, 1–18*. doi:10.1007/s11251-014-9319-4

Gardner, J. (2010). 'Developing teacher assessments: An introduction'. In J. Gardner, W. Harlen, L. Hayward, G. Stobart, M. Montgomery (eds.), *Developing teacher assessment* (pp. 1–11). New York, NY: Open University Press.

Gordon, J.; Halász, G.; Krawczyk, M.; Leney, T.; Michel, A.; Pepper, D.; Wiśniewski, J. (2009). 'Key competences in Europe: Opening doors for lifelong learners across the school curriculum and teacher education'. *CASE network Reports*, (87).

Gotch, C.; M.; French, B. F. (2014). 'A Systematic Review of Assessment Literacy Measures', *Educational Measurement: Issues and Practice*, Vol. 33, No. 2, pp. 14-18.

Graham, S.; Hebert, M.; Harris, K. R. (2015). Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal*, *115*(4), pp. 523–547. https://doi.org/10.1086/681947

Grayson, H. (2014). *KeyCoNet's review of the literature: A summary*. KeyCoNet.

GRDS (2012). *L'école commune. Propositions pour une refondation du système éducatif*. Paris: La Dispute.

Haggerty, K.; Elgin, J.; Woolley, A. (2011). *Social-Emotional Learning Assessment Measures for Middle School Youth*. Seattle: Social Development Research Group, University of Washington.

Halasz, G. (2016). 'Research for CULT Committee: Evaluation of education at the European level'. Report requested by the European Parliament's Committee on Education and Culture. Brussels: European Parliament.

Halász, G.; Michel, A. (2011). 'Key Competences in Europe: interpretation, policy formulation and implementation'. *European Journal of Education* 46, no. 3, 2011: 289-306.

Harlen, W. (2007). *Assessment of Learning*. London: Sage Publications.

Hao, S.; Johnson, R. L. (2013). Teachers' classroom assessment practices and fourth-graders' reading literacy achievements: An international study. *Teaching and Teacher Education*, *29*, 53–63. https://doi.org/10.1016/j.tate.2012.08.010

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

Hattie, J.; Timperley, H. (2007). 'The power of feedback', *Review of Educational Research, 77 (1)*, pp. 81–112.

Hayward, L. (2015). 'Assessment is learning: the preposition vanishes', *Assessment in Education: Principles, Policy & Practice*, 22:1, pp. 27-43.

Heritage, M. (2010*). Formative assessment: Making it Happen in the Classroom, Corvin Press*, Thousand Oaks, California.

Herman, J. L.; Osmundson, E.; Silver, D. (2010). *Capturing quality in Formative assessment practice: Measurement challenges*, CRESST Report 770, National Centre for Research on Evaluation, Standards and Student Testing, Los Angeles.

Herzog-Punzenberger, B.; Le Pichon-Vorstman, E.; Siarova, H. (2017). 'Multilingual Education in the Light of Diversity: Lessons Learned', NESET II report, Luxembourg: Publications Office of the European Union. doi: 10.2766/71255.

Hill, P.; Barber, M. (2014). *Preparing for a renaissance in assessment*, Pearson, New York. Available at: http://gr8dbl.doverbay.ca/wp-content/uploads/2015/04/Preparing_for_a_Renaissance_in_assessment.pdf.

Hipkins, R. (2007). *Assessing Key Competences: Why would we? How could we?*, Review prepared for the Ministry of Education.

Hopfenbeck, T. N.; Florez Petour, M. T.; Tolo, A. (2015). 'Balancing tensions in educational policy reforms: large-scale implementation of Assessment for Learning in Norway', *Assessment in Education: Principles, Policy & Practice*, Vol. 22, No. 1, pp. 44-60.

Hopfenbeck, T. N.; Tolo, A.; Florez, M. T.; El Masri, Y. (2013). *Balancing Trust and Accountability? The Assessment for Learning Programme in Norway: A governing complex education systems cases study.* OECD Education Working Papers, No. 97. Paris: OECD Publishing.

Hou, H. T.; Chang, K. E.; Sung, Y. T. (2007). 'An analysis of peer assessment online discussions within a course that uses project-based learning'. *Interactive Learning Environments, 15(3), pp.* 237-251.

Illeris, K. (2015). 'The Development of a Comprehensive and Coherent Theory of Learning.' *European Journal of Education*, Vol. 50, No 1, pp. 29-40.

Johnson, L.; Adams Becker, S.; Estrada, V.; Freeman, A.; Kampylis, P.; Vuorikari, R.; Punie, Y. (2014). *Horizon Report Europe: 2014 Schools Edition*. Luxembourg: Publications Office of the European Union, & Austin, Texas: The New Media Consortium.

Johnson, L.; Adams, S.; Haywood, K., (2011). *The NMC Horizon Report: 2011*. K-12 Edition. Austin, Texas: The New Media Consortium

Jonsson, A.; Lundahl, C.; Holmgren, A. (2015). 'Evaluating a large-scale implementation of Assessment for Learning in Sweden'. *Assessment in Education: Principles, Policy & Practice*, Vol. 22, No. 1, pp. 104-121.

Kaufman, J. H.; Schunn, C. D. (2010). 'Students' perceptions about peer assessment for writing: their origin and impact on revision work'. *Instructional Science*, 1-20.

Kearney, S.; Perkins, T.; Kennedy-Clark, S. (2015). 'Using self- and peer-assessments for summative purposes: analysing the relative validity of the AASL (Authentic Assessment for Sustainable Learning) model', *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2015.1039484

KeyCoNet (2014). *Catalogue of initiatives*. Brussels: European Schoolnet. Available at: http://keyconet.eun.org/c/document_library/get_file?uuid=e29c058b-01be-4d08-b77c-85925069d007&groupId=11028.

Kirova, A.; Hennig, K. (2013). 'Culturally Responsive Assessment Practices: examples from an intercultural multilingual early learning program for newcomer children', *Power and Education*, Vol. 5, No. 2, pp. 106-119.

Klapp, A. (2015). 'Does grading affect educational attainment? A longitudinal study', *Assessment in Education: Principles, Policy & Practice, 22:3*, pp. 302-323.

Kokotsaki, D.; Menzies, V.; Wiggins, A. (2016). 'Project-based learning: A review of the literature'. *Improving Schools*, *19*(3), 267–277. https://doi.org/10.1177/1365480216659733

Koloi-Keakitse, S. (2016). 'Assessment training: a precondition for teachers' competencies and use of classroom assessment practices', *International Journal of Training and Development*, Vol. 20, No. 2, pp. 107-123.

Komarkova, I.; Conrads, J.; Collado, A. (2015). Entrepreneurship Competence: An Overview of Existing Concepts, Policies and Initiatives. In-depth case study report. JRC technical reports.

Korthagen, F. A. J. (2010). 'How teacher education can make a difference', *Journal of Education for Teaching*, Vol. 36, No. 4, pp. 407-423.

Kyllonen, P. C. (2012). 'Measurement of 21st Century Competencies within the Common Core State Standards.' Retrieved from http://www.usc.edu/programs/cerpp/docs/Kyllonen_21st_Cent_Skills_and_CCSS.pdf.

Lam, R. (2017). 'Taking stock of portfolio assessment scholarship: From research to practice'. *Assessing Writing*, *31*, 84–97. https://doi.org/10.1016/j.asw.2016.08.003

Lau, A. M. S. (2016). ''Formative good, summative bad?' – A review of the dichotomy in assessment literature'. *Journal of Further and Higher Education*, *40*(4), 509–525. https://doi.org/10.1080/0309877X.2014.984600

Lee, A. M. (2013). 'Assessment OF… AS… FOR…learning'. Retrieved from: https://annemichellemlee88.wordpress.com/2013/02/14/%E2%80%A2assessment-of-as-for-learning/

Lee, I.; Mak, P. (2014). 'Assessment as learning in the language classroom'. In (Ed.), *Assessment as learning*. Hong Kong: Education Bureau.

Lieberman, A.; Pointer Mace, D. H. (2008). 'Teacher Learning: the Key to Educational Reform', *Journal of Teacher Education*, Vol. 59, No. 3, May/June, pp. 226-234.

Livingston, K.; Hutchinson, C. (2017). 'Developing teachers' capacities in assessment through career-long professional learning', *Assessment in Education: Principles, Policy & Practice*, Vol. 24, No. 2, pp. 290-307.

Looney, J.; Michel, A. (2014). *KeyCoNet's conclusions and recommendations for strengthening key competence development in policy and practice*. Final report. KeyCoNet.

Looney, J. (2011a). *Integrating formative and summative assessment: Progress toward a seamless system?* OECD Education Working Papers, No. 58, Paris: OECD Publishing.

Looney, J. (2011b). 'Alignment in complex systems: Achieving balance and coherence', *OECD Education Working Papers*, No. 64, OECD Publishing, Paris, www.oecd.org/edu/workingpapers.

Maier, U.; Wolf, N.; Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education, 95*, pp. 85–98.

Martin, R. (2008). 'New possibilities and challenges for assessment through the use of technology'. In F. Scheuermann & A. G. Pereira (eds.*), Towards a Research Agenda on Computer-Based Assessment*. Luxembourg: Office for Official Publications of the European Communities.

McLaren, S. V. (2012), Assessment is for learning: Supporting feedback, *International Journal of Technology and Design Education*, Vol. 22, No. 2, pp. 227-245.

McMahon, T. (2010). 'Peer feedback in an undergraduate programme: Using action research to overcome students' reluctance to criticise'. *Educational Action Research, 18(2)*, pp. 273-287.

Messick, S. (1989). 'Validity'. In Linn, R.L. (Ed.) *Educational measurement*. New York, NY: Macmillan.

Michel, A.; Pons, X. (2017), 'Mirror, mirror on the wall, tell me I'm right!' (Ed.) *European Journal of Education, 52 (2)*, pp. 124-130.

Miyamoto, K.; Huerta, M. C.; Kubacka, K. (2015). 'Fostering Social and Emotional Skills for Well-Being

and Social Progress', *European Journal of Education*, Vol. 50, No. 2, pp. 147-159.

Montenegro, E.; Jankowski, N. A. (2017). Equity and assessment: Moving towards culturally responsive assessment, (Occasional Paper No. 29). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Morris, A. (2011). *Student standardised testing: Current practices in OECD countries and a literature review*, OECD Education Working Paper No. 65.

Moss, P. A.; Pullin, D.; Gee, J. P.; Hartel, E. H. (2005). 'The Idea of Testing: Psychometric and Sociocultural Perspectives', *Measurement*, Vol. 3, No. 2, pp. 63-83.

Mottier Lopez, L. ; Laveault, D. (2008). 'L'évaluation des apprentissages en contexte scolaire : développements enjeux et controverses', *Mesure et évaluation en éducation*, Vol. 31, No. 3, pp. 5-34.

Mottier Lopez, L. (2014). 'L'évaluation pédagogique va-t-elle enfin marcher sur ses deux pieds ? Les enseignements de l'histoire récente de l'école primaire genevoise. In D. Laveault (Ed.), « Les politiques d'évaluation en éducation. Et après ? » *Éducation et francophonie*, XLII (3), pp. 85-101.

Mottier Lopez, L. (2006). 'Interroger la pratique du portfolio en situation scolaire dans un perspective « située » de l'apprentissage'. *Mesure et évaluation en éducation*, Vol. 29, No. 2, pp. 1-21.

Mottier Lopez, L. (2015). 'L'évaluation formative des apprentissages des élèves : entre innovations, échecs et possibles renouveaux par des recherches participatives'. *Questions vives*, No. 3.

Muskin, J. A. (2015). 'Student Learning Assessment and the Curriculum: issues and implications for policy, design and implementation'. In: Current and Critical Issues in the Curriculum and Learning, Unesco International Bureau of Education.

Newton, P. E. (2007), 'Clarifying the purposes of educational assessment'. *Assessment in Education*, Vol. 14, No. 2, pp. 149-170.

Newton, P. E.; Baird J. (2016). 'The great validity debate', *Assessment in Education: Principles, Policy & Practice, 23:2*, pp. 173-177.

Nicol, D. J.; Macfarlane-Dick, D. (2006). 'Formative assessment and self-regulated learning: a model and seven principles of good feedback practice', *Studies in Higher Education*, 31(2), pp. 198–218.

Nicol, D. (2007). 'E-assessment by design: using multiple-choice tests to good effect'. *Journal of Further and Higher Education 31, pp.* 53–64. doi:10.1080/03098770601167922

Nunan, D. (2010). 'Technology Supports for Second Language Learning', *International Encyclopedia of Education* (3rd ed.,Vol. 8,pp. 204–209) (Oxford, Elsevier).

OECD (2005). *The Definition and Selection of Key Competences. Executive Summary*.

OECD (2012). *Equity and Quality in Education: Supporting Disadvantaged Students and Schools*, Paris: OECD Publishing.

OECD (2013). *Synergies for Better Learning. An International Perspective on Evaluation and Assessment*, OECD Publishing, Paris.

OECD (2014). *TALIS 2013 Results: An International Perspective on Teaching and Learning*, TALIS, Paris: OECD Publishing.

OECD (2015a). *The Missing Entrepreneurs 2015: Policies for Self-employment and Entrepreneurship*, OECD Publishing, Paris.

OECD (2015b). *Skills for Social Progress: The Power of Social and Emotional Skills*, OECD Skills Studies, OECD Publishing.

OECD (2016). *Innovating Education and Educating for Innovation: The Power of Digital Technologies and Skills*, OECD Publishing, Paris.

Olovsson, T. G. (2015). 'Changes In The Assessment Process In Swedish Compulsory School Classrooms'. *Procedia - Social and Behavioral Sciences 191,* pp. 424 – 431.

Ontario report (2010). *Growing success: Assessment, evaluation and reporting in Ontario schools*. 1st Edition, Covering Grades 1 to 12. http://www.edu.gov.on.ca/eng/policyfunding/growSuccess.pdf

Panadero, E.; Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review 9*, pp. 129–144.

Pellegrino J., W. (2016). 'Introduction to Special Section of Educational Psychologist on Educational Assessment: Validity Arguments and Evidence—Blending Cognitive, Instructional, and Measurement Models and Methods', *Educational Psychologist, 51:1*, pp. 57-58.

Pellegrino, J. W.; DiBello, L. V.; Goldman, S. R. (2016). 'A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments'. *Educational Psychologist*, *51*(1), 59–81. https://doi.org/10.1080/00461520.2016.1145550

Pellegrino, J.; Hilton, M. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington D.C.: National Academy of Sciences.

Pepper, D. (2007). *Assessment for Disabled Students: An International Comparison*. London: QCA.

Pepper, D. (2011). 'Assessing Key Competences across the Curriculum – and Europe', *European Journal of Education*, 46, 3.

Pepper, D. (2013). *KeyCoNet 2013 Literature Review: Assessment for key competences*. KeyCoNet 2013.

Prøitz, T. S. (2015). 'Learning Outcomes as a Key Concept in Policy Documents throughout Policy Changes', *Scandinavian Journal of Educational Research*, Vol. 59, No. 3, pp. 275-296.

Putwain, D. W.; Connors, L.; Woods, K.; Nicholson, L. J. (2012). 'Stress and anxiety surrounding forthcoming Standard Assessment Tests in English schoolchildren', *Pastoral Care in Education*, Vol. 30, No. 4, pp. 289-302.

Ramdass, D.; Zimmerman, B. J. (2008). 'Effects of Self-Correction Strategy Training on Middle School Students' Self-Efficacy, Self-Evaluation, and Mathematics Division Learning', *Journal of Advanced Academics*, 20, 18–41.

Ramirez-Corona, N.; Ramirez Apud, Z.; Lopez-Malo, A.; Palou, E. (2013). 'Assessing Metacognitive Awareness during Problem-Solving in a Kinetics and Homogeneous Reactor Design Course'. *Proceedings of the 2013 American Society for Engineering Education Annual Meeting*. Atlanta, GA, USA.

Rasmussen, J. (2013). 'Competence goal-driven education in school and teacher education', International Conference on Learning and Teaching 2013: Transforming Learning and Teaching to Meet the Challenges of 21st Century Education.

Raveaud, M. (2004). 'From assessment practices to conceptions of equity: France and England compared', ECER.

Ravotto, P. (2011). 'Competence-based learning in europe & the sloop2desc model'. In G. Fulantelli, L. Oprea. *Preparing the teachers for a competence-based education system*.

Redecker, C. (2013). *The Use of ICT for the Assessment of Key Competences*, European Commission, Joint Research Centre, Institute for Prospective Technological Studies, JRC Scientific and Policy Reports. Luxembourg: Publications Office of the European Union.

Redecker, C.; Johannessen, Ø. (2013). 'Changing Assessment — Towards a New Assessment Paradigm Using ICT'. *European Journal of Education, Vol. 48*, No. 1.

Roschelle, J.; Feng, M.; Murphy, R. F.; Mason, C. A. (2016). 'Online Mathematics Homework Increases Student Achievement'. *AERA Open, 2(4)*, 2332858416673968. https://doi.org/10.1177/2332858416673968

Roscoe, R. D.; Chi, M. T. H. (2007). 'Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions'. *Review of Educational Research, 77*, pp. 534-574.

Salomon, G.; Globerson, T. (1989). 'When teams do not function the way they ought to'*. International Journal of Educational Research, 13*, pp. 89–99.

Sansone, N.; Ligorio, M. B.; Buglass, S. L. (2016). 'Peer e-tutoring: Effects on students' participation and interaction style in online courses'. *Innovations in Education and Teaching International, 1–10*. https://doi.org/10.1080/14703297.2016.1190296

Santos, P.; Cook, J.; Hernández-Leo, D. (2015). M-AssIST: Interaction and Scaffolding Matters in Authentic Assessment. *Educational Technology & Society*, *18*(2), 33–45.

Sargent, C. (2014). *Teacher Guide: Assessment of Key Competences in Education*, KeyCoNet.

Schneider, C.; Bodensohn, R. (2017). 'Student teachers' appraisal of the importance of assessment in teacher education and self-reports on the development of assessment competence', *Assessment in Education: Principles, Policy & Practice*, Vol. 24, No. 2, pp. 127-146.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: multiple choice question examination versus assignment essay, *Higher Education, 35*, pp. 453–472.

Shepard, L. A. (2005). Linking Formative Assessment to Scaffolding. *Educational Leadership, 63 (3)*, pp. 66–70.

Shepherd, C.; Hannafin, M. J. (2013). 'Reframing Portfolio Evidence Empowering Teachers through Single-Case Frameworks', *Journal of Thought*, Vol. 48, No. 1, pp. 33-51.

Shute, V. J. (2008). Focus on formative feedback*. Review of Educational Research, 78*, pp. 153-189.

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

Shute, V. J.; Wang, L.; Greiff, S.; Zhao, W.; Moore, G. (2016). 'Measuring problem solving skills via stealth assessment in an engaging video game'. *Computers in Human Behavior, 63*, pp. 106–117.

Siddiq, F.; Hatlevik, O. E.; Olsen, R. V.; Throndsen, I.; Scherer, R. (2016). 'Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy'. *Educational Research Review*, *19*, 58–84. https://doi.org/10.1016/j.edurev.2016.05.002

Silfver, E.; Sjoberg, G.; Bagger, A. (2016), 'An 'appropriate' test taker: the everyday classroom during the national testing period in school year three in Sweden', *Ethnography and Education*, Vol. 11, No. 3, pp. 237-252.

Sitthiworachart, J.; Joy, M. (2008). 'Computer support of effective peer assessment in an undergraduate programming class'. *Journal of Computer Assisted Learning, 24(3)*, pp. 217-231.

Sjögren, A. (2009). 'The long run consequences of being graded in elementary school', paper presented at the Final Conference of the RTN Network "Economic of Education and Education Policy in Europe" (EEEPE), hosted by the Centre for Economic Performance, LSE, London.

Smith, K. (2011). 'Professional development of teachers – A prerequisite for AfL to be successfully implemented in the classroom', *Studies in Educational Evaluation*, Vol. 37, pp. 55-61.

Soland, J.; Hamilton, L. S.; Stecher, B. M. (2013). *Measuring 21st century competencies: guidance for educators*. RAND Corporation, November 2013.

Søndergaard, H. (2009). Learning from and with peers: The different roles of student peer reviewing, Paris.

Spector, J. M.; Ifenthaler, D.; Sampson, D.; Yang, L. J., Mukama, E.; Warusavitarana, A.; Dona, K. L. et al. (2016). 'Technology Enhanced Formative Assessment for 21st Century Learning'. *Journal of Educational Technology & Society*, *19*(3), 58–71.

Stenlund, T.; Eklöf, H.; Lyrén, P. E. (2017). 'Group differences in test-taking behaviour: An example from a high-stakes testing program'. *Assessment in Education: Principles, Policy & Practice, 24, pp. 4–20*. doi:10.1080/0 969594X.2016.1142935

Stiggins, R. (2005). 'Formative Assessment to Assessment for Learning : A Path to Success in Standards-Based Schools', *TH Phi Delta Kappan*, Vol. 87, No. 4, pp. 324-328.

Stobart, G. (2008). Attitudes and assessment. [Editorial]. *Assessment in Education: Principles, Policy & Practice, 15(1),* pp. 1 - 2.

Strijbos, J. W.; Sluijsmans, D. (2010). 'Unravelling peer assessment: Methodological, functional, and conceptual developments'. *Learning and Instruction, 20(4),* pp. 265-269.

Sturgis, C. (2014). 'Progress and Proficiency: Redesigning Grading for Competency Education', *International Association for K-12 Online Learning*.

Takayama, K. (2013). 'OECD, 'Key competencies' and the new challenges of educational inequality'. *Journal of Curriculum Studies*, *45*(1), 67–80. https://doi.org/10.1080/00220272.2012.755711

Tawil, S.; Cougoureux, M. (2013). *Revisiting Learning: The Treasure Within Assessing the influence of the 1996 Delors Report*, UNESCO Education Research and Foresight.

Tchibozo, G. (2011). 'Emergence and outlook of competence-based education in European education systems: an overview', *Education, Knowledge and Economy, 4:3*, pp. 193-205.

Terrail, J.-P. (2016). *Pour une école de l'exigence intellectuelle. Changer de paradigme pédagogique*. Paris: La Dispute.

Thorsen, C. (2014). 'Dimensions of norm-referenced compulsory school grades and their relative importance for the prediction of upper secondary school grades'. *Scandinavian Journal of Educational Research, 58*, 127–146.

Thorsen, C.; Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation, 18*, 153–172.

Tiana, A.; Moya, J.; Luengo, F. (2011). 'Implementing Key Competences in Basic Education: Reflections on Curriculum Design and Development in Spain'. *European Journal of Education 46 (3)*, pp. 307–322.

Topping, K. J. (2009). 'Peer assessment'. *Theory into Practice, 48(1),* pp. 20-27.

Travers, N. N. (2012). 'What is next after 40 years? Part 1: Prior learning assessment: 1970–2011', *Journal of Continuing Higher Education*, *60*, pp. 43–47.

Turner, S. L. (2014). 'Creating an Assessment-Centred Classroom: Five Essential Assessment Strategies to Support Middle Grade Student Learning and Achievement', *Middle School Journal*, Vol. 45, No. 5, pp. 3-16.

UNESCO (1996). The four pillars of education described in Chapter 4 of Learning: The treasure within: A report to UNESCO of the International Commission on Education for the twenty-first century.

UNESCO (2015). *2013 Asia-Pacific Education Research Institutes Network (ERI-Net) Regional Study on transversal competencies in education policy & practice*. Available at: http://unesdoc.unesco.org/images/0023/002319/231907E.pdf.

Van der Klei, F. M.; Vermeulen, J. A.; Schildkamp, K.; Eggen, T. J. H. M. (2015); 'Integrating data-based decision-making, Assessment for Learning and diagnostic testing in formative assessment', *Assessment in Education: Principles, Policy and Practice*, Vol. 22, No. 3, pp. 324-343.

VanTassel-Baska, J. (2014). 'Performance-Based Assessment. The Road to Authentic Learning for the Gifted', *Gifted Child Today*, Vol. 37, No. 1, pp. 41-47.

Volkwein, J.F. (2003). 'Implementing outcomes assessment on your campus', *Research and Planning EJournal*, Vol. 1, No.1.

Voogt, J.; Pareja Roblin, N. (2012). 'A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies', *Journal of Curriculum Studies, 44:3*, pp. 299-321.

Wang, H. C.; Chang, C. Y.; Li, T. Y. (2008). 'Assessing creative problem-solving with automated text grading'. *Computers and Education, 51(4),* pp. 1450-1466.

Wang, A. I; Zhu, M.; Sætre, R. (2016). 'The Effect of Digitizing and Gamifying Quizzing in Classrooms'. *The 10th European Conference on Games Based Learning.*

Weissberg, R. P.; Durlak, J. A.; Domitrovich, C. E.; Gullotta, T. P. (2015). 'Social and emotional learning: Past, present, and future'. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, and T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York, NY: Guilford.

Wiliam, D. (2006). 'Formative Assessment: Getting the Focus Right', *Educational Assessment*, Vol. 11, No. 3 and 4, pp. 283-289.

Wiliam, D. (2007). 'Content *Then* Process: Teaching Learning Communities in the Service of Formative Assessment', in Reeves, D. (ed.), *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning*. Bloomington (U.S.): Solution Tree Press.

Wiliam, D.; Leahy, S. (2014). 'Sustaining Formative Assessment with Teacher Learning Communities'. Learning Sciences, Dylan Wiliam Center. Available at: http://www.dylanwiliamcenter.com/files/pdf/Sustaining-TLCs-20140829.pdf?aliId=89372614.

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

Williams, M.; Moser, T.; Youngblood, J. II; Singer, M. (2015). 'Competency-Based Learning: Proof of professionalism', *Academy of Business Journal*, 17, pp. 150–161.

Wilson, M.; Scalise, K.; Gochyyev, P. (2015). 'Rethinking ICT literacy: From computer skills to social network settings', *Thinking Skills and Creativity*, Vol. 18, pp. 65-80.

Wouters, P.; van Oostendorp, H. (2013). 'A Meta-analytic review of the role of instructional support in game-based learning'. *Computers & Education, 60*, 412–425.

Wu, H.-M.; Kuo, B.-C.; Wang, S.-C. (2017). 'Computerized Dynamic Adaptive Tests with Immediately Individualized Feedback for Primary School Mathematics Learning'. *Journal of Educational Technology & Society*, *20*(1), 61–72.

Yastibas, A. E.; Yastibas, G. C. (2015). 'The Use of E-portfolio-based Assessment to Develop Students' Self-regulated Learning in English Language Teaching'. *Procedia - Social and Behavioral Sciences*, *176*, 3–13. https://doi.org/10.1016/j.sbspro.2015.01.437

HOW TO OBTAIN EU PUBLICATIONS

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu).

NC-01-16-892-EN-N

ISBN: 978-92-79-70077-4