

1

21st Century competencies: Challenges in education and assessment

By Natalie Foster

(OECD)

This chapter reviews several frameworks of so-called “21st Century competencies”, reviewing their main ideas and the vision of education they seek to promote. It also elaborates on the interrelations between 21st Century competencies – both among each other and to disciplinary learning – before discussing some of the key challenges that this vision of education presents for contemporary education systems. In particular, this chapter discusses challenges in the context of their assessment, including defining assessment constructs and learning progressions, generalisability of assessment claims, task and item design, interpreting scoring and evidence, reporting and validation.

Introduction

What is worth knowing, doing and being has been subject to a global conversation since before the turn of the 21st Century. Success in global contemporary society demands a wider set of competencies that go beyond the traditional literacies of reading, mathematics and science. Information and communication technologies (ICTs) have radically transformed our societies connecting people around the world and delivering unprecedented amounts of information to us, in turn giving rise to new forms of decentralised and autonomous learning. Young people today must not only learn to participate in a more interconnected, digital and rapidly changing world, they must also learn to develop their own agency and make decisions that contribute to individual and collective well-being. They need to understand and appreciate different perspectives, interact and collaborate successfully with others, and take responsible action towards creating a cohesive and sustainable future for all.

The nature of work is also changing. A declining proportion of the labour market in OECD economies is engaged in jobs consisting of routine work and manual labour (OECD, 2016^[1]). The past two decades have seen a significant shift towards economies and societies that increasingly rely on human knowledge to produce new goods and services, underpinned by skills such as creative thinking, innovation and complex problem solving.

These significant trends have important consequences for schooling, teaching and learning. The knowledge and skills that today's students need to thrive in rapidly changing labour markets and to live with others as responsible, democratically- and socially-engaged citizens are changing. Value has shifted away from memorising content towards developing interdisciplinary skills (so-called "21st Century skills" or competencies) and acquiring deeper learning outcomes or transferable knowledge.

Over the past 20 years a growing body of research has examined this global narrative, producing a variety of international frameworks that describe the knowledge, skills and attitudes that young people need for active and effective participation in the emerging global knowledge society (Pellegrino and Hilton, 2012^[2]; Fadel and Groff, 2018^[3]; Binkley et al., 2011^[4]; Scott, 2015^[5]; World Economic Forum, 2015^[6]; European Commission, 2019^[7]). The OECD's Learning Framework 2030 goes one step further, emphasising the need to cultivate students' agency as a key goal of a 21st Century education so that young people are able to fulfil their potential and actively contribute to the well-being of their communities and the planet (OECD, 2018^[8]). A number of works have also comparatively analysed such frameworks (Voogt and Roblin, 2012^[9]; Scott, 2015^[5]; Chalkiadaki, 2018^[10]; Joynes, Rossignoli and Amonoo-Kuofi, 2019^[11]).

While the literature reveals a general consensus on what 21st Century competencies are and why they are important, there remain significant challenges to the adoption of this agenda in practice. First, there is no singularly agreed-upon approach for identifying the competencies that should be prioritised in formal education nor how specific competencies are defined or delimited in relation to others. Second, shifting a greater focus on the development of 21st Century competencies throughout formal education requires accompanying shifts across curricula, pedagogy and assessment, as well as ensuring that these systems closely align. Yet several open questions remain about how best to learn, teach and assess 21st Century competencies in the classroom.

This chapter begins by examining frameworks of 21st Century competencies. It discusses the twin issues of a lack of theory and definition and the need for education system alignment in some detail, with a particular focus on their implications for educational assessment. It then identifies six interconnected assessment challenges for developing assessments of 21st Century competencies: 1) construct and learning progression definition; 2) generalisability; 3) task and item design; 4) interpreting and scoring evidence; 5) reporting; and 6) validation. The chapter concludes by summarising the implications of these interconnected assessment challenges in terms of the requirements they impose on next-generation assessments of 21st Century competencies.

What are 21st Century competencies?

Before going further, it is useful to establish what exactly the term “21st Century competencies” means. There is a diversity of terminologies employed interchangeably within this relatively crowded space: “21st Century skills/competencies”, “soft skills”, “interdisciplinary skills” and “transferable skills”, to name just a few. This terminological ambiguity also extends to the ways in which different frameworks identify and define specific competencies (e.g. ICT literacy vs. digital literacy vs. media literacy). For the sake of clarity, this chapter uses the term 21st Century competencies to refer to the broad vision of education set forth by the frameworks cited above and to the various competencies that they describe. They are generally understood to refer to the knowledge, skills and attitudes necessary to be successful for living and working in the 21st Century global knowledge economy, to participate appropriately in an increasingly diverse society, to use new technologies effectively, and to adapt to change and uncertainty.

Although frameworks vary, they tend to describe 21st Century competencies as being:

- transversal (i.e. relevant or applicable in many fields);
- multidimensional (i.e. encompassing knowledge, skills and attitudes); and
- associated with higher-order skills and behaviours that represent the ability to transfer knowledge, cope with complex problems and adapt to unpredictable situations (Voogt and Roblin, 2012^[9]).

Beyond general convergence around these core characteristics, frameworks identify, organise and classify 21st Century competencies in different ways. Some group competencies based on their conceptual features, for example cognitive, interpersonal and intrapersonal competencies (Pellegrino and Hilton, 2012^[2]). Others group competencies according to their purpose or context of use, for example ways of thinking, ways of living in the world, ways of working and tools for working (Binkley et al., 2011^[4]). Abstracting from the specificities of each framework, some broadly distinct categories of competencies do consistently emerge (Figure 1.1). While these six broad categories capture the essence and exhaustive lists of competencies identified across different frameworks, note that not all frameworks include each category, nor do they always assign specific competencies to the same broader categories.

Figure 1.1. Broad categories of 21st Century competencies



Identifying common categories of competencies provides some useful insight into the broader goals of education that these frameworks seek to promote, but they nonetheless remain strongly interlinked in the sense that engaging one “type” of competence often requires engaging other “types” simultaneously. For example, problem solving (usually categorised as a cognitive competence) also requires individuals to

monitor their progress and adapt accordingly (i.e. metacognitive competencies) and likely some degree of persistence (i.e. intrapersonal competence) in order to reach a successful solution.

Regardless of the way in which specific competencies are categorised across the 9 frameworks reviewed here, critical thinking, creative thinking, communication and ICT-related competencies are consistently identified as those that young people need to develop. This also largely reflects the 21st Century competencies that are most commonly cited within national curricula documentation (Care, Anderson and Kim, 2016^[12]). All frameworks identify the importance of civics and citizenship, although some regard these as a cross-cutting knowledge area (along with others like financial, health, environmental and global literacies) rather than a specific type of competence. Most frameworks also identify problem solving, collaboration, metacognition and self-regulated learning, as well as some intrapersonal competencies.

The intersection of digital competencies with other 21st Century competencies is also addressed within each framework. This is because the proliferation of ICTs in both personal and professional life forms one of the central arguments for developing 21st Century competencies, primarily through their functionalities that enhance our capacity for communication, collaboration, use of knowledge and information-finding. However, frameworks differ in how they conceptualise digital competencies: some consider them to be their own distinct category of competencies (Partnership for 21st Century Learning, 2019^[13]; Binkley et al., 2011^[4]), whereas others employ more integrative approaches where the development and use of ICT knowledge and skills is embedded within other 21st Century competencies such as critical thinking, problem solving, communication or collaboration (Voogt and Roblin, 2012^[9]).

A multi-faceted challenge for education systems

Nearly one quarter of the way through the 21st Century, the idea of “21st Century competencies” is no longer particularly new. So why has there not been more progress towards achieving this vision of education? In their analysis of national education documents in 102 countries, Care, Anderson and Kim (2016^[12]) found that although a majority of countries acknowledge 21st Century competencies within their broader educational vision statements there is wide variation across countries in terms of the specific skills or competencies they reference. One major obstacle is the lack of a clear and universal definition of 21st Century competencies that moves beyond general rhetoric – one that specifies both what is included within this umbrella term and how those competencies relate (or not) to one another.

There is also clearly a lack of shared understanding of how these competencies develop and how they can be taught. Even if most countries acknowledge 21st Century competencies in some way, relatively fewer explicitly integrate them within their curricula or provide clear developmental progressions for them (Care, Anderson and Kim, 2016^[12]). This in turn provides little practical guidance to educators in terms of designing and implementing educational approaches to teaching, learning and assessing these competencies. A second major challenge is therefore how to actually implement a 21st Century competencies agenda in practice. Real change towards adopting this agenda has implications across all aspects of education systems – from curriculum to pedagogy to assessment – and requires all three to be well-aligned.

Challenges in curriculum and pedagogy

While a more in-depth discussion about curriculum and pedagogy is beyond the scope of this chapter, they are nonetheless central components of any 21st Century competencies agenda and are intricately connected to how these competencies can and should be assessed. One ongoing debate in the field concerns disciplinary versus interdisciplinary approaches to teaching and learning 21st Century competencies. While these competencies are widely accepted as being interdisciplinary, what it means to problem solve, think critically or be creative in one context may be very different in another context. In other

words, being able to successfully engage 21st Century competencies when embedded in different domains depends, at least to some extent, on a foundation of relevant knowledge in that domain. Pellegrino and Hilton (2012, p. 4_[2]) suggest that 21st Century competencies represent “transferrable knowledge” that comprises both “content knowledge in a domain and also procedural knowledge of how, why, and when to apply this knowledge”.

In a review of the literature on four different 21st Century competencies, Lai and Viering (2012_[14]) concluded that domain-specific knowledge is an important requisite – although to different extents depending on the specific competence. Critical thinking, for example, requires a foundation of domain-specific knowledge (with some even refuting the existence of domain-general critical thinking processes), while experts tend to agree that creativity has both domain-specific and domain-general components (Lai and Viering, 2012_[14]). Yet even for competencies with some domain-general components, the context in which they are engaged can influence the degree to which those components are relevant. For example, both convergent and divergent thinking processes are important for creative thinking across domains, but convergent thinking might be relatively more important in scientific or engineering domains than in artistic domains.

This has implications for both curriculum and pedagogy. First, curriculum designers need to know what foundational knowledge facilitates the development and application of 21st Century competencies within different domains. Curricula need to integrate 21st Century competencies in the context of particular content knowledge and to treat both as equally important. Second, and related, more effort needs to be directed towards defining developmental progressions so that students are taught the right foundational knowledge to support the integration of 21st Century competencies at the right time – and so teachers can be informed about what to reasonably expect from students at different stages of education. When students first encounter new ideas or concepts, their understanding is shallow and often bound to specific examples. Students need to develop and organise their conceptual knowledge and understanding sufficiently to facilitate its application to novel situations by engaging 21st Century competencies. The teaching strategies that allow students to do this integrate carefully-designed direct instruction with hands-on inquiries that actively engage students in using material they have learnt with higher-order competencies of increasing complexity (Darling-Hammond et al., 2019_[15]). In an iterative process, learning that engages higher-order competencies allows knowledge to be understood deeply enough to be recalled and used for other purposes in novel situations (Learning Policy Institute and Turnaround for Children, 2021_[16]).

All of this means that opportunities to engage 21st Century competencies must be integrated systemically and strategically throughout the curriculum. One way to do this is through student-centred learning methods, such as problem-based or project-based learning, that empower students to work on real-world problems in authentic and genuinely meaningful ways. These participatory approaches to learning enable students to research and evaluate information using different resources and to actively construct their own knowledge and skills instead of passively absorbing and memorising information. Real-world problems are also rarely confined to a single content area, which make them ideal contexts for engaging interdisciplinary 21st Century competencies. Problem-based and project-based learning approaches encourage students to make connections between content areas and engage competencies like critical and creative thinking, problem solving and collaboration (Paniagua and Istance, 2018_[17]).

Despite the fact that student-centred pedagogical approaches are widely acknowledged as beneficial, they impose additional demands on educators. They require that teachers invest time in devising engaging, student-centred lesson plans connected to the curriculum and manage a more interactive classroom in which students collaborate with each other and engage in autonomous research. At the system level, this requires a greater investment in teacher development and training, as well as ensuring that educators are given the autonomy to integrate the curriculum in ways that make sense for their classrooms and that they feel empowered to do so.

Challenges in assessment

There is little point in investing heavily in curriculum and educator training reform without investing in assessment to evaluate what is (or is not) being accomplished in the classroom. Curricula, pedagogy and assessment are intricately connected and must be aligned in well-functioning education systems. Assessments – especially large-scale – are important signposts indicating what students should learn and what they can do. Shifts in curricula and pedagogy can thus be driven by changes in an education system's assessment focus and by the educational gaps that they reveal, in turn informing policymaking and reform. Moreover, explicitly focusing assessment on these competencies requires that they are clearly defined and requires being specific about what exactly they involve at different educational levels, therefore contributing to establishing a shared understanding of these competencies and how they should be taught.

However – despite some promising examples at scale (e.g. the Programme for International Student Assessment) – there is currently a lack of systemic understanding in how to measure or capture the attainment of 21st Century competencies (Care et al., 2018^[18]; Vista, Kim and Care, 2018^[19]). This is because several of the challenges related to defining or integrating 21st Century competencies in curriculum and pedagogy have similarly complex implications for assessment. Six major interconnected assessment challenges, described below in Table 1.1, are discussed in detail in the remainder of this chapter. These assessment challenges are: 1) defining constructs and learning progressions; 2) generalisability; 3) task and item design; 4) interpreting and scoring evidence; 5) reporting; and 6) validation.

Table 1.1. Challenges for the assessment of 21st Century competencies

Challenge	Source of complexity	Implication(s) for assessment
Defining constructs and learning progressions	21st Century competencies are complex and multidimensional, involving cognitive, metacognitive and affective processes. They are also strongly interconnected when engaged authentically. Robust learning progressions are also generally lacking.	<ul style="list-style-type: none"> The constituent variables need to be clearly defined at the beginning of any assessment design process to inform task design, the evidence to be collected, and the claims to be made about student performance. It can be difficult to isolate and interpret evidence for discrete constructs. It can be difficult to identify processes and outcomes (i.e. evidence) linked to different levels of mastery.
Generalisability	21st Century competencies require some degree of domain-specific knowledge to be meaningfully engaged, and they may also be understood and defined differently in different domains.	<ul style="list-style-type: none"> Domain-general assessments lack validity. The greater the domain-specificity of an assessment, the weaker the generalisability of its claims about students' capacity to engage 21st Century competencies outside of that specific domain context.
Task and item design	Students need to work on more open-ended, interactive and authentic tasks in order to engage 21st Century competencies and demonstrate their proficiency.	<ul style="list-style-type: none"> Simple item types cannot fully reflect the range of authentic scenarios that engage complex constructs nor capture the range of evidence of proficiency (e.g. interpersonal competencies are evidenced primarily through behaviours when interacting with others). Multiple instruments or item types are generally required to gather information about all relevant aspects of the construct.
Interpreting and scoring evidence	Students need to work on more open-ended, interactive and authentic tasks in order to engage 21st Century competencies. Processes and behaviours are also key aspects of performance.	<ul style="list-style-type: none"> Evidence generated by process data can be challenging to interpret and use for scoring. Open-ended items without a pre-determined list of "correct" responses may require human scoring (with implications in terms of time and cost).
Reporting	21st Century competencies are complex and multidimensional, involving cognitive, metacognitive and affective processes.	<ul style="list-style-type: none"> Unidimensional score scales are inadequate, but multiple scales (or scales with clearly distinct dimensions) are difficult to achieve given the constraints of large-scale assessments (e.g. limited testing time). Some constructs might not be best described by linear point scales, but the exploration of alternative reporting methods has been scarce so far.
Validation	21st Century competencies are complex and multidimensional, involving cognitive, metacognitive and affective processes. Processes and behaviours are also often associated with performance.	<ul style="list-style-type: none"> How constructs are defined and how student performance is interpreted are more susceptible to cultural bias than traditional tests of knowledge. More open, interactive and process-oriented task and scoring models impose additional burdens in terms of establishing a validity argument.

Defining constructs and learning progressions

21st Century competencies are complex and multidimensional constructs. They involve a combination of cognitive, metacognitive and affective processes, and are supported by a set of knowledge, skills and attitudes. These multiple components are often strongly inter-related with other 21st Century competencies and most authentic, real-world contexts require individuals to engage several competencies simultaneously – making it difficult to clearly distinguish mastery in one competency from another. For example, problem solving involves aspects of metacognition, self-regulated learning and persistence – and depending on the context and typology of the problem, it could also involve elements of creative thinking and collaboration, as well as domain-specific knowledge (see *Generalisability* section below).

For any assessment to fully represent and measure its target construct, all of the constituent elements of the construct must be clearly defined and captured reliably through the assessment instrument(s). As set forth in the Introduction, a sound theoretical framework must describe the kinds of evidence that need to be generated and collected to sustain claims about students' performance and to design appropriate tasks and define proficiency scales that reflect the different levels of competence mastery (Wilson et al., 2011^[20]; Ercikan and Oliveri, 2016^[21]). However, the complexity of these constructs and the conceptual crowding within the broader discourse on 21st Century competencies means it is difficult to break down constructs into discrete and independently measurable components, as well as isolate and attribute evidence generated by students to one particular competence or another (Ercikan and Oliveri, 2016^[21]).

This definition issue is challenging not only in terms of identifying exactly what to measure but also in terms of how to interpret performance. A lack of well-defined learning progressions mean it is harder to identify the processes and outcomes that students demonstrate at different levels of competence mastery, as well as to design and locate assessment tasks that sample students' competency at various levels of complexity or sophistication. This is particularly relevant in the context of assessing 21st Century competencies since significant evidence is also expressed through students' behaviours and processes not only their final output (Care et al., 2018^[18]).

Generalisability

While 21st Century competencies can be applied across subject domains, they are all bound by some extent to the context in which they are applied. This tension between domain-specificity and domain-generalisability has important implications for assessment. First, assessment designers need to be clear about whether and how the target construct changes across domains – for example, whether one aspect of the construct is relatively more important in some domains of application compared to others – and they need to design items that can elicit relevant kinds of evidence accordingly. This of course relies on a strong foundation of theory about the nature of the construct both within and across domains.

Second, the role and importance of domain-specific knowledge in the assessment needs to be considered, with important trade-offs in terms of developing authentic tasks and making generalisable claims about student performance. One approach is to limit the relevance of domain knowledge to the target construct by situating tasks in neutral and accessible contexts or by focusing measurement on domain-general processes. However, this limits the authenticity of the tasks and assessment claims as 21st Century competencies are rarely exercised in reality in contexts where no relevant knowledge is beneficial. Providing the knowledge that test takers need directly within the task prompt could be one way to mediate this limitation but test takers with existing, well-organised knowledge schemas or knowledge of domain-relevant strategies might nonetheless have an advantage.

A different approach is to acknowledge that it is neither possible nor desirable to disentangle 21st Century competencies from domain-specific knowledge and instead integrate their measurement within domain-specific assessments (see Chapter 4 of this report for an example of this approach) – although this has consequences on the generalisability of inferences made about student performance beyond the given

assessment domain (Ercikan and Oliveri, 2016^[21]). One way to mediate the limitations of this approach is to employ a sampling design: in other words, recognise that knowledge is a relevant component of 21st Century constructs and develop assessment tasks across several domain contexts so that the assessment can provide a more comprehensive view of students' strengths and weaknesses across domains (Lai and Viering, 2012^[14]). This approach clearly has practical implications in terms of developing sufficient items across contexts and gathering enough information from students, both within and across different domains, to be able to draw valid and reliable conclusions about student performance.

Item characteristics and task design

Most large-scale assessments rely on traditional item formats, such as multiple choice, true/false statements or close-ended responses, to elicit indicators about students' underlying abilities. While static and close-ended items are easy to code and score, they inherently limit what can be captured about students' performance and essentially target the reproduction of content knowledge. For constructs like mathematics knowledge, the link between test indicators and construct is fairly direct: a correct response demonstrates knowledge of the topic. But these items are not optimal for generating indicators that capture the complexity and multi-component nature of 21st Century competencies, especially as these constructs are defined (at least in part) by behaviours or processes (Lai and Viering, 2012^[14]; Care et al., 2018^[18]; Vista, Kim and Care, 2018^[19]).

Assessments of 21st Century competencies need to generate evidence that indicates not just what students know but also how they deal with complex situations and iterate towards a solution. For example, "good" self-regulated learning or collaboration is characterised as much by behaviours, attitudes and ways of thinking as it is by eventual (successful) outcomes. The challenge for assessment is that simple indicators of knowledge or outcomes do not capture well these underlying processes. As such, test items for measuring 21st Century competencies need to focus on making students' behaviours and thought processes visible (Lai and Viering, 2012^[14]).

This key issue requires innovation in task and item design in two interconnected ways. First, assessment tasks must mirror the kinds of authentic situations that require 21st Century competencies – not only to stimulate the processes and behaviours from which to generate indicators of the construct (Care, Anderson and Kim, 2016^[12]) but also to ensure that claims about student abilities actually reflect performance in real-life contexts (Care et al., 2018^[18]; Lai and Viering, 2012^[14]; Ercikan and Oliveri, 2016^[21]). The literature generally agrees that 21st Century competencies empower individuals to address new and complex problems and adapt to unpredictable situations, meaning authentic tasks are best situated within the context of open-ended, ill-structured problems (see also Chapter 2 of this report). What's more, some types of 21st Century competencies require situations that involve others with whom to interact (e.g. collaboration, communication) or that trigger some kind of personal state, emotion or level of investment (e.g. persistence, conflict resolution, self-regulated learning). These types of authentic problem situations are clearly more difficult to generate within the constraints of a controlled test environment and in a way that engages all students to the same extent (i.e. that elicits evidence from all students).

Second, and closely related, test items need to be open-ended and interactive so that they can make visible test takers' behaviours and thinking processes. For many 21st Century competencies, this means providing students with tools for doing and making and a test environment that enables them to engage in the entire process of idea conception to implementation by providing them with choices and opportunities to explore and iterate upon their ideas. These kinds of affordances cannot be sufficiently provided by the static, closed-response item types typically used in large-scale assessment. While technology provides new opportunities to address these needs (see Chapters 5 and 7 of this report), designing and validating technology-enhanced items also demands more time and financial resources.

Interpreting and scoring evidence

The scoring of items in any assessment is closely related to the task and item design, the claims that the assessment aims to make, and the definition of the construct and its learning progressions (Csapó et al., 2011^[22]). Assessment tasks must elicit relevant evidence from students, but this evidence needs to be interpreted, clearly connected to the construct and levels of performance (as defined by construct maps or learning progressions) and accumulated using some kind of statistical model. Traditional assessments focusing on knowledge reproduction are easy to code and score: if students select or write the correct response then they receive credit. However, this simple scoring model can rarely be applied in the context of measuring 21st Century competencies that are largely defined by thought processes and behaviours, and for which it is not possible to define a concise, finite list of correct responses.

In recent decades, advances in computer-based assessment mean that evidence of student behaviours and thought processes can be captured through process (or log-file) data. But the interpretation of this evidence is far from straightforward: similar patterns of behaviour may mask real differences in thinking processes and approaches. For example, a prolonged period of recorded inactivity could be an indicator of disengagement or of a student who is deep in thought. Such behaviours may also be more susceptible to cultural differences, emphasising the importance of validation activities in computer-based assessments that make use of process data (see *Validation* section below, and also Chapter 12 of this report). Even when the interpretation of the behaviour is clear, it can be challenging to establish a scoring hierarchy among different behaviours and strategies as the optimal choice may depend on a host of other factors including the test taker's level of prior knowledge, motivation and even personality traits. For example, Roll et al. (2014^[23]) demonstrated that while productive help-seeking was an optimal self-regulated learning strategy for most learners in an online problem-solving environment, trial-and-error was actually the most beneficial strategy for those with the lowest levels of prior knowledge.

New and more advanced analytical models are required to use evidence derived from process data, possibly in combination with more traditional types of outcome data. Authentic assessments of 21st Century competencies require test takers to engage in extended, performance-based tasks using interactive tools and resources – but these affordances introduce complexity in scoring and data analysis by giving test takers choice (meaning they may not experience the test in the same way, thus threatening comparability) and by introducing dependency across items (whereby students' prior decisions and actions determine later possibilities and outcomes). While making good choices during the assessment is a part of the construct, it may nonetheless confound other metrics. Innovative analytical models therefore need to account for these dependencies across items and the implications of additional constructs (such as choice) that are in play (see Chapter 8 of this report). New methods are also required to model changes in a students' proficiency over an extended task, as interactive and resource-rich test environments afford possibilities for students to learn over the course of the assessment.

It is not just process data that can pose interpretation and scoring challenges – for some 21st Century competencies it is also challenging to interpret students' outcome data. For example, in an assessment of creative thinking, there may be an infinite number of possible creative outcomes or solutions that cannot be pre-defined. In these cases, automated scoring methods are not possible without integrating some sort of sophisticated artificial intelligence or machine learning model; and human scoring of such complex and open-ended responses can also be unreliable and require a large investment in time and financial resources. Moreover, without clearly defined learning progressions for these competencies, it can be difficult to classify and score outcomes that reflect different levels of mastery.

Reporting

Once evidence from a test has been identified, collected and interpreted, it needs to be accumulated and reported in ways that are appropriate and useful for the intended purpose of the assessment. The reporting

of student performance in large-scale assessments is especially important as these assessments are used to inform policymaking and system-level reform (Vista, Kim and Care, 2018^[19]).

21st Century competencies are complex and multidimensional constructs meaning that assessments intending to measure them should aim to make claims on students' ability across those different dimensions. One challenge in terms of reporting relates to how to generate evidence in such a way that the data scale together to provide an overall picture of student performance on the test, while at the same time providing insights on students' relative strengths and weaknesses. In large-scale assessment, test developers have prioritised achieving a single, reliable scale that summarises students' overall performance. Even when sub-scales are developed to measure different dimensions of the construct, these sub-scales are generally highly correlated with each other so that high (or low) performing students on the overall scale tend also to be high (or low) performing across all the sub-scales. More actionable information on strengths and weaknesses would require developing a more diverse set of item types that target the different dimensions of the construct, but this poses practical challenges including higher development costs and requiring a longer assessment time.

Another reporting challenge concerns how to describe complex behaviours in a way that provides actionable information to the users of the assessment data. For example, digital tests make it possible to record the different strategies that students adopt to solve complex, open problems. Analysing this process data can reveal different “profiles” of problem solvers, for example those who rapidly test ideas versus those who pause and reflect before attempting a solution. These data on solution processes provide a window into how students reason and construct their knowledge, giving potentially useful insights on the quality of instruction they have received. However, data on processes might be hard to convert into a linear point scale as some processes are not necessarily preferable to others. While some methodologies, such as cluster analysis, can be used to identify and describe profiles (e.g. different types of problem solvers), such reports are more complex to understand for policymakers and other users of the assessment data.

Validation

Innovative assessments of complex constructs like 21st Century competencies require more comprehensive validation processes that go beyond psychometric considerations of the reliability of scores (Vista, Kim and Care, 2018^[19]; Care et al., 2018^[18]; Ercikan and Oliveri, 2016^[21]; Ercikan et al., 2016^[24]). Like all assessments, innovative assessments need to be built on strong validity arguments that demonstrate that they measure what they intend to (construct validity), that student performance in the test is related to performance in relevant and authentic real-life situations (external validity), and – in large-scale assessment – that performance across student groups is comparable (fairness and cross-cultural and cross-linguistic comparability).

One set of challenges relates to whether 21st Century competencies are understood and expressed in similar ways across cultures and student groups. While it may be reasonable to assume that some competencies, like problem solving or critical thinking, involve similar thinking processes across cultures and student groups, others – such as those with stronger inter- or intra-personal components – are likely to be more sensitive to cultural or gender differences in terms of how they are expressed. This idea also extends to the ways in which different cultures or student groups may value different 21st Century competencies and consider them appropriate in certain situations (e.g. defining whether an output is creative or whether a problem requires a creative solution). It is therefore important that the target populations of any assessment of 21st Century competencies share a common understanding of the target construct to ensure there is a good degree of construct equivalence (Ercikan and Oliveri, 2016^[21]).

Another set of challenges relates to validating test tasks and items and ensuring fairness among student groups. Many of these validation issues apply to all assessments: task contexts should aim to be equally familiar to students from different backgrounds, while avoiding socio-cultural, gender or other types of biases; task instructions (and translations, where applicable) should be expressed in the most appropriate

way to ensure that students clearly understand what is required of them; and response modes should be simple and intuitive so that construct-irrelevant factors do not unduly influence performance. In technology-enhanced assessments that provide more open and interactive test environments, the user interface design and user experience need careful attention so that the test remains accessible to all students. Students from different socio-cultural backgrounds may lack familiarity with such digital assessment platforms and their affordances, which may create sources of incomparability and jeopardise cross-cultural comparability. Cognitive laboratories and log data analysis can provide insight into students' thinking processes and test taking experiences that can be used for the purpose of validating tasks (see Chapter 12 of this report), but these exercises require a greater human and financial investment during the test development process (Ercikan and Oliveri, 2016^[21]; Care, Anderson and Kim, 2016^[12]).

The third set of validity challenges relates to the evidence and scoring models employed in innovative assessments of 21st Century competencies. A central tenet of score comparability across cultural and language groups is measurement invariance, which refers to the degree to which similar constructs are being measured and scores are comparable for test taker groups. In innovative assessments that integrate evidence from process data into scoring models, one challenge is ensuring that those processes are equivalent across student groups. However, students from different socio-cultural backgrounds or with different levels of prior knowledge may not use the tools and affordances of the test environment similarly, or they may not use the same strategies to solve tasks. This differential engagement can threaten the validity of conclusions and comparisons about student performance if evidence of those processes is interpreted and scored as evidence of mastery of the construct. International assessments further compound this challenge as linguistic or cultural differences may also lead to behavioural differences. For a comprehensive discussion of these issues in large-scale, innovative assessments of 21st Century competencies, see Chapters 11 and 12 of this report.

Finally, as technology-enhanced assessments are able to capture more complex process data, more complex and automated analyses for interpreting and scoring data using machine learning and artificial intelligence algorithms have also been developed (DiCerbo, 2020^[25]). However, if the data sources used to create these scoring algorithms are not representative of all cultural and student groups then the resulting scores may not have equivalent validity and accuracy for all groups. One challenge for implementing such methods at scale is ensuring that all relevant student groups are adequately represented in training the algorithms, and that the scoring algorithms are sufficiently evaluated for validity, accuracy and comparability of scores for all of the diverse student populations.

Conclusion

21st Century competencies are increasingly recognised as key competencies for today's young people to develop so that they can effectively participate in the global knowledge economy, thrive in an increasingly diverse society, use new technologies effectively, adapt to change and uncertainty, and continue to engage in lifelong learning. Nonetheless, there remain significant challenges to the adoption of this agenda in practice across all aspects of the education system, from curriculum to pedagogy to assessment. This chapter focused on discussing six major challenges associated with designing valid assessments of complex 21st Century competencies. Doing so requires innovation throughout the entire assessment development process: from defining the conceptual framework, to task design, test delivery, validation, scoring, analysis and reporting.

Developing assessments of 21st Century competencies first requires clearly defining the target construct and establishing the theoretical underpinning of the assessment. The extent to which domain-specific knowledge supports the target construct also needs to be addressed in the conceptual framework as this will affect how assessment tasks are contextualised and the extent to which student performance in the assessment can be generalised. Without a clear conceptual framework, well-defined learning progressions

or construct maps, it is difficult to interpret and score student performance. This challenge is further amplified by the fact that 21st Century competencies are characterised as much by process as by outcome, that there is often no single or pre-defined “correct” response and that constructs are multidimensional (so performance may not be uniform across all dimensions). Because of the relative emphasis on students’ processes, assessment tasks need to be more open and interactive to provide opportunities for students to demonstrate how they engage in those processes. Task designers must therefore identify the types of complex yet accessible problems and situations that call for students to engage 21st Century competencies as well as develop test environments that allow students to respond in authentic ways and that generate interpretable evidence about their ways of thinking and doing. Finally, these challenges all pose additional demands in terms of establishing the validity argument for innovative assessments to ensure that tasks and scoring methods are equally accessible for different student groups and free from cultural, gender and linguistic bias.

References

- Binkley, M. et al. (2011), “Defining twenty-first century skills”, in Griffin, P., B. McGaw and E. Care (eds.), *Assessment and Teaching of 21st Century Skills*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2324-5_2. [4]
- Care, E., K. Anderson and H. Kim (2016), *Visualizing the Breadth of Skills Movement Across Education Systems*, The Brookings Institute, Washington, D.C., https://www.brookings.edu/wp-content/uploads/2016/09/global_20160916_breadth_of_skills_movement.pdf (accessed on 23 February 2023). [12]
- Care, E. et al. (2018), *Education System Alignment for 21st Century Skills: Focus on Assessment*, The Brookings Institute, Washington, D.C., <https://www.brookings.edu/wp-content/uploads/2018/11/Education-system-alignment-for-21st-century-skills-012819.pdf> (accessed on 12 February 2023). [18]
- Chalkiadaki, A. (2018), “A systematic literature review of 21st century skills and competencies in primary education”, *International Journal of Instruction*, Vol. 11/3, pp. 1-16, <https://doi.org/10.12973/iji.2018.1131a>. [10]
- Csapó, B. et al. (2011), “Technological issues for computer-based assessment”, in Griffin, P., B. McGaw and E. Care (eds.), *Assessment and Teaching of 21st Century Skills*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2324-5_4. [22]
- Darling-Hammond, L. et al. (2019), “Implications for educational practice of the science of learning and development”, *Applied Developmental Science*, Vol. 24/2, pp. 97-140, <https://doi.org/10.1080/10888691.2018.1537791>. [15]
- DiCerbo, K. (2020), “Assessment for learning with diverse learners in a digital world”, *Educational Measurement: Issues and Practice*, Vol. 39/3, pp. 90-93, <https://doi.org/10.1111/emip.12374>. [25]
- Ercikan, K. and M. Oliveri (2016), “In search of validity evidence in support of the interpretation and use of assessments of complex constructs: Discussion of research on assessing 21st century skills”, *Applied Measurement in Education*, Vol. 29/4, pp. 310-318, <https://doi.org/10.1080/08957347.2016.1209210>. [21]
- Ercikan, K. et al. (2016), “Use of evidence-centered design in assessment of history learning”, in Braun, H. (ed.), *Meeting the Challenges to Measurement in an Era of Accountability*, Routledge, New York, <https://doi.org/10.4324/9780203781302-18>. [24]
- European Commission (2019), *Key Competences for Lifelong Learning*, Publications Office of the European Union, Luxembourg, <https://data.europa.eu/doi/10.2766/569540> (accessed on 28 February 2023). [7]
- Fadel, C. and J. Groff (2018), “Four-dimensional education for sustainable societies”, in Cook, J. (ed.), *Sustainability, Human Well-Being, and the Future of Education*, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-78580-6_8. [3]

- Joynes, C., S. Rossignoli and E. Amonoo-Kuofi (2019), *21st Century Skills: Evidence of Issues in Definition, Demand and Delivery for Development Contexts*, Institute for Development Studies, Brighton,
https://assets.publishing.service.gov.uk/media/5d71187ce5274a097c07b985/21st_century.pdf
 f (accessed on 16 March 2023). [11]
- Lai, E. and M. Viering (2012), “Assessing 21st century skills: Integrating research findings”, *Paper presented at the National Council on Measurement in Education, Vancouver, B.C.*, Pearson,
http://images.pearsonassessments.com/images/tmrs/Assessing_21st_Century_Skills_NCME.pdf. [14]
- Learning Policy Institute and Turnaround for Children (2021), *Design Principles for Schools: Putting the Science of Learning and Development Into Action*,
https://k12.designprinciples.org/sites/default/files/SoLD_Design_Principles_REPORT.pdf
 (accessed on 16 March 2023). [16]
- OECD (2018), *The Future of Education and Skills 2030*, OECD Publishing, Paris,
[https://www.oecd.org/education/2030-project/about/documents/E2030%20Position%20Paper%20\(05.04.2018\).pdf](https://www.oecd.org/education/2030-project/about/documents/E2030%20Position%20Paper%20(05.04.2018).pdf) (accessed on 1 March 2023). [8]
- OECD (2016), “Automation and independent work in a digital economy”, *Policy Brief on the Future of Work*, OECD Publishing, Paris. [1]
- Paniagua, A. and D. Istance (2018), *Teachers as Designers of Learning Environments: The Importance of Innovative Pedagogies*, Educational Research and Innovation, OECD Publishing, Paris, <https://doi.org/10.1787/9789264085374-en>. [17]
- Partnership for 21st Century Learning (2019), *A Framework for Twenty-First Century Learning*,
<http://www.p21.org/> (accessed on 1 March 2023). [13]
- Pellegrino, J. and M. Hilton (2012), *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, National Academies Press, Washington, D.C.,
<https://doi.org/10.17226/13398>. [2]
- Roll, I. et al. (2014), “On the benefits of seeking (and avoiding) help in online problem-solving environments”, *Journal of the Learning Sciences*, Vol. 23/4, pp. 537-560,
<https://doi.org/10.1080/10508406.2014.883977>. [23]
- Scott, C. (2015), “The futures of learning 2: What kind of learning for the 21st century?”, *Education, Research and Foresight: Working Papers*, UNESCO,
<https://unesdoc.unesco.org/ark:/48223/pf0000242996>. [5]
- Vista, A., H. Kim and E. Care (2018), *Use of Data From 21st Century Skills Assessments: Issues and Key Principles*, The Brookings Institute, Washington, D.C.,
<https://www.brookings.edu/wp-content/uploads/2018/10/EffectiveUse-Vista-Kim-Care-10-2018-FINALforwebsite.pdf>. [19]
- Voogt, J. and N. Roblin (2012), “A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies”, *Journal of Curriculum Studies*, Vol. 44/3, pp. 299-321, <https://doi.org/10.1080/00220272.2012.668938>. [9]

Wilson, M. et al. (2011), “Perspectives on methodological issues”, in Griffin, P., B. McGaw and E. Care (eds.), *Assessment and Teaching of 21st Century Skills*, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-94-007-2324-5_3. [20]

World Economic Forum (2015), *New Vision for Education: Unlocking the Potential of Technology*, https://www3.weforum.org/docs/WEFUSA_NewVisionforEducation_Report2015.pdf [6]
(accessed on 16 March 2023).



From:

Innovating Assessments to Measure and Support Complex Skills

Access the complete publication at:

<https://doi.org/10.1787/e5f3e341-en>

Please cite this chapter as:

Foster, Natalie (2023), “21st Century competencies: Challenges in education and assessment”, in Natalie Foster and Mario Piacentini (eds.), *Innovating Assessments to Measure and Support Complex Skills*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/3637901c-en>

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.