



B | Center for
Universal Education
at BROOKINGS

OPTIMIZING
ASSESSMENT FOR
ALL

October 2018

Use of data from 21st century skills assessments: Issues and key principles

Alvin Vista, Helyn Kim, and Esther Care

Alvin Vista is a fellow at the Center for Universal Education at Brookings

Helyn Kim is a postdoctoral fellow at the Center for Universal Education at Brookings

Esther Care is a senior fellow at the Center for Universal Education at Brookings

Optimizing Assessment for All (OAA) is a project of the Center for Universal Education at the Brookings Institution. The aim of OAA is to support countries to improve the assessment, teaching, and learning of 21st century skills through increasing assessment literacy among regional and national education stakeholders; focusing on the constructive use of assessment in education; and developing new methods for assessing 21st century skills.

Acknowledgments

The Brookings Institution is a nonprofit organization devoted to independent research and policy solutions. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.

Brookings gratefully acknowledges the support provided by Porticus.

Brookings recognizes that the value it provides is in its absolute commitment to quality, independence, and impact. Activities supported by its donors reflect this commitment.

TABLE OF CONTENTS

Executive Summary	1
Introduction	3
What are the major functions of assessment?	7
Teaching and learning in the classroom	8
Monitoring and accountability	9
The context and purpose of 21st century skills assessment	14
Providing insights into the teaching and learning processes	14
Aligning the system	16
What are the specific challenges to measuring 21st century skills?	18
Assessment design	19
Validation issues	20
What are the implications of 21st century assessments for data use and reporting?	22
Current practices of data use and reporting in education	23
Flow of data in the current data use process	23
Role of data in decision-making and informing policy	25
Evolving practices of data use and reporting for 21st century skills	25
General principles for using assessment data	29
Design data collection processes that are aligned with purposes and agenda at all levels	29
Establish a clear link between the captured form and the intended reported form of data	30
Principles for specific education stakeholders	32
National policymakers and decision-makers	34
Researchers	34
District and school leaders	34
Teachers	35
Parents and students	35
Conclusion and recommendations	36
Endnotes	38
References	39
Supplementary Bibliography	42

Executive summary

With the learning goals of education shifting to include a broader range of skills, the challenge globally is how to support students in developing these skills. The components of the education system must be aligned to support the development of 21st century skills, and the qualitatively different structure of these skills requires some completely new approaches, both in the measurement aspect and collection of assessment data. A major issue that confronts education systems is a deficiency in the effective use of collected student learning outcomes data. Notwithstanding the huge sums that are dedicated to its collection, a proportional commitment is not made to its strategic analysis or its dissemination.

The main purpose of this publication is to provide guidance on how data from 21st century skills assessment can be used and interpreted in terms of learning outcomes to inform teaching and learning. Towards this purpose, we put forward actionable recommendations that are both applicable and relevant to the current state of assessing these 21st century skills to enhance learning outcomes, as well as forward-looking in anticipating the future of assessment. In this publication, we consider the purposes of collecting student achievement data associated with 21st century skills, discuss how these data are currently used in

various contexts and the challenges associated with each, and finally provide key principles for effective data use both generally and specific to major stakeholders.

Beginning with a discussion of what demarcates 20th and 21st century skills in the context of assessment, we consider the main purposes of the practice. These purposes are roughly dichotomized across formative and summative types of assessment. Formative assessments are undertaken throughout the teaching and learning process, with the direct purpose of improving the learning outcomes of those students being assessed. Summative assessments are typically conducted at the end of learning processes to evaluate students' learning outcomes by comparing them with some validated standards or benchmark. However, it is the purpose of the assessment, rather than the type, that leads to different uses of assessment data—for teaching and learning, as well as for monitoring and accountability.

The current state of teaching and assessment of 21st century skills is outlined, with acknowledgement of some major issues. These include our lack of understanding and knowledge of how 21st century skills can be taught, and the possible lack of alignment

between traditional curricula and a 21st century skills learning agenda. Although the use of large-scale assessment data for system accountability and monitoring is well-established, specific information about student performance expectations for 21st century skills is provided by only a handful of education systems that have these skills formally embedded in their curriculum, and therefore, have the mechanisms for system-wide data collection, use, and dissemination.

Challenges specific to assessment of 21st century skills may be one reason why education systems are having difficulty translating policies into actual practice in schools and classrooms. These include the inherent nature of transferable skills that can be demonstrated across different situations and in response to different contexts. Such skills require assessment approaches that are either sufficiently broad or sufficiently dynamic to capture this essential quality. Most educational assessments tend to reward “correct” answers to clearly defined questions. Although there are several instruments and advanced assessment approaches that have been demonstrated to capture 21st century skills, the challenge is how to use these systemically and ensure they are not only valid and reliable, but also practical, in the contexts that they are to be used. Additional challenges include lack of comprehensive operational definitions, lack of standards for making evidence-based inferences, threats to generalizability, cross-cultural validity of the definitions, and measurement errors.

Finally, focusing on the main purpose of this publication, a set of general principles are presented and discussed in detail. These general principles, in summary, are the following:

- 1) **Design data collection processes that are aligned with purposes and agenda at all levels.**
Avoid redundancies in data collection that occur when collecting too much or overlapping data. Align different approaches across the system to meet the goal of enhancing learning.
- 2) **Establish a clear link between the captured form of data and the intended reported form of data.**
Ensure that the data-capture process is systematic. Clearly specify the format and structure of the reporting framework. Provide considerable infrastructure and support systems for implementation. Build a timely and regular assessment program that can provide sufficient assessment data at an individual level. Always take into consideration that the quality of the data itself depends on several factors, from measurement precision to the representativeness of the sample, and failure to take these limitations into account can result in misleading interpretations and conclusions.

Complementing these main general principles are data use principles that are relevant to specific stakeholders, focusing primarily on researchers, district and school leaders, and teachers. Relating to the use of 21st century assessment data, we find that 21st century assessments require solid research support for researchers, need a data-driven instructional systems model for improving instructions for district and school leaders, and require the assessment process to be contextualized and incorporated into teaching and learning processes innovatively for teachers.

Introduction

The main purpose of this publication is to serve as a practitioner- and policy-oriented resource focused on the specific issues of data use in the context of assessing 21st century skills.

The changes in our economy and society in this century have placed a greater emphasis on the skills that citizens need to be successful. This diverse set of skills, often referred to as 21st century skills, and including critical thinking, creativity, problem solving, communication, and socio-emotional skills, among others, are in high demand as the need for rote or routine-based knowledge decreases due to automation in the workplace (Rotherham & Willingham, 2010). In education specifically, there is the concern of a global learning crisis—that students are not fully prepared with the skills they need to thrive in today’s rapidly changing world (The Education Commission, 2017). Consequently, international, non-governmental, private sector, academic, and governmental organizations around the world are focusing their attention

on addressing this challenge. For example, the United Nations set 17 Sustainable Development Goals, one of which provides for inclusive and quality education for all children (United Nations, 2015). Others, such as Partnership for 21st Century Skills (P21), Organisation for Economic Cooperation and Development (OECD), and Assessment and Teaching of 21st Century Skills (ATC21S), have developed theoretical frameworks that identify and describe 21st century skills. While the skills themselves are not new, the recent attention and interest in these skills by multilateral organizations have created a new focus on how to measure them with the same rigor as traditional learning domains. Moreover, countries around the world are broadening their learning goals beyond the acquisition of traditional academic skills such as literacy and numeracy, to include 21st century skills such as problem solving, critical thinking, and collaboration (Care, Anderson, & Kim, 2016). There is a move toward integrating 21st century skills in education systems in order to equip students with the skills they need to have successful

lives and livelihoods. As the goals of the education system change, this brings about many challenges. One of these challenges is to review how assessment is used to measure a more diverse set of learning goals.

According to the National Research Council's Committee on the Foundations of Assessment (NRC, 2001, p. 1), "Educational assessment seeks to determine how well students are learning and is an integral part of the quest for improved education. It provides feedback to students, educators, parents, policymakers, and the public about the effectiveness of educational services." Various stakeholders (e.g., education sector providers, governmental bodies, international agencies, and researchers) conduct assessments to inform or make changes to the education system that improve learning. This central concept of using assessment to improve learning is related to the broader concept of using measurement to inform the current *state* for the purposes of changing that *state*. In the context of education, the change most people are looking for is improvement of learning outcomes. Resting on the premise that we need to improve the learning of 21st century skills, we need to develop 21st century assessments to evaluate the extent to which changes to the education system are effective.

The term "21st century assessment" encompasses multiple aspects of learning assessment, including the skills assessed and the methods used. While some aspects of 21st century assessment are new, principles of good assessment apply as much in this century as they did in the last. The distinction between 20th and 21st century assessments lies in the increasing use of newer technologies and psychometric methods, and in the increasingly diverse number and type of skills being assessed. Most 20th century assessments are analyzed using classical test theory approaches, such as treating counts of correct responses as the primary indicator¹ of performance and reporting on percent

correct. While modern measurement approaches such as item response theory and structural equation modeling techniques have been developed in the mid-20th century, it was only through the advent of personal computers that their use began to be practical at school level. More importantly, the spread of these modern measurement approaches has enabled the development of test instruments that go beyond correct-incorrect scoring and percent-correct reporting (see, for example, Hambleton & Jones, 1993). When applied to the assessment of 21st century skills, modern techniques enable the development of multidimensional² measurement tools and the capture of rich response data (e.g., multiple response types on a single item, process data, etc.). The rapid pace of digital technology has enabled rich and authentic platforms for these 21st century skills to be demonstrated and measured, such as in-game environments and digital spaces that allow manipulation of 3D objects. The combination of digital technology and modern measurement approaches means that indicators of these competencies can now be captured and measured in real time.

There are a number of approaches to assessment of 21st century skills, and examples are provided in Table 1. These examples will be revisited throughout this document. With the main focus of this publication being the use of data from 21st century skills assessment, the examples are used to provide tangible evidence of the new methods and tools as well as examples of the problems, as identified in this publication.

Given the relative newness of assessing 21st century skills, at least when it comes to integrating them within education systems, many questions still remain unanswered:

- What are the major functions of assessment broadly?

Table 1

Example approaches and implementations of 21st century skills assessment

Approach	Example implementations		
	Program/project	Description	Target 21st century skill/s
Assessing 21st century skills through conventional methods that are commonly used for traditional learning domains (e.g., numeracy and literacy).	International Civic and Citizenship Education Study (ICCS; Schulz et al., 2016)	Survey consisting of a cognitive (or knowledge) test and an affective-behavioral questionnaire. The cognitive test uses multiple choice and open-ended response items. The questionnaire is composed mainly of Likert-type and categorical response items.	Civic knowledge, attitude and engagement, and behavior related to citizenship
Embedded assessments that are woven into the fabric of the learning environment. Mainly using conventional measurement methods (e.g., multiple choice items) but the environment and format are augmented by technology.	SimScientists (Quellmalz, et al., 2009)	A platform that combines science lessons with embedded assessment. Presented in a digital slide-type format with some animation and interactivity. The assessment components are embedded as multiple choice items with automated scoring and feedback.	Content-based (science) critical thinking
Finite state systems and tasks with fixed manipulable elements and interactions among the elements (e.g., single-player digital games with fixed interactive elements).	MicroDYN (Greiff, et al., 2012) and MicroFIN (Funke, 2001)	A set of tasks done in a digital environment where players manipulate multiple variables to solve a complex problem. An example task for MicoDYN requires participants to manipulate variables for advertisement strategies to maximize the popularity of several target products.	Complex problem solving (a multidimensional construct composed of three dimensions: <i>information retrieval, model building, and forecasting</i>)
Game or task based assessments with open or flexible state systems that capture, collect and analyze background <i>process data</i> ³ . Although the manipulable elements in the environment are fixed, the interactivity across the elements and players is non-finite.	Assessment and Teaching of 21st Century Skills (ATC21S; Griffin & Care, 2015)	Pair-based set of tasks done in an asymmetric digital environment where players interact with each other to collaboratively solve a problem. The environment is designed so that the task cannot be solved by one player alone (i.e., allocation of resource and information are asymmetric) and therefore requires both players to work together. Both players can manipulate the task elements and act within the task environment openly and flexibly.	Collaborative problem solving (a multidimensional construct composed of <i>cognitive processes</i> and <i>social processes</i> dimensions)

- What is the purpose of measuring 21st century skills?
- What are the specific challenges to measuring 21st century skills?
- What are the implications of 21st century skills assessments for data use and reporting?
- How can the skills be assessed in such a way that results are meaningful and useful for various stakeholders?

This publication explores these questions and describes how data from 21st century skills assessment can be used and interpreted in terms of learning outcomes to inform teaching and learning.

What are the major functions of assessment?

Assessment has various purposes, and therefore, should be intentionally designed and used in a way that is consistent with the intended purpose. For instance, assessments can be used for *formative* or *summative* purposes or both. Specifically, formative assessment may be conducted at the beginning as well as throughout the teaching and learning process with the intent to identify, monitor, and improve student learning through ongoing feedback. Formative assessment, due to its function, is typically undertaken by the teacher who administers, evaluates, and uses the information—such assessment is invariably classroom-based.

On the other hand, summative assessment, such as a mid-year or end-of-year examination, is typically conducted at the end of a unit, lesson, or program to evaluate student learning by comparing against some standard or benchmark. Summative assessment is used by a wide variety of stakeholders, and may be classroom-based or undertaken in larger scale contexts. For instance, summative classroom-based assessments will typically be used to measure achievement; larger scale contexts assessments will provide information for use by educational leaders or

policymakers to evaluate systems or obtain information about whether standards have been met. Common across summative classroom-based and larger scale contexts assessment is the provision of achievement data by the student. This means that a single type of assessment may have the potential to be used for multiple purposes.

As such, the terms ‘formative’ and ‘summative’ are less about the type of assessment and more about the functions they serve; in practice, the distinction between the two may not be clear or even meaningful (Newton, 2007). The focus, therefore, when designing and using assessments, should be less about the type and form of assessment and more about the functions they serve.

In summary, assessments can serve various functions: supporting instructional improvement and promoting student learning in the classroom; providing evidence for accountability review at the classroom, school, or provincial/national levels and monitoring system performance. Each of these functions is discussed below, with specific reference to 21st century skills.

TEACHING AND LEARNING IN THE CLASSROOM

The primary purpose of classroom teaching is to improve student learning. This can be supported through using assessment results to provide effective feedback to students; adjusting teaching to consider results; and actively involving students in their own learning and assessing (Black & Wiliam, 1998). Both formative and summative assessments can be used to improve learning, but they differ mainly in scope. Formative assessment is a continuous process, while the scope of summative assessments is linked with a specific learning objective.

Whether assessment is able to improve learning, however, is related to the quality of the data and how well the information is used to feed back into the teaching and learning. Issues include:

- classroom-based assessment practices can lead to surface and rote learning that focuses on recall of isolated information or test-taking strategies (e.g., teaching to the test);
- overemphasis on the outcome of the assessment (e.g., grades) rather than on learning;
- use of normative rather than criterion-referenced approaches to assessment, placing the focus on rankings rather than on determining each student's learning progress.

Current assessment practices reflect uneven implementation and embedding of 21st century skills in the teaching and learning process. In some education systems where 21st century skills have been specified or are at least implicit in the curricula, there is some evidence of formal teacher training on developing assessment tools and using assessment data to improve teaching (Asia-Pacific Education Research

Institutes Network, 2015). In a study of nine countries in the Asia-Pacific region that examined assessment policies and practices of 21st century skills, or transversal competencies (TVC) as referred to in the study, teachers from nine countries indicated that they had some access to TVC assessments (Care & Luo, 2016). The teachers reported that information generated from the tools was used by them both formatively and summatively. Despite the reports from this qualitative study, details of actual implementations remain unclear in terms of how teachers are using data from 21st century skills assessments to inform their teaching and learning practices.

To have a better understanding of the broad picture regarding the current state of skills assessment, not only in these countries, but across the region, a 2018 study with eight Asian countries/territories (Bhutan, Cambodia, Hong Kong, Malaysia, Mongolia, Nepal, Pakistan, and Vietnam) examined examples of currently existing assessment tasks, tests, and test items related to TVC in these participating studies, both at the national and classroom levels (Care, Vista & Kim, 2018). Findings show that the most common (and only) functions for national level tools that were provided by the eight countries/territories were for accountability purposes and summative reporting at the systems level. Classroom-based tools, on the other hand, were reported to be used for both summative and formative purposes, ranging in formats, such as open-ended questions, rating scales, and multiple choice items, and captured various TVC, including critical and innovative thinking, global citizenship, and interpersonal skills. However, a majority of these tools and items were not developed specifically to measure the targeted TVC. Of the 58 total tools and items contributed by the countries/territories, less than a quarter were identified as being specifically designed to capture TVC. Although the study highlights the fact that countries are beginning to identify opportunities in their current education systems for assessment of TVC that are more in line with changing education goals,

it is not clear that assessment data are actually being used for teaching and learning in the classroom.

Challenges in using data for teaching and learning 21st century skills in the classroom are, in part, related to the fact that an explicit focus on these skills at the national level is relatively new. Consequently, there may be lack of technical understanding of how to integrate assessment activities and use the information (Care & Luo, 2016). A major teacher training and professional development need is to build the capacity of teachers to teach and assess 21st century skills (Saavedra & Opfer, 2012). In Singapore, with the idea that “21st century learners call for 21st century teachers,” a model of teacher education for the 21st century was developed with a strong intention “to provide teachers with the best support for their work in 21st century classrooms” and develop the necessary skills, attitudes, and depth and breadth of content knowledge (National Institute of Education, 2009). These efforts to build teacher capacity and invest in high-quality professional development are essential if 21st century skills are to be the focus of education systems.

Another issue may be misalignment or lack of clear signaling between what is identified at national policy level and what is happening in schools and classrooms. According to the previously mentioned study in the Asia-Pacific region, at the system level, Hong Kong, India, Malaysia, Mongolia, Republic of Korea, Thailand, and Vietnam reported that they had conducted in-service teacher training on assessment of these competencies (Care & Luo, 2016). But at the school level, not all school leaders were aware of system-level mandates on TVC assessment, and there was variability across countries and schools within these countries as to whether teachers were informed of the policy (Care & Luo, 2016). Some teachers are not receiving guidance or materials regarding 21st century skills assessment, which may indicate a lack of consistent implementation within countries.

MONITORING AND ACCOUNTABILITY

Assessment can be used to monitor how a system is performing and evaluate its quality or effectiveness, which in turn serves to make or influence decisions. For example, national assessment results can be used to provide information about how students are performing against standards and whether they have met some basic level of proficiency; data from national assessments can be used to guide resource allocation and target underperforming schools. These are primary functions of national assessments, and a major focus of ministries when they implement national assessment (Greaney & Kellaghan, 2008). International large-scale assessment (ILSA) programs such as TIMSS and PISA also serve this function. However, because national assessments are directly linked with national education policy and key aspects of a country’s educational system such as the curriculum, while ILSA are more broad-based, each of these large-scale assessments varies in the role they can play in monitoring and accountability (Kellaghan, Greaney, & Murray, 2009).

When it comes to monitoring and accountability in the context of 21st century skills at the systems level, there is very little evidence that much is happening. According to a global scan of available online education policy documents conducted by Care, Anderson, and Kim (2016), countries across the world are increasingly identifying a variety of 21st century skills, such as critical thinking, creativity, social skills, communication, problem solving, and digital literacy, as valued outcomes of formal education learning experiences. In fact, in the most recent update in August 2017 that included over 150 countries (Figure 1; see <http://skills.brookings.edu> for a complete list of countries and data), 76 percent identify specific 21st century skills within their national policy documents. Despite the emphasis on 21st century skills, this has not translated into clearly defined implementation plans, descriptions of appropriate teaching strategies, or development of

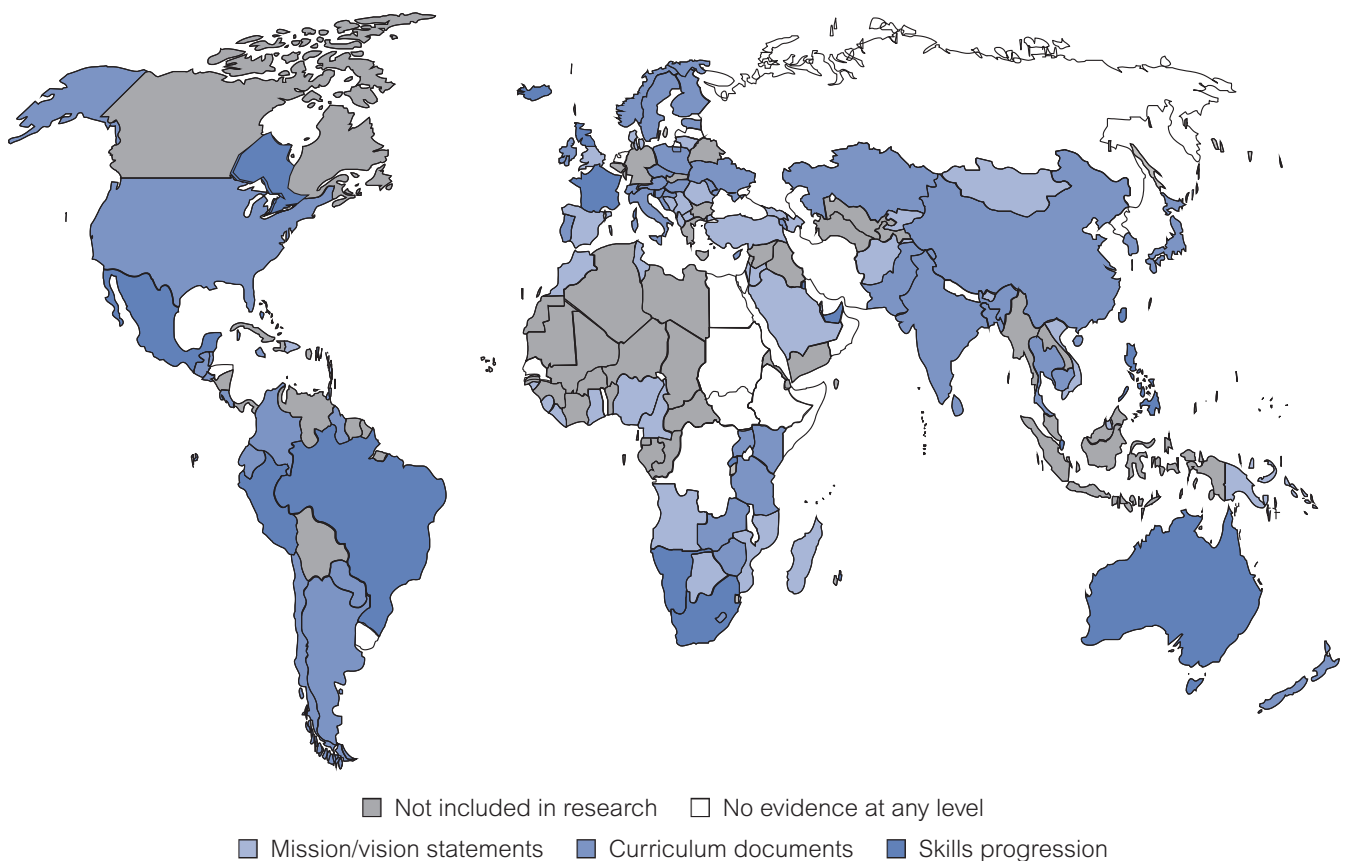
well-designed assessment tools. The issue may be the (lack of) depth and consistency with which countries are integrating these skills throughout their education system. For example, although a majority of the countries (117 countries out of 152) in the scan identified specific 21st century skills in their documents, 53 of these countries only do so in their mission and vision statements or general documents but not in their curricula; 58 countries reference skills in their curricular documents but do not show evidence of progressions

countries may be just beginning to think about and develop approaches to measuring 21st century skills.

Without these learning outcomes being specified clearly in curricular documents, as well as more broadly at education policy level, there can be no monitoring of student progress in skills acquisition. All assessment requires a framework if assessment is to provide meaningful information linked to curricular goals. These frameworks set the overall structure of

Figure 1

Identification of 21st century skills in education policy documents around the globe



of the skills from basic to more complex over time; and only 17 countries, such as Australia, Mexico, Singapore, and Namibia, include skills progressions within their policy documents. These findings suggest that

the assessment program, which are then operationalized more specifically through assessment/test design blueprints. Although many countries are mentioning the need for skills acquisition and development, while

others are identifying opportunities within the curriculum to develop these, there are few countries and/or provinces that are providing sufficiently detailed descriptions of student performance associated with 21st century skills, to enable assessment tools to be developed to capture these. Table 2 provides a list of countries and provinces that include detailed descriptions of student performance—or learning progressions—relating to 21st century skills. Although these countries have identified progressions or the concept of progress in their policy documents, the degree or depth at which these are specified vary across countries. Naturally, the identification itself does not imply that these policy goals are translating into actual practice in schools and classrooms.

To fulfill the monitoring and accountability function of assessment, which is to provide information of performance against standards, an assessment framework specific for 21st century skills is crucial. Systems that implement assessment programs for 21st century skills more formally, for example through a national program, need to have frameworks that explicitly establish what such an assessment program intends to measure and how the assessment will be structured. Although assessment frameworks specific for 21st century skills exist, such as the assessment framework for International Computer and Information Literacy Study (ICILS; Fraillon, Schulz, & Ainley, 2013), this is not yet the case at the national systems level.

Table 2

Countries/provinces and their status for inclusion of learning progressions of 21st century skills in policy documents

Country/Province	Examples of Focus Skills	Presence of 21st Century Skills Progressions
Australia	Literacy, numeracy, ICT, critical and creative thinking, personal and social capability	The curriculum is presented as a progression of learning from Foundation to Year 10 that describes what is to be taught and the quality of learning expected as learners progress.
Brazil	Communication, literacy, reflection, collaboration, critical thinking, life skills	Subject areas are broken down into objectives, which are ranged across four basic domain areas and across each year in school, progressing as students advance in school.
Canada (Ontario)	Critical thinking, problem solving, communication, collaboration	Ontario's curriculum includes skills as outcomes for each subject area and describes skills expectations within those subject for learners at each grade level.
France	Rational and critical thinking, problem solving, creativity, communication	The curriculum is separated into cycles that focus on these skills as they progress from basic knowledge and learning to more consolidated and in-depth learning.
China (Hong Kong)	Collaboration, creativity, information technology skills, problem-solving	The curriculum outlines expected outcomes for students in specific areas and lists specific progression of skills that are expected within the key learning areas.
Iceland	Communication, creative and critical thinking, using media and information	The Icelandic National Curriculum Guide maps progressions of the specific skills students need to develop and how these skills progress over time.

Kuwait	Communication, critical thinking, teamwork, self-evaluation, numeracy	Eight key competences cut across all subject areas. The curriculum materials are presented by subject, and each subject lists key competency targets for students at the close of each grade level.
Mauritius	Civic skills, critical, creative and innovative thinking skills, communication	The curriculum demonstrates a progression of stages within the education system, from foundation to consolidation to orientation.
Mexico	Critical thinking, problem solving, creativity, collaboration, digital skills	There is evidence of skills progression in the curricular proposal, which indicates the key curricular components and their sub-areas across basic education grade levels.
Namibia	Learning to learn, social skills, communication, information and communication technology	Phase competencies, the different levels to be attained in each learning area by the end of each phase, are broken down into detailed statements of what is expected that the learners understand and can do as they progress.
Peru	Social skills, entrepreneurship, communication	The Ministry website has a specific section entitled "How they learn" where details about skills progression at each education level are provided.
Philippines	Information, media and technology skills, learning and innovation skills, communication	The curriculum is conceived as a spiral progression to ensure integrated and seamless learning. Each subject area demonstrates how higher levels of schooling build upon competences developed at lower levels.
Rwanda	Critical thinking, problem solving, communication, cooperation, life skills	The curriculum includes profiles of student competencies at the end pre-primary, primary, and secondary school to capture student progress.
Scotland	Communication, critical thinking, problem solving, creativity	The curriculum describes progressions of achievement for learners within each of the eight subject area. Each subject area also has a desired set of skills outcomes.
Singapore	Critical thinking, civic literacy, collaboration, communication	Subject-specific syllabi provided online describe expected progressions for learners based on how students at various stages think, develop, and learn.
South Africa	Social skills, creativity, problem solving, critical thinking, communication	Curriculum documents describe the specific skills and content at each grade level, and defines progressions across different levels of school.
United Arab Emirates	Creativity, confidence, critical thinking, problem solving	The UAE Student Profile references holistic education systems as models and lists thematic, progressive targets for students by the end of each cycle.

There are of course significant challenges in achieving this function for 21st century skills, which cannot be overstated. Designing and implementing any assessment program for the purposes of system monitoring and accountability is complex even for conventional domains. In the context of monitoring and accountabil-

ity, one main challenge is that the complex nature of 21st century skills requires complex assessment tools that might not lend themselves easily to the restrictions of large scale assessment. The level of information detail required at national or systems level is much narrower than at classroom level. At the systems level,

only key information might be needed, which may consist just of a subset of the level of information needed in a classroom assessment. For example, although an assessment might capture multidimensional constructs (e.g., MicroDYN), the systems-level purpose might require only the information on the overarching construct (in the case of MicroDYN, just a score on *complex problem solving*). There is a case

for a bottom-up scaling where classroom assessments are “trimmed” of some functionality, when adopted for national use; as opposed to an approach of having the tools designed for national use be expanded to capture more detailed information for classroom use. The challenge, though, is in determining how to ensure alignment and the right level of information reaches the different levels of education.

The context and purpose of 21st century skills assessment

Current assessment systems tend to focus on academic subject area, numeracy and literacy, and students' recall of factual knowledge. Most assessments tend to be static in nature, providing "snapshots" of achievement at specific points. When it comes to 21st century skills like problem solving or collaboration, these traditional assessment formats (e.g., multiple choice tests) are less likely to sufficiently capture student ability to engage in complex processes, such as being able to apply what they have learned in one situation to a completely new situation. Therefore, if assessments are to advance the learning and teaching process, they need to provide useful and reliable information about what is it that students are learning and how they are progressing toward mastering their educational goals (Schoenfeld, 2017).

PROVIDING INSIGHTS INTO THE TEACHING AND LEARNING PROCESSES

Despite reform efforts to include a broad range of skills in their national education systems, one of the biggest challenges for ministries of education is implementing the teaching and learning of these skills in the classroom (Care et al., 2016). One potential way to overcome this challenge is for a shifting of norms to occur, where instructional practices are adaptive and where teachers continually seek evidence on where students are in their learning, what problems they may be having, what should come next in their learning if they are to reach the goals, and responding to student

learning state through use of a variety of pedagogical approaches and scaffolds (Corcoran, Mosher, & Rogat, 2009). If, as is asserted by Black and Wiliam (1998), assessments inform teachers about how the nature of student learning can enhance instructional practices, investing in designing transparent assessments of 21st century skills can provide valuable insights into the teaching and learning processes. Teachers do understand students' learning and development based on information derived from conventional classroom assessments. Detailed examinations of learning progressions test, validate, and extend the sequence of skills so teachers and parents have better understanding of how learning and skills develop over time. There is a caveat, however, that learning progressions are broad in nature and individual-level variation requires teachers to be adaptive.

Learning can be envisioned as a development of understanding and skills, which becomes progressively more sophisticated. Over the past several decades, the concept of learning progressions has been gaining momentum as a tool for assessment, teaching, and learning (e.g., Heritage, 2008; Hesse et al., 2015; National Research Council, 2001). Learning progressions have been defined as "descriptions of successively more sophisticated ways of thinking about an idea that follow one another as students learn: they lay out in words and examples what it means to move toward more expert understanding" (Wilson & Bertenthal, 2005, p. 3). Learning progressions, which are empirically grounded and testable hypotheses of how students' understanding become more advanced over time (National Research Council, 2007), provide a carefully sequenced set of building blocks that move the student toward mastery of the goal (Popham, 2007).

Conceptually, learning progressions can enhance assessment and instructional practices by specifically identifying what students know or do not know at particular points along a learning trajectory and using

this information to make instructional adjustments to scaffold students as necessary. The concept and application of learning progressions is not new. Learning progressions have been developed for a variety of subject areas, including English language arts, mathematics (Clements & Sarama, 2007; Hess, 2010; 2011), and science (Corcoran et al., 2009; Mosher, 2011). For example, a learning progression for understanding patterns, relations, and functions in mathematics class may entail: at a basic level, using "concrete, pictorial, and symbolic representations to identify, describe, compare, and model situations that involve change;" at mid-level, describing and comparing "situations that involve change and use the information to draw conclusions...;" and at a more complex level, "approximate, calculate, model, and interpret change..." (Hess, 2010, p. 15). At a national level, learning progressions have been used to create norms and standards for performance. For instance, Australia uses learning progressions for literacy and numeracy that identify the processes and the knowledge, skills, and dispositions involved along a continuum from the basic level to more complex levels (Australian Curriculum, Assessment and Reporting Authority [ACARA], n.d.). By describing common pathways for acquiring specific aspects of literacy and numeracy development, teachers can have a better understanding of where students are currently in their learning and where they need to go next in their development. To use literacy as an example, comprehending texts is identified as one of the sub-elements of literacy. A student at the lowest level is expected to "use behaviours that are not intentionally directed at another person to attend to, respond to or show interest in familiar people, texts, events and activities." A student at a middle level is expected to "use conventional behaviors and/or concrete symbols consistently in an increasing range of environments and with familiar and unfamiliar people to respond to a sequence of gestures, objects, photographs...to complete a task, respond to texts with familiar structures..., and respond to requests." Finally, at the highest level, students are expected to "use

conventional behaviours and/or abstract symbols consistently in different contexts and with different people to work out the meaning of texts with familiar structures...respond to questions, sequence events and identify information from texts with familiar structures, and use information in texts to explore a topic.”

But, the hypothesized progressions that describe the pathways students are likely to follow are unclear for most 21st century skills. For instance, what are the subskills that are needed for collaboration? What does a basic level of collaboration look like? What about more complex forms of collaboration? Attempts at developing learning pathways have been focused on very few 21st century skills that have theoretical support from large bodies of literature. Critical thinking, creativity, and problem solving have been measured in various ways for decades, but it was relatively recent that empirically-based learning progressions have been developed for more complex skillsets (e.g., collaborative problem solving in ATC21S and complex problem solving in MicroDYN). Among ILSAs, it was only in 2012 that creative problem solving was assessed in PISA (OECD, 2014), and in 2015 that PISA assessed collaborative problem solving.

Assessment is “a tool designed to observe students’ behavior and produce data that can be used to draw reasonable inferences about what students know” (Pellegrino, 2014, p. 236). Thus, measuring 21st century skills has the potential to elicit information about what these skills entail, what their building blocks are, and how the learning of these skills might progress, which in turn, can provide insights into the teaching and learning of these skills in the classroom. Moreover, the information can be used in conjunction with relevant theory and research about how students learn skills to support the development of learning progressions specifically for the skills. This could help address the challenges related to implementation and result in clearer links to the teaching and feedback that is needed to enable students to learn; reference points

for assessing students’ levels of progress and problem areas that need support; and in the long run, inform the design of curricula that are well-aligned with what students need to learn within and across different grade levels to progress to more sophisticated forms.

ALIGNING THE SYSTEM

Teaching and learning occurs within a whole education system—across multiple levels, including classroom, school, and nation. Aspects of learning that are assessed and emphasized in the classroom should be consistent with the aspects of learning that are targeted and assessed at the school and national levels—a concept referred to as *vertical coherence* (National Research Council, 2001). The level of information gathered from large-scale and classroom assessments may be different—for instance, in large-scale assessments, the understanding of learning may be coarser, whereas in classroom assessments, the information may be at a finer-grained level. Regardless, assessments within a system, as mentioned above, should provide aligned information across the different levels of education, so that results are consistent, albeit more or less detailed, as one moves up and down the levels of the system (National Research Council, 2001). Assessment must also be well-aligned, and ideally seamlessly integrated, with curriculum and pedagogy, so that the components of the education system are working toward a common set of educational goals—a concept referred to as *horizontal coherence* (National Research Council, 2001). Conceptually, this means that the goals outlined in the curriculum are fully measured, without adding irrelevant aspects; and instruction should match both the goals and the measures (Baker, 2005).

Given that 21st century skills assessments are relatively new, their usefulness and sustainability relies on the alignment and integration within existing national systems, rather than seen as something separate. Systems are not static and will shift as values, goals,

resources, and times change (Baker, 2005), requiring continual re-alignment. When the goal is concise and specific (e.g., walking), with a clear criterion identified (e.g., walk 15 steps), the goal, instruction, and measurement is likely to be consistent. Alignment can be difficult to achieve when the education goals are broad and general, and when there is an issue of transferability to different contexts and situations, as is the case for 21st century skills (Baker, 2005).

However, a focus on 21st century skills, which cuts across subject areas, could serve as a common underpinning to which the system would align. This would mean that the definitions and examples of these skills in different subject areas would need to be clear

and explicit, which poses a challenge due to their complex nature. For instance, what would “be able to identify a problem,” an important subskill of problem solving, look like in a mathematics class for 8-year olds compared to a language and literacy class for 11-year olds? Or, more specifically, what would it look like for a geometry word problem compared to when analyzing a literature passage? And, what are the learning pathways for these skills within and across grade levels? Achieving both vertical and horizontal coherence requires going beyond existing standards that establish what students should learn to a deep understanding of the skills themselves, as well as *how* students learn, in ways that will be useful for guiding instruction and assessment (National Research Council, 2001).

What are the specific challenges to measuring 21st century skills?

The focus on assessment of 21st century skills is increasing as education systems are broadening their educational goals. These skills are complex and their application to real-world situations is more important (in the context of teaching and learning) than their abstract conceptualization. Measuring 21st century skills is substantively different from measuring conventional learning domains such as numeracy and literacy because these skills emphasize what students can do with the knowledge they have acquired, rather than what that knowledge is. Because 21st century skills are *comparatively* more complex, non-routine, and dynamic, the measurement process needs to

take into account their application in real-life and non-familiar situations. For example, although numeracy skills can be measured in a very abstract manner (e.g., using test items that require solving equations), critical thinking skills may require tasks that elicit argumentation while being situated in a real-life scenario (Kuhn, 1992). Also, the skills are interdisciplinary and comprised of multiple interrelated elements. When it comes to subject areas such as literacy, numeracy, and science, there are learning progressions that describe the pathways students are likely to follow in their mastery of the specific subject area (e.g., National Research Council, 2007). This sequence of learning

identifies the essential concepts and skills that are being developed, as well as the expected behaviors and outputs that exemplify what students know and are able to do for each level of progress. There are examples of what this looks like for 21st century skills, such as the learning progression in *civic knowledge* for ICCS. However, many socio-emotional constructs remain elusive to define, not to mention developing clear learning progressions for them. Thus, while learning progressions exist for civic and citizenship *knowledge*, there is still no defined developmental progression for *attitude* and *engagement* in ICCS (Schulz et al., 2016).

For some educational skills, there is a strong and direct link between observable behaviors that are captured by the *proxy measures*⁴ and the target construct or skill. As such, the proxies serve as accurate indicators of the skill. However, as the skill becomes more complex, simple indicators become increasingly inadequate for capturing the volume of information that would be associated with increasing complexity of the construct. In other words, deciding on the indicators that operationally define the levels of increased understanding along the progression toward mastery is difficult.

ASSESSMENT DESIGN

In order to be useful, assessment tasks must be designed in such a way that the results provide evidence linked to student learning and can be used to make inferences and decisions (Pellegrino, 2014). Because the components and processes underlying 21st century skills are interrelated and complex, how they are considered in assessment design and scale development is a major challenge. Implementing assessments that are not well designed can conflict with curricular goals and undermine progress towards meeting them (Schoenfeld, 2017). If a complex skillset is underrepresented in an assessment because of the challenges in measuring it precisely, then tests may only assess fact-based information. This would have long-term

repercussions for meeting larger educational goals. Assessment tasks need to be designed in a way that allows learners of different ability levels to engage in the complex processes. This requires a systematic approach that deconstructs the skills to make more accessible the capture of specific sub-skills, or elements.

Capture of elements is an approach to minimize the problems in development of assessment tasks intended to measure the complex skill. Where a skill is in fact not unidimensional, then traditional test development approaches become limited because complex skill sets, traditionally represented by composite-score scales in classical test theory approaches, are better measured using more modern approaches such as item response modelling (Hambleton & Jones, 1993). Tests that rely on developmental continua, according to Shepard (2018), should be based on well-conceptualized constructs as well as empirical evidence and development work to build learning progressions. These can be used to ensure that there is coherence among curriculum, instruction, and assessment (i.e., horizontal coherence; National Research Council, 2001).

Another challenge has to do with the fact that these skills are generic and transferable, essentially domain-general. However, if assessments are to reflect real-life demands, the role that content knowledge plays (if any), needs to be taken into account. For instance, solving a complex problem in science requires an understanding of the relevant scientific content knowledge. One approach to assessment could be to provide the needed information as part of the assessment; another would be to assume that complex thinking cannot be isolated from the content knowledge; and thus, assessment of 21st century skills would include generic skills as well as specific content knowledge (Ercikan & Oliveri, 2016). This poses the issue of whether 21st century skills can be disentangled or isolated from domain-specific knowledge, or even whether an attempt should even be made (Ercikan & Oliveri, 2016)—and what might this mean in terms of

generalizability of the skills to different contexts. When tasks embed content knowledge, to what extent is student performance on the task reliant on the student's domain-specific knowledge rather than on the generic competencies, such as complex thinking or problem solving? To what extent does student performance on one task within a specific domain generalize to tasks in other content knowledge areas? These questions have implications for the validity of assessments of 21st century skills.

VALIDATION ISSUES

We note that validation is not about a measurement variable or even the measurement tool per se but rather about the interpretations we derive from the measurement process. As Cronbach and Meehl point out in their classic paper, “one does not validate a test, but only a principle for making inferences” (Cronbach & Meehl, 1955, p. 297). This is consistent with the view expressed by Messick (1990), that validity is not an inherent property of the measurement tool itself, but rather a body of evidence for the “empirical evaluation of the meaning and consequences of measurement” (p. 2) implying that there is no single metric for validity. This means that several criteria for validity provide the degree, rather than a definite valid-invalid categorization, to which the interpretation of a test result is appropriate for its intended purpose. This highlights the importance of validation but also emphasizes that it is a process where the aim is to improve along a continuous scale.

Although these validation issues can apply to any assessment tool, some are more relevant in the context of assessing 21st century skills. First, establishing construct validity—how well the assessment measures what it is intended to measure—is challenging when working with complex constructs, with no clear operational definitions of the skills. A related challenge is in establishing the set of standards which can be accepted as evidence for whatever inferences we

make related to the target construct. Standards must be defined *before* the validation process as to “avoid substituting *a posteriori* rationalizations for proper validation” (Cronbach & Meehl, 1955, p. 300). This means that the validation process has to be incorporated in the design of any measurement tool, especially if the target construct is new or not yet well established in the research community. Second, the role of content-specific knowledge in the assessment of 21st century skills poses potential issues for external validity, or the generalizability of inferences about student competencies in one content area to other content areas and, more importantly, to real-life situations (Ercikan & Oliveri, 2016). For instance, if a student can think critically about a particular literature passage, does that mean that the student is able to think critically about local or national political issues? Part of the issue here is in developing assessment tasks that can provide an indication that what a student is able to do is, in fact, domain general, rather than domain specific. Otherwise, there is a danger of over-generalizing the outcomes of the assessment. Once the fundamental issues of validity are addressed, other aspects of the validation process can be investigated. For example, cross-cultural validity—whether the definitions of the constructs are similar in different cultures or whether the skills develop in a similar manner in different cultures—may pose issues depending on the purpose and use of the assessment. Cross-cultural validity is an important issue especially for assessments of socio-emotional attitudes and behaviors. ILSAs that focus on these constructs, such as the ICCS, take these issues seriously and design their tools to take cross-cultural differences into account (Schulz et al., 2016).

For any measuring process, there is always some measurement error. The amount of error is magnified particularly in situations where the measurement is by proxy. This is often the case in education where direct measures are very rare. This is true of 21st century skills where sets of observable indicators (e.g., active communication, responding to prompts) act as proxies

for more complex constructs such as collaboration. The complexity of 21st century skills further magnifies the errors associated with their measurement. The consequence of greater error includes less precision of measurement, and associated issues with developing benchmark levels (e.g., cut-off scores that deter-

mine a minimum proficiency level) and interpreting results—especially for students that are close to the cut-off values. This is less of an issue for aggregate data use but becomes important when individual-level decisions are being made.

What are the implications of 21st century assessments for data use and reporting?

Assessment is increasing in education, and more data are being collected—for instance, as a result of implementing continuous assessment practices (Modupe & Sunday, 2015). Continuous assessment—assessing frequently—is the purposeful way of observing and documenting the work that students are engaging in and using the information collected to understand and extend their learning (Carlson, Humphrey, & Reinhardt, 2003). But, how are the data reported and used? How

data are reported is related to what it can and will be used for. For example, assessments that report one number, as a score or rank, may provide a summary of the level of achievement, but will be of little use for identifying what the student knows and is ready to learn next. On the other hand, assessments that report scores for each component or sub-area, as well as describing with “words and examples what it means to make progress or to improve in an area of learning”

(National Research Council, 2001, p. 137), may be useful for teaching and learning but may be too granular for informing national level policies.

It is relatively recent that the research community started to take a closer and serious look at issues related to data use and data-driven processes specific to 21st century skills assessments. The current state of data use remains patchy as can be expected given that sources of 21st century skills assessment data are relatively sparse. As demonstrated by the Care Vista & Kim NEQMAP (2018) study, there is potential in national education systems, at least in the participating eight Asian countries, to extend current assessment approaches to include assessment of 21st century skills at the national and school levels; yet, the evidence shows that tools developed specifically to capture these complex skills are lacking. Instead, they *happen* to be embedded in assessment of traditional academic domains, which means that data are not being captured for 21st century skills even when the items do tap into the skills. Despite shifting education goals and a desire to equip students with a broad range of competencies that can be transferred and applied to real-world situations, much of the attention is still directed towards academic subject areas. However, we can reasonably expect that data use will remain uniform across domains, whether they are traditional or fall under the 21st century umbrella. We examine current practices and how these would be affected as practices evolve when applied to the assessment of 21st century skills.

CURRENT PRACTICES OF DATA USE AND REPORTING IN EDUCATION

Systematic and large-scale data use in education only emerged in the last few decades, and has been particularly visible through the proliferation of international large-scale assessment programs. Test-based accountability systems (Marsh, Pane, & Hamilton,

2006) have formalized data-driven decision making (DDDM) frameworks which involve: data collection and organization → information analysis and summary → knowledge synthesis and prioritizing → decision making → implementation, and → impact evaluation, in an integrated cycle (Mandinach, Honey, & Light, 2006). The advantage of such frameworks is the applicability across levels of the educational system, from classroom to district to national levels, as well as within level. Education systems across the world have begun to adopt a DDDM framework in varying degrees of structural rigor, formality, and completeness but mainly in the context of traditional or core learning domains. Although the principles are substantially similar, there are only a handful of applications of data-driven decision making arising from 21st century skills assessment and these are mostly in the developed world. For example, the SimScientists program has been rolled out across several school districts in the United States where it was used to measure student progress against state standards and teachers have used the data to adjust instruction (Quellmalz, Silbergliitt, & Timms, 2011).

In examining the current practices of data use and reporting in education, we focus on two aspects: 1) the flow of data and 2) the role of data in decision-making. These provide a base for exploring how these practices evolve as assessment results increasingly include data from 21st century skills assessments.

Flow of data in the current data use process

There is typically a lag between data collection and data use in large-scale assessments, and so all consumers of data adjust their use (or are constrained by it) based on the amount of time that data processing and reporting takes. For example, due to logistical challenges and lengthy time requirements in systems-level data collection, the users of systems-level data often use them for diagnostic or accountability purposes rather than for interventions that would affect

the data providers. Accordingly, when systems-level data are used for instructional improvements, these target the next cohorts of students. This type of data use, common in ILSAs, may be contentious because data from one assessment event, which reflects learning opportunities provided by the system at one point are used to justify actions with different cohorts of students at another time. Over time, it can be presumed that opportunities provided by an education system will vary according to budgetary variations, education reform, etc.

On the other hand, individual or classroom-level data that are collected and processed quickly are often used for interventions that impact the sources of data directly. Classroom-level data are also often primary data (i.e., raw and individual level). If the primary data are part of systems-level assessment, these are also likely to be aggregated as the data move to higher levels. Generally, the flow of data in this context is towards increasing aggregation and separation from the primary sources.

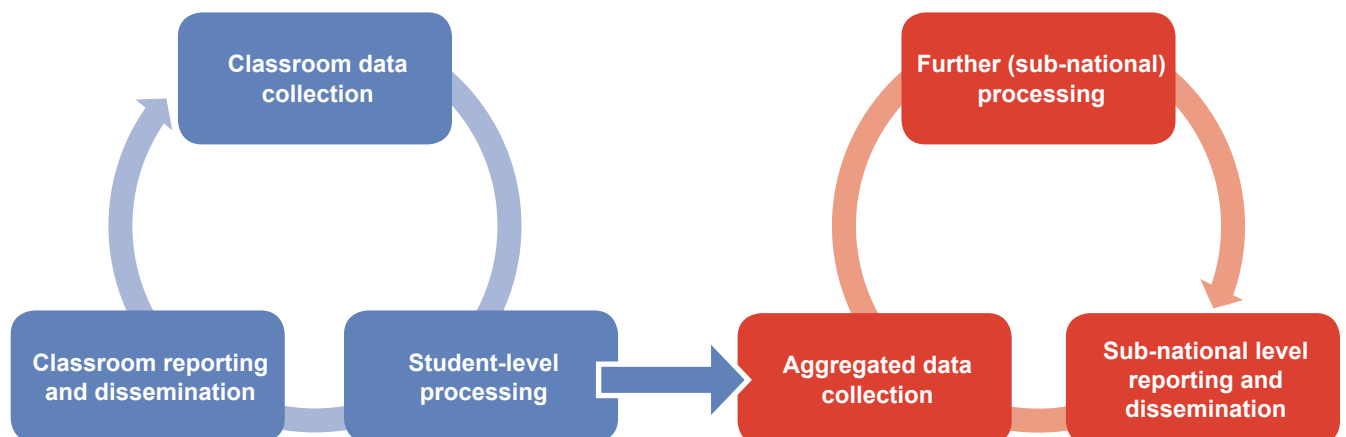
The flow of data can be separated into three main phases. The first phase is data collection while the second phase is data processing where raw data is

transformed in some way, whether through aggregation or analytical transformation. Processed data then flows to the consumers of data in the third phase, where data are used or processed further. Data reporting and dissemination occur in phase three. In most situations, the flow follows a cyclical process, where phases 1 to 3 occur during a cycle and repeat for each new cycle.

However, the data or some subset can continue on to a separate cycle. For example, within the classroom, data can be collected as part of the formative assessment cycle, but the same data can be aggregated and processed as part of the district or sub-national level cycle for accountability purposes. This flow is illustrated in Figure 2, where the data in phase 2 of the classroom-level cycle become part of phase 1 in the sub-national level cycle. This flow, however, may vary depending on the type and purpose of assessment. In ATC21S, where both student-level and population-level data reporting and dissemination were possible, results were processed at individual-level, and student location (or ability estimate) along the measured skill were reported in a learning progression format, while averages of the latent trait estimate would be used when reporting and disseminating population-level

Figure 2

Flow of data across different levels of data use cycles



results. In ICCS, although latent trait estimates are computed at individual-level (and therefore individual results are technically available), only sub-national and national-level results are reported because the assessment program is a survey (i.e., sample-based design).

Role of data in decision-making and informing policy

Recent assessment reforms are beginning to expand the role of assessment from traditional accountability and diagnostic roles to be more integrated into instructional reforms. However, in terms of systems-level policy formation, large-scale educational assessments at either national or sub-national levels remain the primary sources of data. In cases where a country participates in international testing programs (e.g., PISA, TIMSS), data from these are also used in policy formation to varying extents. Data from national assessments are primarily used to inform policies that lead to reforms in education system governance, curriculum planning and implementation, and educational financing.

While large-scale data provide systems-level pictures of student outcomes and are useful in informing policies, their scale means that logistical constraints limit their scope and timeliness. Small-scale data can be targeted and therefore inform school-based policies that affect individual instruction more directly. Formative assessments can also be more flexible and have broader scope (i.e., target more domains) and can evaluate near-term interventions better than large-scale approaches.

EVOLVING PRACTICES OF DATA USE AND REPORTING FOR 21ST CENTURY SKILLS

Data use and reporting for 21st century skills are to date limited. At the large-scale level, in part as a response to the Sustainable Development Goals (SDGs),

the Programme for International Student Assessment (PISA), an international survey from the OECD that aims to evaluate education systems globally, has implemented assessments of creative problem solving (PISA 2012), financial literacy (PISA 2012 and 2015), collaborative problem solving (PISA 2015), and global competence (PISA 2018); and the International Civic and Citizenship Education Study (ICCS 2016) established an assessment of global citizenship intended to provide internationally comparable indicators of civic knowledge and engagement to inform policies and practices (Schulz, Ainley, Fraillon, Losito, & Agrusti, 2016).

Although these large-scale assessments can raise awareness of what skills may be valued at a global level, there is emphasis on scores, ranking, and country comparisons. Additionally, the definitions of the constructs have been developed by a small number of “experts” rather than emanating from long-term use and consensus. The issue of consensus around understanding of these constructs is important, in particular because they are associated with values rather than only cognitive domains. Another important issue is the cross-cultural relevance of the definitions. There are four issues currently with including these skills in large-scale assessments:

1. lack of alignment of the skills with national learning goals
2. cross-cultural acceptability
3. limited understanding of how to teach and skills as an outcome of results from the assessment
4. limited understanding of what are acceptable standards for demonstration of the skills, meaning that interpretation tends to be through comparisons with other countries rather than through contemplation of what is reasonable in each context.

While educational policy relies on and is driven by results from a country's national assessment programs, results from ILSAs are strong motivators and often drive high-level discussions among stakeholders (including policymakers). Granted that these discussions and enthusiastic calls-to-action mostly occur during *negative* results in ILSAs, the four issues above remain particularly problematic. Putting aside these problems, international large-scale assessments could be used strategically to emphasize the need to teach these skills. However, the inherently lower levels of precision currently accessible in the assessment of 21st century skills, means that results should be treated with caution.

Where more fine-grained assessment information can be generated and used, as in the classroom, assessment of 21st century skills may be more useful especially if skills are aligned at classroom level with national learning goals, through the process of developing assessments using a bottom-up approach. Rather than having centrally developed national-level tests forced down to the classroom, it is worth considering an approach where classroom-based tests are scaled to national use. Whether the skills are targeted implicitly or explicitly in the classroom, engaging in the assessment process of these skills can make teachers better aware of how the skill is defined, thereby focusing on teaching the skills more intentionally. Similarly, students may become more aware of the importance of the skills. Accordingly, compared to the other purposes, assessments of 21st century skills may be most appropriate for informing day-to-day instruction, and could address some of the negative practices associated with classroom-based assessment.

The process of how assessment data can be used is summarized below:

- 1) Locate students along a learning progression and identify gaps in achievement

The ability estimates for each student's performance can be mapped onto a developmental continuum that represents the range of competency on a skill, such that the student's location can be identified along a learning progression scale. The student's location along this scale is then reported to the teachers.

- 2) Adapt instructional practices to individual needs and inform instructional improvement

Various professional development models that are based on developmental learning and evidence-based approaches use student data for individualized instruction. One such approach has been developed in Australia for professional learning teams where teachers, in small teams, use student assessment data to set goals, plan tasks, adjust or differentiate instruction as necessary, and monitor individual student progress in the classroom (Care, Griffin, Zhang & Hutchinson, 2014).

- 3) Track and communicate student progress

Tasks may be designed for formative use with fast testing and reporting turnaround. This enables multiple and regular reporting cycles, with each cycle providing an individual-level "Learning Readiness Report" (see example in Figure 3) to track and communicate student progress in a format that is both informative and accessible to lay-people.

- 4) Inform data-driven decision making at classroom- and school-levels

Student-level data can be aggregated and reported at classroom-level through a class-level report (see example in Figure 4). These summary data are used by the teachers to group students for differentiated instruction within their classrooms as well as learn from each other in their professional learning

Figure 3

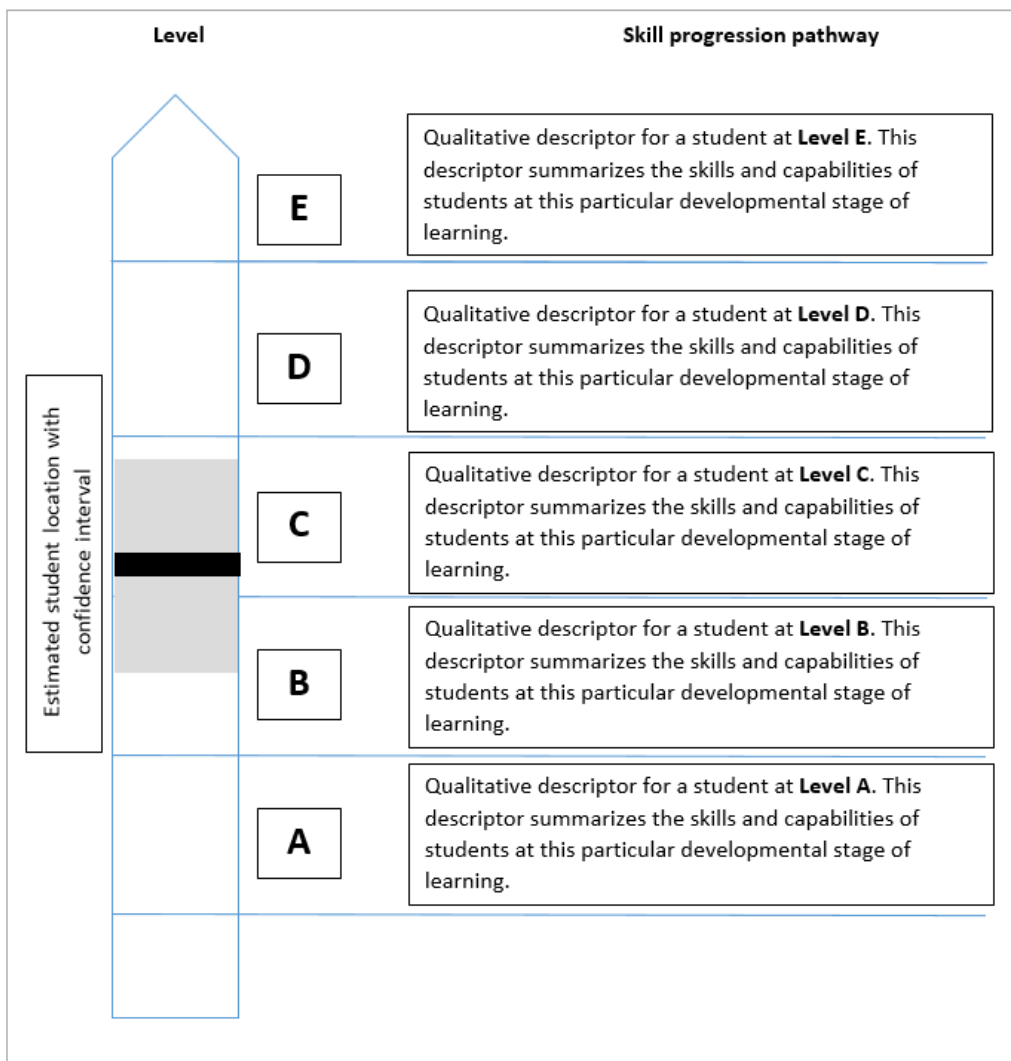
Sample Learning Readiness Report

Learning Readiness Report

Student Name: John Smith

Student Class: Grade X-XXX

Assessment: Learning Domain XXX



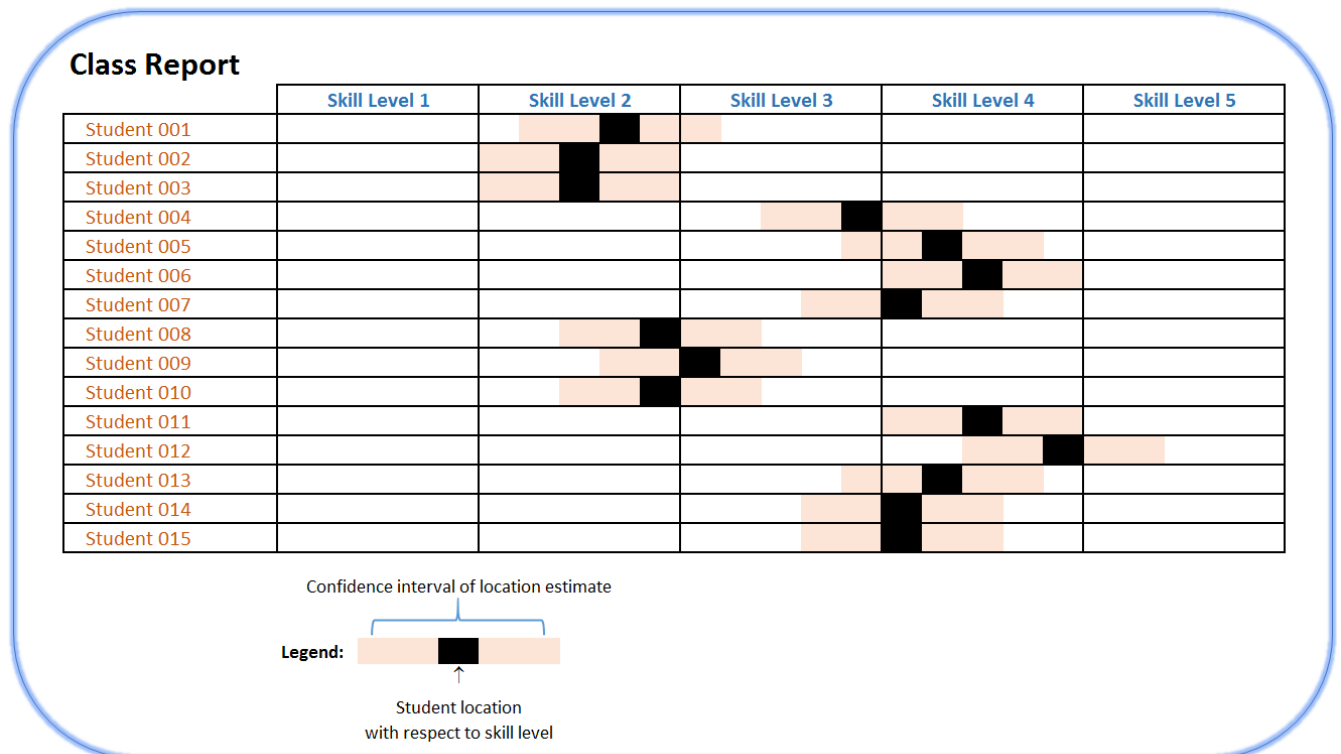
teams to monitor which interventions work well. Effective interventions across the team’s classrooms can then be generalized across multiple teaching teams and eventually to the whole school.

In recent years, technology has increasingly been used as a way to make assessment delivery more efficient. Quite apart from quick turnaround of the assessment cycle, the information—or process data—that can be captured by computer-delivered assessments can also be useful for exploring the nature of skills (the underlying processes and sub-skills) (Ramalingam & Adams, 2018). For example, the process data in the ATC21S assessment project (e.g.,

interactions between student and the problem space gathered through mouse movements; the number of mouse clicks on a particular item; the time taken to complete an item and the steps within each item) provide information that helps us to understand students’ thinking processes and engagement skills (Pöysä-Tarhonen, Care, Awwal & Häkkinen, 2018). This knowledge can then be used to improve construct validity, develop more comprehensive assessment that better target the processes underlying 21st century skills, and ultimately, allow for a deeper understanding of the nature of these skills (Ramalingam & Adams, 2018). This has important implications for the teaching and learning of 21st century skills in the classroom.

Figure 4

Sample class-level report



General principles for using assessment data

Data use practices will continue to evolve, but there are current general and stakeholder-specific data use principles that will remain relevant and applicable to this emerging field of 21st century skills assessment. These principles serve two main purposes: 1) provide a framework for guiding best practices; and 2) develop data literacy by synthesizing the best available research on 21st century skills assessment and data use. The following key principles are relevant for all data stakeholders, whether they are collectors or consumers of 21st century assessment data.

DESIGN DATA COLLECTION PROCESSES THAT ARE ALIGNED WITH PURPOSES AND AGENDA AT ALL LEVELS

Closely related to this principle is the importance of design to avoid redundancies among measures in the same educational system, as well as checking across various sectors (e.g., public-private, research, corporate, academia, etc.) to avoid collecting new

data when these may already have been collected. Cross-sectoral coordination is crucial in the effort to maximize the use of existing data. Reviews of assessment systems in the developing world show that the focus of systematic educational data collection is often on national large-scale educational assessments. These types of assessment are resource intensive and often wasteful in terms of actual data use. In a review of cost-benefit analyses, the costs of implementing large-scale assessments can be as high as 80% of the total assessment expenditures for developing countries, often necessitating external funding support, and yet are only able to provide system-level and not classroom-level indicators (Wagner, 2011). While more applicable to ILSAs, the complexity of assessing 21st century skills has increasing cost implications, making this key principle important in strengthening the link between assessment results and policy changes. Aligning data collection processes with purposes and agenda ensures that cost-benefits are maximized.

The issue of timeliness is also a concern because the focus at systems-level is on assessments that have long data-collection cycles. This can cause a disconnect whenever policymakers change during the interval between the start of an assessment program and when its findings can inform policy (Care & Beswick, 2015). Again, the scope of this issue has been traditionally limited to ILSAs, but the scope has become broader as 21st century skills are becoming integrated into large-scale assessments (e.g., collaborative problem solving in PISA).

This alignment is also important to assessment data specifically. Gipps and Cumming (2005) state that “The key issue is around *fitness for purpose*” (p. 696) in relation to use of assessment. In other words, different assessment approaches serve different purposes, and the approach must be aligned with the purpose. For example, if the purpose is high-stakes, such as tracking the performance of a system and making decisions about policies, then a standardized, summative

assessment that is high in reliability is important; if the purpose is to provide ongoing feedback in the classroom, the assessment approach would be different, and a focus on reliability and standardization is less important.

Each form and approach plays an important role within the education system, and one type is not better than another, but rather, the approaches need to be aligned across the system and across multiple levels (see Table 3) and complement (not substitute for) each other, in order to meet the goal of enhancing student learning.

ESTABLISH A CLEAR LINK BETWEEN THE CAPTURED FORM AND THE INTENDED REPORTED FORM OF DATA

This key principle emphasizes the importance of designing the data-capture process systematically and ensuring that it provides the format and structure that match the intended reporting frameworks. For example, different levels of data need to be taken into account to match the intended aggregation levels that need to be reported for the findings. One specific issue to consider for this key principle is that data aggregation is usually difficult to reverse. Thus, it is better to capture primary data in as granular form as possible because individual data points can be aggregated if the need arise, whereas already aggregated data cannot be easily disaggregated. For example, storing sub-skill scores as well as summary scores is a conservative approach. These can be kept in the database even if some users (e.g., schools or districts) only need the top-most level scores on even just aggregated statistics.

It is also important to adapt the reporting format to both audience and purpose. Most assessment findings, especially at the system level, are reported in aggregate. For example, achievement data are usually

reported as averages for schools and districts. If group differences are reported, these are usually based on broad demographic variables such as school type (public/private), location (urban/rural), and similar grouping variables. Official memoranda that are circulated only within the education ministry or to high-level stakeholders need to be supplemented with audience-targeted reports that are accessible to lay readers. The most common format of large-scale findings are the main report and the technical reports, both of which provide the state of educational performance at a very broad level (i.e., population-level statistics). These types of reporting need to be supplemented with reporting formats that are more appropriate and useful for school-level and classroom-level purposes.

Similarly, national assessments are fed back to schools and teachers, but usually in an aggregate form that provides the average achievement levels of the school but not of the individual students—note that PISA has a reporting scale that maps to a learning progression for problem solving and thus provides qualitative descriptions of skill at school-level (for sampled schools).

For assessments that report student-level results, state-of-the-art reporting formats such as individualized learning progressions (Figure 3) and performance indicators that are specific to construct dimensions

(e.g., MicroDYN) are only beginning to emerge in the field. This is to be expected, as these new reporting formats require considerable infrastructure and support systems to implement. It also requires a timely and regular assessment program to provide sufficient assessment data at an individual level.

Finally, the reporting strategy should consider the limitations of the data in making interpretations and conclusions. Raw data and corresponding results are always neutral, but the interpretations and conclusions are not. There is always a human factor involved in interpretation of data. The consumers of data need to take into consideration margin of error in their interpretation since these can have major implications for policy decision-making. The quality of the data itself also depends on several factors, from measurement precision to the representativeness of the sample. Failure to take this margin of error into account can result in misleading interpretations and conclusions. Given the relative lack of understanding of 21st century skills as learning domains, additional interpretive comment will be needed for the foreseeable future.

Principles for specific education stakeholders

Complementing the principles discussed above, there are data use principles and best practices relevant to specific stakeholders. Stakeholders of educational data can be broadly grouped into collectors of data or consumers of data, although these two are not mutually exclusive. There is evidence that some types of data usage are more frequently observed than others. System review, improvement planning, and student progress tracking are among the most common data uses while staff evaluation (including teacher performance) and instructional review (using data to deter-

mine which aspects of classroom instruction are effective) are the least common across district, school, and even classroom-level users (Means, Padilla, Gallagher, & SRI International, 2010; Newton, 2007). The common uses of data for various stakeholders are summarized in Table 3.

In the following, we discuss additional principles that relate specifically to 21st century assessment data for specific stakeholders listed in Table 2.

Table 3
Educational assessment data usage across stakeholder types

Stakeholder	Collector of data	Consumer of data	Data purpose and usage	Additional comments and illustrative description of data use
National Policymakers and Decision-makers (including ministry, state, or province level staff)	Yes (aggregate data)	Yes (aggregate data)	System review; System decision-making; System implementation	As part of the system review process, policymakers also use assessment data to review and potentially adjust the systems-level measures of quality education. In particular, findings from the ICCS program inform global citizenship and sustainable development education agendas (Schulz et al., 2016). System decision-making includes resource allocation and organizational intervention.
Researchers (e.g., NGOs, government, academia)	Yes (primary and aggregate data)	Yes	System review; System decision-making; Teaching review; Progress review; Diagnostic	Diagnostic use includes doing research on the correlates of achievement and/or performance (i.e., what factors affect student performance on a given assessment).
District and school Leaders (including principals)	Yes (aggregate data)	Yes	System (local) review; System (local) decision-making; School improvement planning; Staff development; Progress review (aggregate)	Local decision-making includes resource allocation and school or classroom-level intervention.
Teachers	Yes (primary and aggregate data)	Yes	Teaching review; Student intervention Progress review	Progress review includes the use of individual assessment data for instructional improvement, locating students on a developmental continuum for formative purposes (e.g., in ATC21S), and performance monitoring.
Parents and Students		Yes (primary data)	Progress review School and career decision-making	Decision-making includes the use of one's assessment results to decide on schools, academic pathway, and eventually career/vocation choices.

NATIONAL POLICYMAKERS AND DECISION-MAKERS

In recent years, ILSAs such as ICCS and PISA have started to include 21st century skills among the domains being assessed. This has increased the interest among policymakers to include these skills in their own curricula and explore systems-level assessments of selected skills. A key principle for policymakers and decision-makers is to **first set the policy goals, taking into consideration systems-level issues, before choosing the particular skills or skillsets that are to be included in the curriculum and assessed at national level.** National policy goals can be complex and involve issues both directly connected to education (including budget constraints and educational priorities) as well as issues that are only incidental (e.g., politics and changes in government) (Wagner, 2011). This key principle avoids *ad hoc* choices among policymakers as well as emphasize the importance of linking the implementation decisions with well-planned policy.

RESEARCHERS

A key principle relevant to 21st century assessment data use for researchers is that **21st century assessments require solid research evidence.** This pertains to: 1) effectiveness of new assessments (i.e., empirical support that the tools capture the complex constructs they intend to measure) and 2) effectiveness of data extraction or capture mechanisms (i.e., that the tools and processes are appropriate to the level of complexity of the data). Notwithstanding the increasing use of education data for decision-making, there is little empirical research on how effective these practices are and what mechanisms impact the efficacy of data use (Coburn & Turner, 2011). Coburn and Turner (2011) recommend the following:

- focus on the relationship between initiatives to promote data use and aggregate outcomes

- focus on describing the activities that are involved with data use initiatives
- investigate how data enters into streams of ongoing action and interaction as they unfold at multiple levels
- seek to understand the role of environmental, organizational, and group context in how the practice of data use unfolds
- investigate data use as a situated phenomenon as it unfolds in real time

More specific to 21st century skills, this key principle suggests that more research is needed on the following areas: 1) defining and validating the skills (such as collaboration or creativity), 2) new approaches to assessing complex constructs, and 3) state-of-the-art reporting methods/formats and their effectiveness (e.g., learning progressions of 21st century skills).

DISTRICT AND SCHOOL LEADERS

District and school leadership have a direct impact on teachers' data use. The key principle for this group is use of a **“data-driven instructional systems” model.** When district and school leadership develop and implement a data-driven instructional systems model, this key principle ensures that the program design process takes into account the complexity and unique challenges of assessing 21st century skills. This is closely related to the DDDM framework discussed previously, but focused on instructional improvement. One example of such a model is structured by Halverson and colleagues (2007) as consisting of six component functions: (a) data acquisition, (b) data reflection, (c) program alignment, (d) program design, (e) formative feedback, and (f) test preparation. These components form a continuous cycle and while originally designed for conventional learning domains, the model applies to other sources of educational data, including 21st

century assessments. Program design has implications for teacher training, allocation of district/school resources, and improvement planning. 21st century skills assessment data literacy among district and school leaders is crucial in order for them to properly design and implement a data-driven instructional model.

TEACHERS

Teachers need to rethink their role as both consumer and collector of data. They are in a unique position in the sense that they use data from all major types and functions of assessment. The challenge is in developing assessments that are aligned with the learning goals outlined in the curriculum, as well as in understanding the progression of learning that is required to reach more sophisticated forms of the skills. As a consequence of the qualitatively different requirements of 21st century assessment processes, such as the importance of capturing indicators of behavior rather than content knowledge, teachers have to adjust their approach to teaching and learning so that these indicators become observable. There is, however, no one-size-fits-all approach and each classroom will have unique contexts that affect how particular behavioral indicators are expressed.

As such, the key principle for data use among teachers is that it is **important to contextualize the as-**

essment to the needs of each individual student and the overall environment of the classroom.

Teachers know their students best and therefore are in the position to customize assessments appropriate to the needs of the classroom.

PARENTS AND STUDENTS

Parents and students are key users of individual-level assessment data. An important principle for parents and students is that **assessment data do not convey static information, and therefore, should be used to promote learning.** Using this principle, parents and students need to be proactive in demanding regular, current, and accurate data on learning and performance. Of course, since assessment results are also essential for many students to gain access to further education, training, or employment, their understanding of assessment needs to be broader than its function purely in the learning and teaching context.

As reporting methods become more advanced, or just novel, especially for more complex 21st century skills (e.g., scores for several indicators across multiple dimensions of a complex construct), raising assessment literacy is also necessary. Raising assessment literacy is an important first step in becoming effective consumers of assessment data. Parents and students should engage with school staff to understand the assessment results that are provided to them.

Conclusion and recommendations

This publication brings together key principles on assessment data use so that stakeholders develop data literacy in the emerging field of 21st century skills assessment. Assessment plays a significant role in education, as it is used to determine what students know and can do with regard to what is expected, and make decisions accordingly. With the focus of education shifting to include a broader range of skills, the challenge globally is how to support students in developing these skills (Care et al., 2016; Care & Luo, 2016). Having assessments of 21st century skills does not ensure that effective learning and teaching will take place, but assessments do provide systematic quantification of teaching and learning. Additionally, teachers can use assessments to promote learning of the skills, if the goals are clear and appropriate as made visible through the curriculum, if reliable and valid information can be gathered about what the student currently knows and is able to do related to the goals from assessment, and if that information is used

to identify ways to scaffold student learning through instruction (Pellegrino, 2014). In other words, the components of the education system must be aligned to support the development of 21st century skills.

The effective use of assessment data is only one component in the teaching and learning process, but it is an essential piece of the whole system. Many of the lessons learned and best practices on general data use in education remain important and applicable in the context of 21st century skills assessment, but the qualitatively different structure of 21st century skills requires new approaches, both in the measurement aspect and collection of assessment data.

One of the key principles emphasizes the importance of designing the data-capture process systematically. In the context of 21st century skills, the data-capture process is not well established compared to traditional domains. It is therefore recommended that robust

data-centered tools are developed to show that the data-capture process for these complex learning goals can be made as systematic as in traditional domains. This recommendation has a two-fold benefit: 1) tool development, especially if done across all levels of the school system from the classroom all the way to national scale, raises awareness through proof-of-concept approaches in tool/task development; and 2) the development process, undertaken collaboratively and through engagement of various stakeholders can have a cumulative effect on building a set of best practices.

Widespread adoption can be slow, just as it was for tools (e.g., mechanized standardized tests) and methods (e.g., item response theory) that were developed for conventional domains, where it took several decades for most modern methods to become standard across systems. Even for core domains such as numeracy and literacy, the adoption of modern data-collection tools and processes has not been universal. However, the build-up becomes faster as more stakeholders become aware of what methods exist and what data-capture processes are possible. To ensure that this recommendation's focus on *awareness raising* is optimized, sets of best practices and resources need to be readily accessible through multinational networks. Policymakers are more likely to adopt new

tools or data-capture processes if empirical evidence that they work is available, and if they see these being adopted by other systems.

Finally, we recommend that data reporting be aligned more closely to the stakeholder purpose and to the needs of the target consumers. It is important to be aware that style of use of data varies considerably across education system levels. For example, while data reporting of generalizable skills can adopt modern structures such as learning progressions, reporting strategies must remain aligned with national education aspirations. An example of how this recommendation can be applied to the implementation of learning progressions is to develop a set of standard qualitative descriptors for the levels of skill. This would enable more efficient data usage with a broader scope because it would ensure uniformity across schools in the system. Data for formative use would be aggregated effectively for summative or accountability use.

In these early days of implementation of 21st century learning goals through national education systems globally, our attention is focused on how to ensure that assessment can facilitate learning rather than merely attempt to grade or rank it.

ENDNOTES

- 1 We use the term “indicator” in a more technical sense, referring to observed (or manifest) variables that indicate or point towards an unobserved (or latent) construct.
- 2 In a measurement context, the term “dimension” refers to an aspect or factor of what is being measured. For example, the term “unidimensional” means that only one latent trait or factor is being measured.
- 3 *Process data* includes distinct key strokes, mouse movements, and all capturable time-stamped user activities in a digital environment. The process data can be analyzed either discretely (looking at specific markers that can be linked with cognitive processes) or holistically (looking at sets of connected markers, such as sequences of actions, that can be linked to more complex cognitive processes).
- 4 Measurement by proxy is a method wherein something that is difficult or impossible to measure directly is replaced by a related but more easily measurable variable. An example would be using infant mortality rate as a proxy for maternal health.

REFERENCES

- Australian Curriculum, Assessment and Reporting Authority (ACARA) (n.d.). *National literacy and numeracy learning progressions*. Retrieved from <http://australiancurriculum.edu.au/resources/national-literacy-and-numeracy-learning-progressions/>
- Asia-Pacific Education Research Institutes Network. (2015). *Preparing and supporting teachers in the Asia-Pacific to meet the challenges of twenty-first century learning ERI-Net Regional Policy Study Series (Vol. 3)*: Bangkok, Thailand.
- Baker, E. L. (2005). Aligning curriculum, standards, and assessments: Fulfilling the promise of school reform. In C. A. Dwyer (Ed.), *Measurement and research in the accountability era* (pp. 315-336). Mahwah, NJ: Erlbaum.
- Black, P., & Wiliam, D. (1998). Assessment in classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7-74. Retrieved from: <https://www.gla.ac.uk/t4/learningandteaching/files/PGCTHE/BlackandWilliam1998.pdf>.
- Care, E., Anderson, K., & Kim, H. (2016). *Visualizing the breadth of skills movement across education systems. Skills for a Changing World*. Washington, D.C.: The Brookings Institution.
- Care, E., Griffin, P., Zhang, Z., & Hutchinson, D. (2014). Large-scale testing and its contribution to learning. In C. Wyatt-Smith, V. Klenowski, and P. Colbert (Eds.), *Designing Assessment for Quality Learning*. Dordrecht: Springer.
- Care E., & Kim, H. (2018) Assessment of Twenty-First Century Skills: The Issue of Authenticity. In E. Care, P. Griffin, M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills. Educational Assessment in an Information Age* (pp. 21-39). Cham: Springer.
- Care, E., Kim, H., & Scouler, C. (2017). 21st century skills in 20th century classrooms. *Educadores*, Oct-Dec, 31-44.
- Care, E., & Luo, R. (2016). *Assessment of transversal competencies: Policy and practice in the Asia-Pacific region*. Bangkok and Paris: UNESCO.
- Care, E., Vista, A., & Kim, H. (2018). *Assessment of transversal competencies: potential of resources in the Asian region*. Bangkok and Paris: UNESCO.
- Carlson, M. O., Humphrey, G. E., & Reinhardt, K. S. (2003). *Weaving science inquiry and continuous assessment: Using formative assessment to improve learning*. Corwin.
- Coburn, C. E., & Turner, E. O. (2011). The practice of data use: An introduction. *American Journal of Education*, 118(2), 99-111.
- Corcoran, T., Mosher, F. A., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. New York, NY: Consortium for Policy Research in Education. Retrieved from: http://www.cpre.org/images/stories/cpre_pdfs/lp_science_rr63.pdf.
- Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the building blocks project. *Journal of Research in Mathematics Education*, 38, 136-163.
- Earl, L., & Katz, S. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Western and Northern Canadian Protocol for Collaboration in Education. Retrieved from: <http://www.wncp.ca/media/40539/rethink.pdf>.
- Edenborough, R. (2005). *Assessment methods in recruitment, selection and performance*. London: Kogan Page.
- Ercikan, K., & Oliveri, M. E. (2016). In search of validity evidence in support of the interpretation and use of assessments of complex constructs: Discussion of research on assessing 21st century skills. *Applied Measuring in Education*, 29, 310-318.
- Frailon, J., Schulz, W., & Ainley, J. (2013). *International computer and information literacy study: Assessment framework*. International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.acer.org/files/ICILS_2013_Framework.pdf.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, 7(1), 69-89.
- Gipps, C. V., & Cumming, J. J. (2005). *Assessing literacies*. In *International handbook of educational policy* (pp. 695-713). Dordrecht: Springer.
- Greaney, V., & Kellaghan, T. (2008). *Assessing national achievement levels in education (Vol. 1)*. Washington, DC: World Bank Publications.
- Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement*, 36(3), 189-213.
- Griffin, P., & Care, E. (2015). Methodology of the ATC21S Project. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills: Methods and Approach* (pp. 3-33). Dordrecht: Springer.
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership*, 17(2), 159.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from https://www.csai-online.org/sites/default/files/Learning_Progressions_Supporting_2008.pdf.

- Hess, K. (December 2011). *Learning Progressions Frameworks Designed for Use with the Common Core State Standards in English Language Arts & Literacy K–12*. National Alternate Assessment Center at the University of Kentucky and the National Center for the Improvement of Educational Assessment, Dover, N.H.
- Hess, K. (December 2010). *Learning Progressions Frameworks Designed for Use with the Common Core State Standards in Mathematics K–12*. National Alternate Assessment Center at the University of Kentucky and the National Center for the Improvement of Educational Assessment, Dover, N.H.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills: Methods and Approach* (pp. 37-56). Dordrecht: Springer.
- Kellaghan, T., Greaney, V., & Murray, S. (2009). *Using the results of a national assessment of educational achievement* (Vol. 5). Washington, DC: World Bank Publications.
- Kuhn, D. (1992). Thinking as Argument. *Harvard Educational Review*, 62(2), 155.
- Mandinach, E. B., Honey, M., & Light, D. (2006). *A theoretical framework for data-driven decision making*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: RAND Corporation.
- Means, B., Padilla, C., Gallagher, L., & SRI International. (2010). *Use of Education Data at the Local Level: From Accountability to Instructional Improvement*. Washington, D.C.: U.S. Department of Education, Office of Planning, Evaluation and Policy Development.
- Messick, S. (1990). *Validity of test interpretation and use*. Report. Educational Testing Service, Princeton, NJ.
- Modupe, A. V., & Sunday, O. M. (2015). Teachers' perception and implementation of continuous assessment practices in secondary schools in Ekiti-State, Nigeria. *Journal of Education and Practice*, 6, 17-20.
- Mosher, F. A. (2011). *The role of learning progressions in standards-based education reform*. CPRE Policy Brief. University of Pennsylvania. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1010&context=cpre_policybriefs.
- National Institute of Education (2009). *A teacher education model for the 21st century*. Singapore: National Institute of Education. Retrieved from https://www.nie.edu.sg/docs/default-source/te21_docs/te21-online-version---updated.pdf?sfvrsn=2
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the foundations of assessment. In J. W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.
- National Research Council (2007). *Taking science to school: Learning and teaching science in grades K-8*. Committee on Science Learning, Kindergarten through eighth grade. R.A. Duschl, H.A. Schweingruber, & A.W. Shouse (Eds.). Washington DC: The National Academies Press.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14, 149-170.
- Pellegrino, J. W. (2014). A learning sciences perspective on the design and use of assessment in education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 233-252). New York, NY: Cambridge University Press.
- Organisation for Economic Co-operation and Development (OECD). (2014). PISA 2012 results: Creative problem solving: Students' skills in tackling real-life problems (Volume V). OECD, Paris, France. Quellmalz, E. S., Timms, M., & Buckley, B. (2009). Using science simulations to support powerful formative assessments of complex science learning. In *Annual Meeting of the American Educational Research Association*. San Diego, CA.
- Pöysä-Tarhonen, J., Care, E., Awwal, N., & Häkkinen, P. (2018). Pair interactions in online assessments of collaborative problem solving: case-based portraits. *Research and Practice in Technology Enhanced Learning*, 13(12), 1-29. <https://doi.org/10.1186/s41039-018-0079-7>
- Quellmalz, E. S., Silbergliitt, M. D., & Timms, M. J. (2011). *How can simulations be components of balanced state science assessment systems* (Policy Brief). San Francisco, CA: WestEd.
- Ramalingam, D., & Adams, R. (2018). How can the use of data from computer-delivered assessments improve the measurement of twenty-first century skills? In E. Care, P. Griffin, & M. Wilson (Eds.), *Assessment and Teaching of 21st Century Skills: Research and Applications* (pp. 225-238). Cham: Springer.
- Rotherham, A. J., & Willingham, D. T. (2010). 21st-century skills: Not new, but a worthy challenge. *American Educator*, Spring, 17-20. Retrieved from <https://www.aft.org/sites/default/files/periodicals/RotherhamWillingham.pdf>
- Saavedra, A. R., & Opfer, V. D. (2012). *Teaching and learning 21st century skills: Lessons from the learning sciences*. New York: RAND.
- Schoenfeld, A. H. (2017). On learning and assessment. *Assessment in Education: Principles, Policy, & Practice*, 24, 369-378.
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). IEA International Civic and Citizenship Education Study 2016 Assessment Framework. Retrieved from http://www.iea.nl/fileadmin/user_upload/Publications/Electronic_versions/IEA_ICCS_2016-Framework.pdf.
- Shepard, L. A. (2018). Learning progressions as tools for assessment and learning. *Applied Measurement in Education*, 31, 165-174.
- The Education Commission. (2017). *Learning generation*. Retrieved from <http://educationcommission.org/wp-content/uploads/2016/09/Commission-2017-Progress-Report.pdf>.
- United Nations. (2015). *Transforming our world: The 2030 agenda for sustainable development*. Retrieved from <https://sustainabledevelopment.un.org/post2015/transformingourworld>.

Woods, K., Mountain, R., & Griffin, P. (2015). Linking developmental progressions to teaching. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills: Methods and Approach* (pp. 3-33). Dordrecht: Springer.

World Economic Forum. (2016). *The future of jobs: employment, skills and workforce strategy for the fourth industrial revolution*. Geneva, Switzerland: World Economic Forum.

SUPPLEMENTARY BIBLIOGRAPHY

- American Psychological Association. (2015). *Top 20 Principles from Psychology for PreK-12 Teaching and Learning*. Washington DC: American Psychological Association.
- Australian Council for Educational Research. (2017). *Principles of Good Practice in Learning Assessment*. UNESCO Institute for Statistics. Retrieved from <http://uis.unesco.org/sites/default/files/documents/principles-good-practice-learning-assessments-2017-en.pdf>.
- Bellanca, J. A., & Brandt, R. (Eds.). (2010). *21st century skills: Rethinking how students learn*. Bloomington, IN: Solution Tree Press.
- Braun, H., & von Davier, M. (2017). The use of test scores from large-scale assessment surveys: psychometric and statistical considerations. *Large-scale Assessments in Education*, 5(1), 17.
- Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, 13, 74-83.
- Ercikan, K., & Lyons-Thomas, J. (2013). Adapting tests for use in other languages and cultures. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbooks in psychology: APA handbook of testing and assessment in psychology, Vol. 3. Testing and assessment in school psychology and education* (pp. 545-569). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/14049-026>
- Lai, E. R., & Viering, M. (2012). *Assessing 21st Century Skills: Integrating Research Findings*. Paper Presented at The National Council on Measurement in Education, Vancouver, BC.
- Metcalfe, S.J., Kamarainen, A., Tutwiler M.S., Grotzer, T.A. & Dede, C. J. (2011). Ecosystem science learning via multi-user virtual environments. *International Journal of Gaming and Computer-Mediated Simulations*. 3(1)86-90.
- National Research Council. (2011). *Assessing 21st Century Skills: Summary of a Workshop*. J.A. Koenig, Rapporteur. Committee on the Assessment of 21st Century Skills. Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. Retrieved from: https://www.sri.com/sites/default/files/publications/imports/21st_century_skills.pdf.
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A framework for conceptualizing and evaluating the validity of instructionally relevant assessments. *Educational Psychology*, 51, 59-81.
- Quellmalz, E. S., Timms, M. J., & Buckley, B. C. (2009). *Using science simulations to support powerful formative assessments of complex science*. Retrieved from http://simscientists.org/downloads/Quellmalz_Formative_Assessment.pdf.
- SEAMEO INNOTECH. (2015). *Assessment Systems in Southeast Asia: Models, Successes and Challenges*. Quezon City: SEAMEO INNOTECH Regional Education Program.
- Silva, E. (2009). Measuring skills for 21st century learning. *Phi Delta Kappan*, 90, 630-633.
- Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st century competencies: Guidance for educators*. Washington, DC: Asia Society.
- The Organisation for Economic Co-operation and Development [OECD]. (2017). PISA 2015 Collaborative Problem Solving Framework. Retrieved from <https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. John Wiley & Sons.
- UNESCO. (2016). *2014 Regional study on transversal competencies in education policy and practice (Phase II)*. Asia-Pacific Education Research Institutions Network (ERI-NET). Paris and Bangkok, UNESCO. <http://unesdoc.unesco.org/images/0024/002440/244022E.pdf>.
- UNESCO. (2017). *Global Accountability Monitoring Report 2017/2018. Accountability in Education: Meeting Our Commitments*. Paris and Bangkok, UNESCO

