

BROOKINGS

OPTIMIZING
ASSESSMENT FOR
ALL

November 2018

Education system alignment for 21st century skills

Focus on assessment

Esther Care, Helyn Kim, Alvin Vista, and Kate Anderson

Esther Care is a Senior Fellow at the Brookings Institution

Helyn Kim is a Postdoctoral Fellow at the Center for Universal Education at Brookings

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

Kate Anderson was an Associate Fellow at the Center for Universal Education at Brookings at the time of her contribution to this publication.

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.

Acknowledgements

We would like to extend our sincere appreciation to Dr. Scott Paris as a peer reviewer of an early version of this publication.

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.

In addition, 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 and the analysis and recommendations are not determined or influenced by any donation.

Table of Contents

Executive Summary	3
Introduction.....	4
Educational shift toward a 21 st century skills agenda	6
This is where the story starts	6
Is this a major shift?	8
Country emphasis on key 21CS	8
Regional level shift	10
What is needed to incorporate a 21CS agenda into an education delivery system?.....	13
Curriculum	13
Pedagogy	14
Assessment.....	14
Knowledge of the learning goals	14
What are challenges to incorporating a 21CS agenda?.....	15
The role of learning progressions in assessment of 21CS	26
Other assessment considerations	27
Conclusion	33
Challenge 1: Understanding the nature of 21CS	33
Challenge 2: Developing learning progressions of 21CS	33
Challenge 3: Designing appropriate and authentic assessment of 21CS	33
Re-aligning the system: Examples of country reform efforts.....	34
References.....	35

EXECUTIVE SUMMARY

There has been a major shift in educational learning goals—as seen most recently by Goal 4.7 of the Sustainable Development Goals—focused on global citizenship education and education for sustainable development. The shift concerns recognition of the need for education systems to equip learners with competencies such as problem solving, collaboration, critical thinking, and communication. The focus on these "21st century goals" is visible in education and curricular reform, and has been promoted by global discussion of changing work and societal needs. This paper describes global, regional, and national examples of this shift, and then focuses on implementation challenges. The paper focuses most explicitly on the issue of assessment but asserts that any major reform in an educational philosophy shift must ensure alignment across the areas of curriculum, pedagogy, and assessment.

The paper identifies several challenges to implementation of this educational shift. These include the need for clear understanding of the necessary skills—beyond mere identification of definition and description. This is essential if education systems are to reform curricula to integrate the new learning goals that the skills imply. A second challenge is the need for clear descriptions of what different levels of competencies in skills might look like. Although a few education systems have developed early frameworks which include increasing levels of competency, there are no generic examples that describe how some of these skills "progress." Such descriptions would enable teachers to know what to reasonably expect of a child in the early years of elementary school versus of a child in later years in terms of collaborative behavior or critical thinking. A third challenge lies in the obstacles that these first two hurdles pose to the development of assessments of 21st century skills (21CS). Without an absolutely clear understanding of a learning domain, or "construct," designing assessment frameworks and tasks are impossible. Without an understanding of what increasing levels of competency in a skill look like, it is not possible to draft the assessment tasks that will target different levels.

Educational assessment is both ubiquitous and unpopular. Despite increasing visibility of concepts such as "assessment for learning" or "formative assessment," which describes the constructive use of assessment to inform teaching, the primary use of assessment by national education systems remains summative—for use in certification, identification of eligibility for education progress, and system accountability. The assessment of 21CS, still in its infancy, does not lend itself easily to the modes of assessment that typically populate summative assessment approaches.

The paper identifies possible assessment approaches, using examples to highlight effective strategies for assessment of the skills, while acknowledging the technical difficulties associated with "capture" of behaviors in scoring and reporting them. In order to appreciate the implications of the nature of the skills for assessment, Gulikers, Bastiaens, and Kirschner's (2004) authenticity framework is used to evaluate the adequacy of specific assessment tools designed to measure these skills. This leads into a discussion of use of learning progressions both to model the development of complex skills, and as a scoring and reporting mechanism. Both expert-driven and empirical approaches to development of learning progressions are described, making clear that these progressions are central to moving the 21CS agenda forward.

A central issue in educational assessment concerns whether the same learning domain is being measured across the different populations where it may be administered. According to the vision of the Sustainable Development Goals, this means that all assessments should be appropriately targeted for different ability levels, and also for individuals from different cultures and sub-groups. Following a discussion of the cross-cultural issues relevant to assessment of 21CS, the paper looks at three countries—Australia, Kenya, and the Philippines—to identify how they are approaching the assessment and teaching of 21CS in their basic education sectors. The countries' varied emphases on curriculum, pedagogy, and assessment are of particular interest as a majority of countries around the world explore how to approach these challenges. These examples lead to the conclusion that learning progression models are key to ensuring alignment through the education delivery system. This requires a great deal of research both in academia and in the basic education sector before comprehensive programs are put in place, but it is a start.

INTRODUCTION

The 21st century has introduced new imperatives into education practices, stimulated by increasing concern about global inequities and lack of fairness. As demonstrated first by the Millennium Development Goals, and currently by the Sustainable Development Goals (United Nations, 2016), the concerns are about both access to education and quality of education. The 21st century policies about equity have consequences for how education systems must cater to all children and youth. Education systems need to provide educational experiences relevant to the 21st century world that youth face, and this means the introduction of new learning domains—domains that are characterized by transferable skills and competencies. Despite global, regional, and national aspirations toward a 21CS learning agenda, they are not yet translating into full-scale implementation at the school and classroom levels. In this paper, we briefly discuss the educational shift toward this agenda and what is needed to incorporate it within national education systems. We next focus on the overarching challenge of aligning curriculum, assessment, and pedagogy, and specifically, challenges including 1) understanding the nature of 21CS; 2) developing learning progressions of 21CS; and 3) designing appropriate and authentic assessment of 21CS.

The rapid changes we are experiencing globally, regionally, nationally, and locally are associated with increasing mechanization, and information generation and exchange. Hundreds of statements have been made over recent decades about changing landscapes (Schleicher, 2012) and each new era, such as the “New Industrial Revolution” or “New Information Age” (e.g., Anderson, 2012; Schmidt & Cohen, 2013). Although many of these identify the positive developments that these might foreshadow, many also identify concerns about the human element.

“Forget all the talk about machines taking over...What happens in the future is up to us.” (Schmidt & Cohen, 2013)

"In the industrial age and in analog clocks, a minute is some portion of an hour which is some portion of a day. You know, in the digital age, a minute is just a number. It's just 3:23. It's almost this absolute duration that doesn't have a connection to where the sun is or where our day is." Douglas Rushkoff

"Technology made large populations possible; large populations now make technology indispensable."
Joseph Wood Krutch (1893-1970)

Education has become the main mechanism providing individuals with the knowledge, skills, and competencies needed by the society of the day but educational provision typically lags behind the emergence of need. Many authorities (Robinson & Aronica, 2015) who draw comparisons between classrooms of yesteryear and today, have pointed out the few changes in classroom design and management. However, notwithstanding use of images of teachers standing in a clearly instructional role and students passively taking in information, there is a wide range of changed practices in many classrooms. Unfortunately, these changed practices are inequitably distributed. The most inexperienced and newly qualified teachers are those who are sent to rural and remote regions; similarly, these regions are historically and currently the least resourced in terms of buildings and teaching aid infrastructure (e.g., Adedeji & Olaniyan, 2011; Monk, 2007). Therefore, some of the most needy students are the least well-provided for in education.

What knowledge, skills, and competencies are society demanding of citizens today? Although there are differences across national values, cultures, and socio-economic character—which mold country economic and education policy—there is a common drive for individuals who are literate and numerate, with knowledge of global societies, who understand the scientific principles that underlie how the physical world operates, and who have the competencies and skills to function adaptively and effectively within their immediate environments, globally, and virtually. The Organization for Economic Co-operation and Development (OECD) (2015), employer-relevant organizations such as the World Economic Forum (2016), and national education systems (Care, 2018a), have made statements to this effect in response to employer concerns about the competencies and characteristics that individuals bring to the workplace, to national concerns about the socio-economic issues facing societies and the economy, and to community concerns about youth preparedness to contribute to society. The consequence is a growing emphasis on the need for students to emerge from education with competencies beyond the knowledge accumulation that was highly valued in 20th century education.

Table 1. Similarities across national key learning areas of selected countries

	Cambodia	England	Namibia	Singapore
<i>Primary</i>				
Common Subjects	Khmer Language Foreign Languages Mathematics Science Computer Social Studies Arts Education Physical Education and Sports	English Languages Mathematics Science Computing History Art and Design Physical Education	Languages Mathematics Natural Sciences Technology Social Sciences Arts Physical Education	English Languages Mathematics Science Social Studies Arts education Physical Education
Uncommon Subjects	Health Education Local Life Skills	Design and Technology Music Geography	Commerce Environmental Studies	Life skills Character and citizenship education Values in Action Music
Source	Curriculum framework of general education and technical education	The national curriculum in England: Key stages 1 and 2 framework document	The National Curriculum for Basic Education	Primary School Curriculum
<i>Lower Secondary</i>				
Common Subjects	Khmer Language Foreign Languages Mathematics Science Social Studies Information and Communication Technology Physical Education and Sports Arts Education	English Languages Mathematics Science History Computing Physical Education Art and Design	Languages Mathematics Natural Sciences Social Sciences Technology Physical Education Arts	English Languages Mathematics Sciences History Computer Studies Physical Education Art
Uncommon Subjects	Health Education Local Life Skills	Music Citizenship Design and Technology Geography	Commerce	Food & Consumer Education Design & Technology Fundamentals of Electronics Life Skills Character and Citizenship Education Geography Literature Music Media Studies Business Studies
Source	Curriculum framework of general education and technical education	The national curriculum in England: Key stages 3 and 4 framework document	The National Curriculum for Basic Education	Express Course Curriculum; Normal Course Curriculum

Through the late 1990s and particularly into the first two decades of this century, there have been thousands of articles, chapters, books, and blogs written about qualities needed for the 21st century. Google Scholar, at the time of writing,¹ lists 1,070,000 references on “21st century skills” since 1997 alone. Currently, the majority of education systems structure their educational offerings by subject or discipline areas (such as language, mathematics, history, science, etc.) across grade levels (Table 1). This approach tends to be ubiquitous across both high functioning education systems, as identified by OECD Program for International Student Assessment (PISA) results, and those which do not achieve the same level of learning outcomes for their students.

EDUCATIONAL SHIFT TOWARDS A 21ST CENTURY SKILLS AGENDA

This is where the story starts

Beyond the core knowledge and concepts that a basic education² provides, or that technical and vocational education within a formal education system provides, society demands that education systems equip graduating students with the ability to use and apply core knowledge and concepts. This would manifest through young people solving problems, communicating clearly, making evidence-based decisions, working together, and thinking creatively—all within the socio-cultural context of their societies. These competencies, combined with the attitudes, values, and ethics of their societies, have now become explicit aspirations of the formal education sector as illustrated by selected countries in Table 2. Aspirations do not vary greatly according to country educational achievement or economic income.

EDUCATION IN THE 21ST CENTURY – HOW IS IT DIFFERENT FROM THE 20TH CENTURY?

It's different by virtue of shifting learning goals; technologies that influence how we behave, perform, and produce; inclusion of students from populations not previously well represented; and globalization that affects some but not all learning environments. Assessment within the education provision in the 21st century is seen increasingly as playing a constructive role—supporting teaching and learning, and providing feedback to the education system about how it is performing.

Nearly all student information we collect, from which we infer student abilities, is derived from written output. For literacy and numeracy, this approach makes sense, since both the skill and outcome can be demonstrated in the same way as that in which the fully developed skill can typically be demonstrated. As interest has grown in competencies such as information communication technologies, we have again seen some convergence between the medium in which a competency is assessed and that in which it will typically be demonstrated—in rich digital environments, where interactivity and detailed behavioral actions can be captured and processed. Where do we go now, though, when education systems are increasingly geared toward the teaching and learning of processing skills and social skills? These are not easily sampled through pen and paper media.

¹ May, 2018

² In this paper, basic education is defined as the core education offered by most countries from around 5-15 years of age, or from Grades 1-10.

Table 2. National education mission statements from selected countries

	Mongolia	Portugal	Hong Kong	Finland
Mission / Vision Statement	The vision of the program is to create a familial, educational, and social environment that enables every child to be creative, confident, decision-making, co-operative lifelong learners, and a citizen of national language, culture, and traditions.	<p>MOE's vision for curriculum:</p> <ul style="list-style-type: none"> • The construction of consciousness of personal and social identity; • The participation of civic life in free, responsible, solidary, and critical capacity; • Respect and value of diversity of individuals and groups' belongings and choices; • Valuing the different ways of knowing, communicating, and expression; • Development of aesthetic; Appreciation of the world; • Development of intellectual curiosity, taste for knowing, work, and study; • Construction of ecological conscience and valuing the preservation of natural and cultural goods; • Valuing relational dimensions of learning and of ethical principles that regulate the relationship with knowledge and with others. 	The mission of the Education Bureau, The Government of the Hong Kong Special Administrative Region is: To enable every person to attain all-around development in the domains of ethics, intellect, physique, social skills, and aesthetics according to his/her own attributes so that he/she is capable of life-long learning, critical and exploratory thinking, innovating and adapting to change; filled with self-confidence and a team spirit; willing to put forward continuing effort for the prosperity, progress, freedom, and democracy of their society, and contribute to the future well-being of the nation and the world at large.	The purpose of education referred to in the Basic Education Act is to support pupils' growth into humanity and into ethically responsible membership of society and to provide them with knowledge and skills needed in life. Furthermore, the aim of pre-primary education, as part of early childhood education, is to improve children's capacity for learning.
Source	Upright Mongolian Child National Program (in Mongolian)	National Curriculum for Basic Education: Essential Competencies	Basic Education Curriculum Guide	The Basic Education Act of Finland

Is this a major shift?

Education systems are designed to meet the needs of society, although not always the needs of all within it. It has been widely assumed that students who progress successfully through a basic education system will emerge with competencies valued and needed by society. Dissatisfaction with the degree to which this goal is being achieved has contributed to education systems now being explicit about new learning domains and including them as specific goals of the education experience.

How do we refer to these new learning domains? Are they 21st century competencies or skills? Are they general capabilities, soft skills, life skills, or social-emotional skills? Or are they intra-personal and inter-personal skills? Are they global competences? Do they include values, attitudes, and ethics? Are these terms distinguishable from each other?

The term “21st century skills” has been interpreted in many ways, but is generally considered to denote a combination of skills that are important in a modern society and workforce (Ercikan & Oliveri, 2016). The terms “transferable” (UNESCO, 2012) or “transversal” (Care & Luo, 2016) competencies encompass some of the same skills that can be applied across multiple situations, in contrast to technical vocational skills which are specific to particular occupations. “Soft skills” is another term used in some regions; it is commonly understood to refer to a combination of interpersonal and intrapersonal skills, and may include emotional characteristics, attitudes, and values. In this paper, we are concerned with learnable and teachable competencies that have been identified by countries’ national education systems as part of their core aspirations. Although there is a diverse range of competencies identified by different education systems, they generally include skills such as critical thinking, collaboration, communication, problem solving, and digital literacy. For convenience, throughout this paper, we refer to these transferable or generalizable skills as 21CS. This is not to deny that other important skills that have traditionally been hallmarks of education systems are transferable—for example, literacy, and numeracy.

There are many reports that discuss the difficulties in terminology, labeling, and frameworks and structures. This paper is agnostic as to the specific knowledge, skills, attitudes, values, and ethics (Binkley et al., 2012) that should be sought. It is similarly agnostic as to evaluation of particular frameworks or structures. (For review of frameworks, see Dede (2010) and current work being undertaken at Harvard [<https://easel.gse.harvard.edu/taxonomy-project>]). This paper is focused squarely on those competencies that are most frequently identified in the aspirations of national ministries of education, as found in a large-scale mapping of over 150 education systems (Care, Anderson & Kim, 2016).

Country emphasis on key 21CS

A large-scale mapping study of nations’ aspirations to equip students with these 21CS revealed how widespread the shift to their explicit identification has become (Care & Kim, 2018). Countries are broadening their educational provision beyond a sole focus on academic achievement. However, the degree to which aspirations have translated to implementation within education systems varies across countries. In the study, how far countries have moved toward implementation was explored through several indicators:

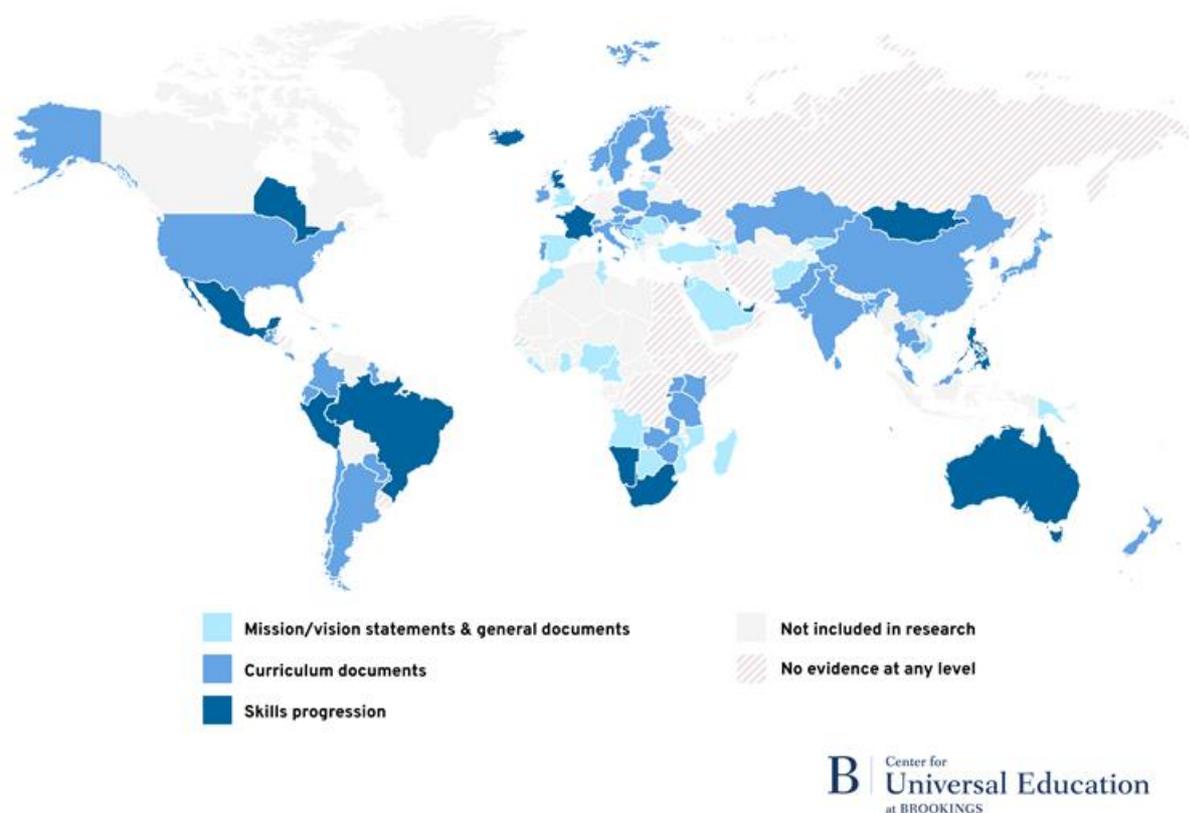
- Whether a country identified specific 21CS in their education mission, vision statements, or associated policy documents, such as national education plans;
- Whether a country identified specific 21CS within the curriculum;
- Whether a country described how 21CS develop and progress over time from basic to more complex forms and through the education levels.

Mission and vision statements, and associated national policy documents, portray a country’s educational aspirations for individual learners, or may reflect social and economic goals and national values. If specific competencies are identified, it suggests that governments value these. Within the curriculum, embedding the 21CS suggests intention to develop and teach these skills. But, beyond aspirations and intentions, countries need to move beyond surface recognition that 21CS increase in sophistication over the years of childhood and adolescence, to focus on description of their learning trajectories. This progress over time as students move through different education levels needs articulation in the same way as the more traditional skills such as literacy and numeracy do. This is essential for implementation—teaching of these skills in the classrooms—to occur.

Figure 1 shows findings from data across 152 countries, collected between May 2016 and August 2017. All data are sourced back to each country’s websites, or to multilateral organizations such as the Global Partnership in

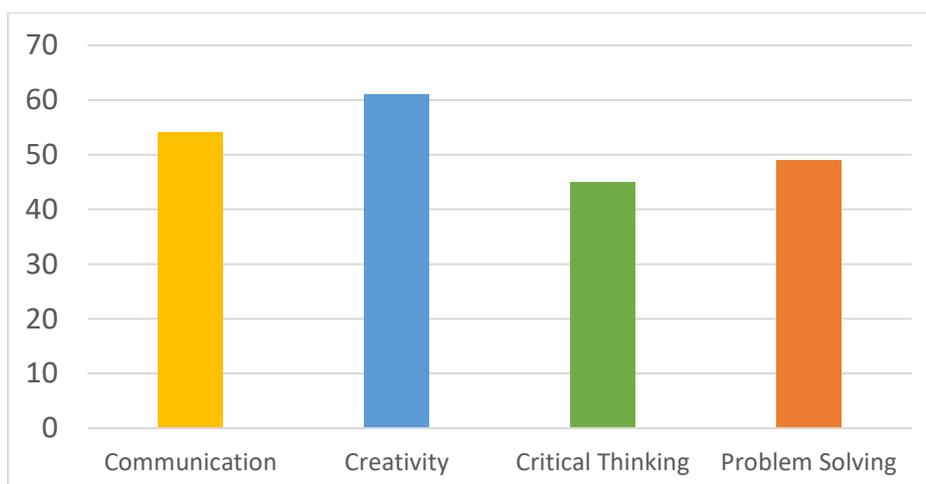
Education, World Bank, and UNESCO IBE (all of which publish the data with direct permission from relevant governments). Of the 152 countries included in the data, 53 (35 percent) countries—including Spain, Morocco, Madagascar, and Dominican Republic—identified specific skills in their mission or vision statements and/or general policy documents but not in their curricula (although in some countries, curricula were unavailable online). Fifty-eight (38 percent) countries—including Chile, Norway, India, New Zealand, and Zambia—have specific skills embedded within their curricular documents but do not show evidence of progressions of skills. Only 17 (11 percent) countries—including Australia, Mexico, Singapore, Iceland, and United Arab Emirates—provide descriptions of how skills develop and progress over time, and across different educational levels. In 25 (16 percent) countries—including Iran, Russia, Democratic Republic of Congo, and Egypt—there was no evidence of the presence of any of the three indicators described above.

Figure 1. Explicit identification of skills in national documents across 152 countries



Communication, creativity, critical thinking, and problem solving were the four most frequently identified skills within national policy documents of the 152 countries (Figure 2). Other skills identified include information technology, social, and entrepreneurship skills—indicating that countries are explicitly identifying a wide range of skills beyond the academic.

Limitations in the study should be considered. Only publicly available information was collected, meaning that countries may have national policy documents that identify 21CS, but do not make these available online to the public (e.g., national curriculum may not be online), and/or the information available online may not be up to date. Nevertheless, countries globally are moving toward an explicit focus on equipping students with a broad range of skills. The key issue, made clear by the data, concerns how implementation is actually planned and structured.

Figure 2. Four most frequently identified skills in national policy documents

Regional level shift

Another set of studies, which has focused at a regional level on the shift to competencies, provides similar findings. The Asia-Pacific Education Research Institutes Network (ERI-Net) coordinated by UNESCO's Education Bureau in Bangkok, undertook a series of three studies (UNESCO, 2015a, 2016a, 2016b) to explore how “transversal competencies” were being taken up by national systems. ERI-Net drew on a group of 10-11 countries, varying slightly across the studies. They included Australia, China (Shanghai; Beijing), Hong Kong SAR, India, Japan, Republic of Korea, Malaysia, Mongolia, Thailand, Viet Nam, and the Philippines. These studies were followed by two additional UNESCO studies (Care & Luo, 2016; Care, Vista & Kim, 2018) implemented through the Network on Education Quality Monitoring in the Asia-Pacific (NEQMAP), which explored assessment of transversal competencies, and drew on these countries as well as Cambodia, Nepal, and Pakistan.

For the first study, a tentative framework was drafted to capture countries' identification of the transversal competencies they valued. This first study highlighted the multiple rationales for the shift, taking into consideration personal, national, and global perspectives across economic, social, and humanitarian discourses. Notably, none of the participating countries mentioned acquisition of knowledge and “cognitive skills” as the primary goals of education. The most cited transversal competencies were critical, innovative, and reflective thinking; reasoned decision-making; communication; and collaboration (Figure 3). However, there were additional competencies that were unique to countries and which tended to be more strongly associated with attitudes, values, and ethics. The initial framework was modified accordingly for use in the subsequent studies (Table 3).

Figure 3. Number of countries and economies by skills and competencies

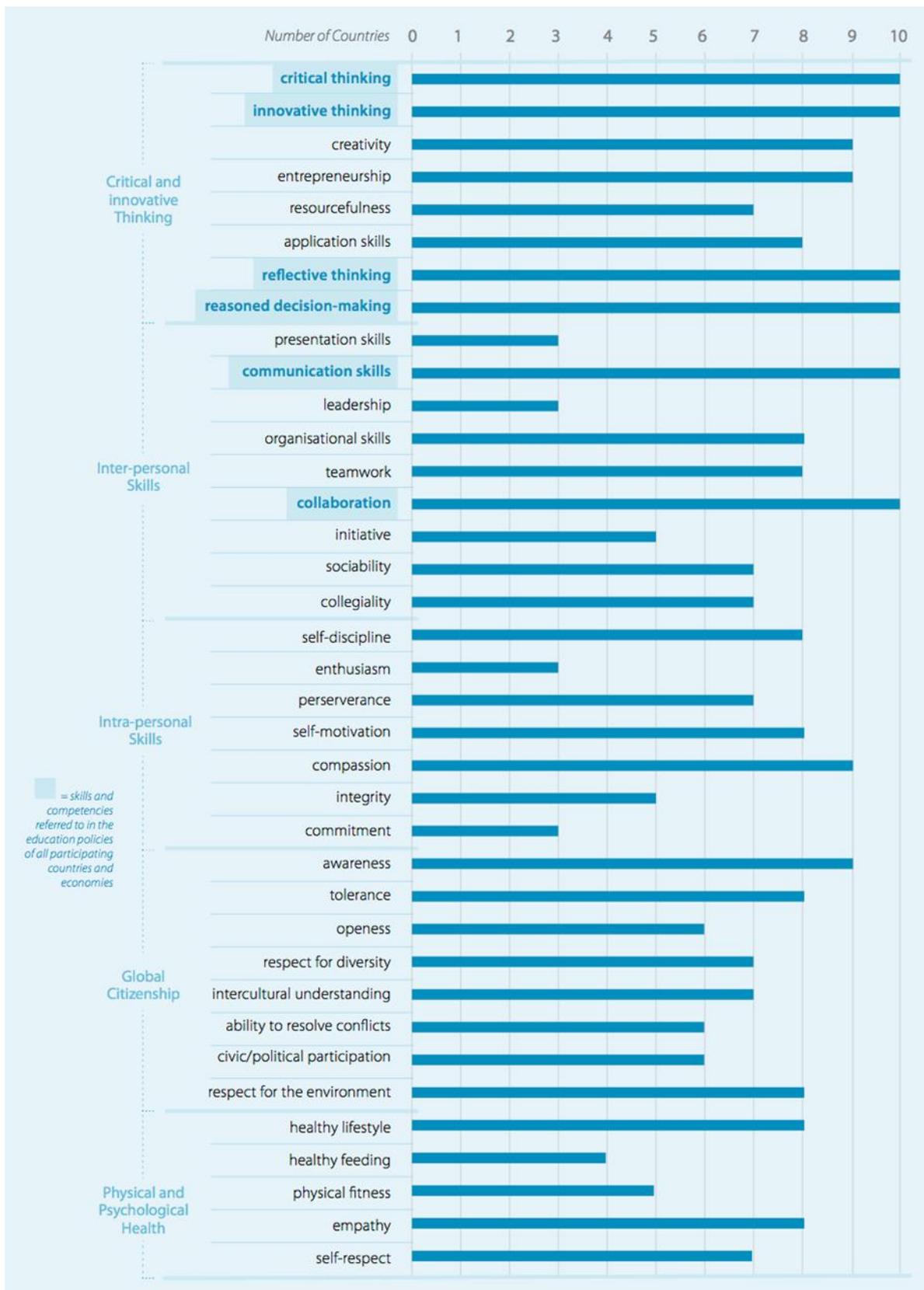


Figure adapted from UNESCO (2016a). *School and teaching practices for twenty-first century challenges: Lessons from the Asia-Pacific region (Phase II): Regional synthesis report*. Paris, France: UNESCO. Retrieved October 17, 2016 from <http://unesdoc.unesco.org/images/0024/002440/244022E.pdf>.

Table 3. ERI-Network definition of transversal competencies

Domains	Examples of key skills, competencies, values, and attitudes
1. Critical and innovative thinking	Creativity, entrepreneurship, resourcefulness, application skills, reflective thinking, reasoned decision-making
2. Interpersonal skills	Communication skills, organizational skills, teamwork, collaboration, sociability, collegiality, empathy, compassion
3. Intrapersonal skills	Self-discipline, ability to learn independently, flexibility and adaptability, self-awareness, perseverance, self-motivation, compassion, integrity, self-respect
4. Global citizenship	Awareness, tolerance, openness, responsibility, respect for diversity, ethical understanding, intercultural understanding, democratic participation, conflict resolution, respect for the environment, national identity, sense of belonging
5. Media and information literacy	Ability to obtain and analyse information through ICTs, ability to critically evaluate information and media content, ethical use of ICTs
6. Others (e.g., physical health, religions)	Appreciation of healthy lifestyle, respect for religious values

Table adapted from UNESCO (2016a). *School and teaching practices for twenty-first century challenges: Lessons from the Asia-Pacific region (Phase II): Regional synthesis report*. Paris, France: UNESCO. Retrieved October 17, 2016 from <http://unesdoc.unesco.org/images/0024/002440/244022E.pdf>.

Notwithstanding educators' strong endorsement of transversal competencies for holistic development (UNESCO, 2016b), valuing traditional pathways to academic excellence can pose challenges to implementation. The UNESCO studies (UNESCO, 2016b; Care & Luo, 2016) found that:

- Few professional development programs for teachers address the competencies, and those that do typically do not provide examples of classroom approaches to lesson planning and implementation;
- The competencies themselves are not well-defined or described, and how they are to be integrated within the curriculum is not clear;
- The tension between an examinations-focused system and whole child development is not addressed; and
- Assessment frameworks for the competencies, including standards and guidelines, are not available.

Additional sources of information about the global shift are reported by Siarova, Sternadel, and Mašidlauskaitė (2017) and Nusche (2016), drawing on country reports compiled on OECD countries and the European Commission's Recommendation on Key Competences in 2006, and by the Partnership for 21st Century Learning³ initiative reflecting movement in the U.S. landscape. Nusche concluded that most European Union states had changed their primary and secondary curricula to reflect the complex competencies outlined in the European Commission's Recommendation, as well as highlighted in UNESCO's four "Pillars of Education" –learning to know, learning to do, learning to be, and learning to live together (DeLors Report, 1996).

From these global and regional studies, there is strong evidence of a widespread shift in national education aspirations toward explicit recognition of the value of 21CS. Although some global and national communities might still argue which 21CS are more valued than others, and how they might be classified or categorized, there is no question that governments have taken up the challenge to ensure that future students will be better equipped to cope effectively and constructively with the demands of our world.

The power of intergovernmental organization proclamations, such as Education For All and the Sustainable Development Goals, has been well recognized. Major initiatives have been mobilized to support equal and quality

³ www.p21.org

access to education for all. It has become clear that access to quality education in literacy and numeracy alone is not sufficient, a recognition endorsed through Sustainable Development Goal 4.7:

“by 2030 ensure all learners acquire knowledge and skills needed to promote sustainable development, including among others through education for sustainable development and sustainable lifestyles, human rights, gender equality, promotion of a culture of peace and non-violence, global citizenship, and appreciation of cultural diversity and of culture’s contribution to sustainable development.” (UNESCO, 2015b)

Sustainable Development Goal 4 for quality education recognizes cognitive learning objectives, socio-emotional objectives, and behavioral learning objectives (UNESCO, 2017). Despite this recognition, a global mobilization of efforts to respond to the 21CS shift is non-existent, and individual countries struggle alone to plan the shift.

Technology, globalization, and demographic shifts are changing every aspect of society (Education Commission, 2016). The skills students needed by the time they completed schooling in previous generations, such as the ability to memorize facts, are no longer sufficient for the demands of the workforce and society in the 21st century. Skills such as critical thinking, collaboration, problem solving, and social skills are recognized by employers and educators alike as important 21st century skills (Child Trends, 2015). Education systems have acknowledged the shifting aims of education by increasingly including a broad range of skills in their policy and curriculum statements (Care et al., 2016). While this shift in policies is important, many countries have yet to implement these policies, and are still using 20th century teaching methods and assessments to teach 21CS (Care et al., 2018). This is holding back progress on learning for millions of children around the world.

WHAT IS NEEDED TO INCORPORATE A 21CS AGENDA INTO AN EDUCATION DELIVERY SYSTEM?

When education goals reflect qualitatively different learning domains, and are mirrored in curriculum, identification of appropriate pedagogical strategies and assessment methods aligned with these different goals must be considered. Much has been written about the need for alignment of education’s three delivery mechanisms – curriculum, pedagogy, and assessment. The publication of Wyse, Hayward, and Pandya’s handbooks on *Curriculum, Pedagogy, and Assessment* (2016) is one explicit acknowledgement of the importance of the links among them. The goals of education are what shapes the three delivery mechanisms. Although alignment is critical, most research tends to treat them independently. There are thousands of studies that focus on curriculum, instructional practices, or assessment but few that consider their interactions. Of course, organizationally, many education systems are the largest public institution in a given country, and are structured in many diverse departments, units, and centers. Because of organizational dysfunction and lack of communication within such huge bureaucracies, current aid efforts are building stronger institutions that are transparent and accountable, such as Australia’s Department of Foreign Affairs and Trade, Australia,⁴ and the World Bank.⁵ The separation of the different arms of education departments and ministries often follows not only the early childhood, elementary, and secondary divides, but also the education delivery divides of curriculum development and dissemination, learning delivery, and assessment. These institutional divides, although required for efficiency, frequently obstruct communications among these groups, and may be responsible for philosophy of education non-alignment. For a reform that seeks to implement qualitatively different education goals, like the 21CS shift, these divides are problematic.

Curriculum

Curricular reforms by countries in recent decades have mainly reflected a desire to move away from a relatively narrow set of cognitive skills and subject areas, responding formally to the desire to develop competencies that value the transformation and application of learning. Of course, there are also reforms that are locally stimulated,

⁴ <https://dfat.gov.au/geo/philippines/development-assistance/Pages/development-assistance-in-philippines.aspx>

⁵ <http://www.worldbank.org/en/topic/governance>

such as introduction of mother tongue-based education in many countries. The use of the term “competencies” as a focal point for learning goals has led to a focus on student learning outcomes. In turn, this has required curriculum reform that identifies evidence-based behaviors and skills to demonstrate competence. The European Center for the Development of Vocational Training (CEDEFOP, 2012) has made clear that the design and delivery of outcomes-oriented curricula statements influence the teaching and learning process. In particular, CEDEFOP found that outcomes-oriented curricula promote learner-centered pedagogies because outcomes are measures of student learning. Similarly, many countries have adopted an *Understanding by Design* (Wiggins & McTighe, 2005) approach to their curriculum revision. This provides for an approach to curriculum that develops the education goals and then identifies the necessary precursors in terms of curricular detail. This is one aspect of alignment—to ensure that curricular goals are met by the curriculum content.

Pedagogy

Learner-centered pedagogies have represented a shift from focus on teacher delivery of the curriculum to the learning of the student. The rise of learner-centered and constructivist pedagogies can be seen in part as stimulated by inputs from psychologists about the nature of learning, particularly through organizations such as the American Psychological Association in the 1990s (e.g., APA, 1997; Alexander & Murphy, 1998). Focus on the learner draws on information processing models, as well as acknowledgement of the role of motivation in learning (e.g., Eccles & Wigfield, 2002). Acceptance of information processing models in particular is relevant to the learning of 21CS. Deconstruction of complex skillsets into their contributing processes provides a learning pathway or roadmap to skills acquisition. Constructivist pedagogies also focus on students’ active engagement in learning and construction of knowledge—these approaches emphasize exploration and discovery, and focus on processes. A shift from traditional pedagogies that highlight content, structure, and information acquisition to learner-centered and constructivist pedagogies represents a necessary alignment with the nature of 21CS learning goals.

Assessment

Through the last century, education assessment was traditionally guided by curricular goals and by norm-referenced practices. The majority of testing in classrooms was designed to identify student acquisition of content, and the ability to demonstrate this acquisition. This summative function of assessment has provided the means by which students can gain access to additional years of education and training. A similar mode of assessment has been used for system evaluation purposes. In recent decades, the function of assessment has been extended to its use to guide teaching—a formative function (Black & William, 1998). This approach requires assessment to generate more descriptive information, and is well-aligned with a 21CS teaching and learning goal.

Knowledge of the learning goals

Alignment across the curriculum, pedagogy, and assessment assumes knowledge of the learning goals. However, for 21CS, learning goals explicated in the form of complex competencies present challenges such as identifying specific grade-level expectations for students. Subject-based curricula are familiar to educators, both the curriculum writers and teachers. Amendment of these curricula according to differing views about approaches to history, for example, or to stay up to date with scientific principles and knowledge, has been a relatively straightforward process. Where the learning goals are not so familiar, as in the case of 21CS, the process becomes problematic. Although it might be accepted that 21CS are important, there is no body of research curriculum writers can refer to that helps them identify what levels of 21CS should reasonably be expected at different stages in the formal education process. What level of problem solving capability or collaborative capabilities should be expected of students at different grades or ages, for example? How to establish education standards—what students should know and be able to do—in the absence of a long history of teaching and learning these competencies is a hurdle to curriculum development and assessment. Although the logic of standards-based systems is straightforward, it is dependent on comprehensive knowledge about the content and skills to be acquired.

“Standards define the knowledge and skills—competences—students are expected to have attained at different stages of education. The curriculum covers the objectives identified in standards, and student assessments and school evaluations focus on attainment of standards.” (Looney, 2011)

What are challenges to incorporating a 21CS agenda?

Historically, the majority of education systems have relied on a content and knowledge-based approach to teaching and learning, with the emphasis on memorizing and recalling facts. The focus has been on academic disciplines—language, mathematics, natural sciences, and history—together with art, music, and physical education. Although religious and humanist values over the centuries (before the call for universal education) might have called for morality, character, and social values, the reality is that in the elementary and lower-secondary school years, information acquisition has been valued. This has been demonstrated through tests and examinations practices, which tend to reward correct answers. Classroom pedagogical practices have been aligned with these goals—focused on presentation of knowledge to students, strategies to aid memorization and storage of knowledge, and exercises to aid in representation of the information on demand and in specific ways. As students progress through the education system, fluency in application of abstract, routine skills becomes more visible, but application of strategies in dynamic situations where standard approaches might not be viable remains invisible.

Facts are the constants in learning. Regardless of environment, the answer to a knowledge question remains the same. This has implications for how it can be taught and learned, and an important characteristic is the dichotomization of identification of true or false, correct, and incorrect. Transferable skills, however, are dynamic—they are fluid and take different forms. They are brought to bear to manipulate facts, and to build new knowledge, relationships, objects, and understandings. How skills manifest is therefore dependent on situational factors. These skills are about applications in different contexts and environments, and in response to different stimulus situations. The skills are about adaptability. These two very different sets of learning goals—content and skills—in education require different teaching and learning strategies to facilitate their acquisition. The increased valuing of 21CS therefore has major implications for education system infrastructure. Different learning domains demand different instructional approaches. History is not taught in the same way as drama; science is not taught in the same way as a sport. And of course, the different instructional strategies that are demanded by the different learning goals need to be aligned with different assessment strategies—some learning domains can be assessed appropriately using pen and paper tests, others require different approaches.

The Care et al. (2016) mapping of nations' vision statements and associated curricular information revealed that the vast majority of countries that identified 21CS in their statements did not describe notions of increasing competence. Although in part this is presumed due to some countries not uploading this information for public access, there remains a gap between statement of aspiration and evidence indicating that different levels of these skills have been recognized as a core component of the curricular shift. We offer three challenges to incorporating a 21CS agenda into education systems.

CHALLENGE 1: UNDERSTANDING THE NATURE OF 21CS

The lack of knowledge about 21CS domains is our first challenge to teaching and assessment of 21CS. This has implications for how curriculum, pedagogy, and assessment play out in the classroom. Identifying, defining, and describing a learning domain must precede exploration of how to integrate it into a learning system. The education system's learning goals are typically deconstructed into learning objectives—what is the expectation of what students will know or can do when they have completed a particular unit, module, or topic? The teacher then considers what kinds of activities might be used to teach the objectives and what will reinforce them. This identification of activities or tasks is an essential step to reveal whether the students have achieved the objectives, or where they are along the path to achievement. To be able to implement this process, in-depth knowledge of the learning domain is a pre-requisite—for the teacher, the assessment developer, and the curriculum designer.

How can we not know about a learning domain? The answer lies in the fact that although we can identify a skill when we see it—for example a person typing—the components that underlie this are not seen. We have “constructed” a belief that there are underlying mechanisms which account for display of skills in different ways, situations, and levels of quality. Defining the construct is the first order of business. Is typing defined by being able to type 300 words a minute? What about the accuracy of that typing, and the grammatical correctness of the typing? These are all part of typing skills. Our business is to hypothesize the full nature of a skill (or construct), linking empirical evidence to it by recognizing behaviors demonstrated across multiple occasions and situations. Although there is substantial research evidence around identifying the subskills underlying particular 21CS: 1) there is not an adequate body of research that focuses on how these subskills interact and develop (e.g., Reimers & Chung, 2016); and 2) much of the research lies in academic literature focused on psychological issues rather than within the context of education systems (Pellegrino & Hilton, 2012).

Notwithstanding that education research linking theories of development to teaching and learning of 21CS in the classroom might be scarce, we describe some general understandings—informed by theory and research across a variety of skills—of three 21CS frequently identified by national education systems, including collaboration, critical thinking, and problem solving (Care & Kim, 2018).

Collaboration

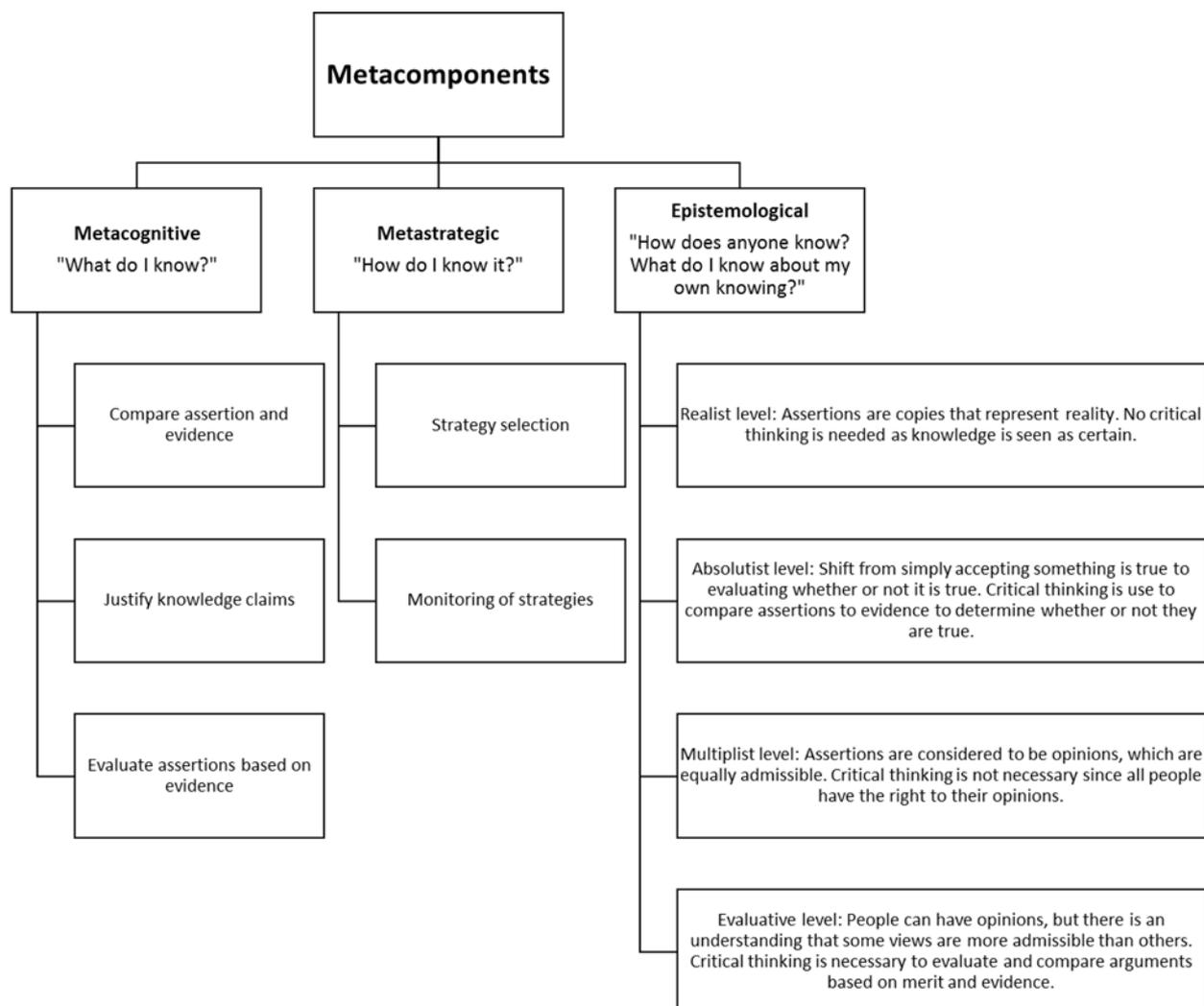
Collaboration has been described as a learning skill (Partnership for 21st Century Learning, 2015), an interpersonal skill (National Research Council, 2011), a way of working (ATC21S, 2014); and a way to learn about how individuals think (Enyedy & Stevens, 2014). There is agreement that collaboration is a key skill that improves student learning in school, performance at work, and throughout one's personal life (Lai & Viering, 2012). Collaboration builds on effective communication skills by inserting these into interpersonal situations (Greenstein, 2012). Collaboration occurs when meeting a goal requires more than what any one individual is able to manage alone and needs to pool resources with others. Collaboration therefore involves a construction of shared meaning that involves an iterative cycle of sharing, confirming, repairing conceptions, and managing the task at hand (Roschelle & Teasley, 1995). Primary dimensions of collaboration are social interdependence, interpersonal skills, and task-related processes. Specifically, collaboration includes knowledge, skills, and attitudes (Binkley et al., 2012; Child & Shaw, 2016; Greenstein, 2012; Partnership for 21st Century Learning, 2015) such as:

- Interacting effectively with others and having meaningful conversations;
- Knowing when it is appropriate to listen or to speak (social regulation);
- Working effectively in diverse teams (e.g., conflict resolution and team management);
- Introducing new ideas and sharing of resources;
- Exercising flexibility and willingness to be helpful in making necessary compromises to accomplish a common goal;
- Assuming shared responsibility for team work;
- Perspective taking; and
- Valuing the individual contributions made by each team member.

Critical thinking

Critical thinking is intentional, goal-directed, and reflective (Lewis & Smith, 1993). It comprises mental processes, strategies, and representations that are used to evaluate, make judgments, and learn new concepts (Sternberg, 1986). Critical thinking also involves evaluating the thinking process (Halpern, 1998). Critical thinking is not a developmental phenomenon where a sequence of competencies emerges at certain ages, but rather a progression in which the constituent processes become more coordinated and sophisticated over time (Kuhn, 1989). In general, critical thinking involves both non-executive and executive processes, but much of the focus has been on metacomponents—or the higher order executive processes used to plan, monitor, and evaluate (Sternberg, 1986). These metacomponents can be further broken down into three broad categories: metacognitive, metastrategic, and epistemological (Kuhn, 1999). Metacognitive knowing has to do with declarative knowledge—"what do I know?" Metastrategic knowing has to do with procedural knowing—"how do I know it?" Epistemological knowing has to do with a broader understanding of knowing and knowledge—"how does anyone know?" and "what do I know about my own knowing?" Figure 4 shows the three categories of metacomponents and their subcomponents as they relate to critical thinking processes.

Figure 4. Three categories of metacomponents of critical thinking (adapted from Kuhn, 1998).



Specifically, critical thinking consists of a complex set (Binkley et al., 2012; Willingham, 2007) of attitudes and dispositions, and thinking skills, such as:

- Willingness to engage in and persist at a complex task;
- Flexibility and open-mindedness;
- Willingness to abandon non-productive strategies in an attempt to self-correct;
- Verbal reasoning;
- Argument analysis;
- Hypothesizing and undertaking a systematic process to check evidence;
- Estimating likelihood and uncertainty; and
- Decision-making.

Problem solving

Real-world problems and goals are rarely well identified, and the information relevant to them is equally unclear. Problem solving is the basic cognitive process for identifying the nature of problems, assessing different options, and making informed choices when there is no clear or routine solution (Greenstein, 2012; Mayer, 2013). Problem solving requires acquiring and evaluating information to solve different kinds of complex, and sometimes non-familiar problems in both conventional and innovative ways. There are many frameworks that identify approaches to problem solving—the OECD (2013a) approach was mentioned earlier in this paper. Most frameworks include processes such as describing the problem, gathering information, and considering multiple perspectives; identifying gaps, generating hypotheses, and developing procedures to test these evaluating outcomes; and reviewing. There are many similarities between critical thinking and problem solving, with the most salient difference being the

existence of a problem as the structuring element for problem solving. This goal directedness provides an organizing feature for the sets of processes that are brought to bear.

Greenstein (2012), who provides guidance in assessment of 21CS for teachers, outlines steps in problem solving useful for classroom implementation. Each step includes several components which enable five main processes:

1. Understand the problem
2. Brainstorm all possible solutions
3. Devise a plan
4. Carry out the plan
5. Evaluate the results

Despite the fact that these domains are known and relatively well-described, there is still a lack of understanding as to how they manifest at different levels of competency. In other words, unlike literacy and numeracy, learning progressions of these skills, which provide comprehensive descriptions of how the skills develop from basic to sophisticated forms, are not in evidence. This has major implications for teaching and assessment, and brings us to the second challenge of incorporating a 21CS agenda as a major challenge for countries around the world.

CHALLENGE 2: DEVELOPING LEARNING PROGRESSIONS FOR 21CS

Development of skills implies a process through which increasingly complex applications of skills are gradually acquired and demonstrated. Some curricular domains present content in successive grade levels that are not heavily contingent on previous grade level studies—for example, Indonesian history versus the history of Yugoslavia (although ideally there is a gradual acquisition of historiography through the grades). In the case of 21CS, acquisition of increasing competency presumes acquisition of sets of processes that build on each other.

Historical approaches to understanding skills developments include those of Bloom and colleagues, and Biggs and colleagues. One widely used approach to *cognitive* skills development, created by Bloom and Krathwohl (1956), has been used since the mid-1950s. Later revised by Anderson and Krathwohl (2001), and with inclusion of the *affective* domain, these taxonomies have heavily influenced teaching and learning. Rather than differentiating between successively more sophisticated skills, Bloom's taxonomy defines complexity through identification of different actual skills, each of which is viewed as successively more difficult, and builds on previous acquisition. This approach is associated with concepts of developmental stages rather than with a gradual and progressive acquisition of increasingly sophisticated competencies. A somewhat similar approach that identifies sets of processes as denoting different developmental ways of thinking, as opposed to gradual development and sophistication of contributing dimensions, is the "Structure of the Observed Learning Outcome" (SOLO; Biggs & Collis, 1982) taxonomy. Conceived as a way of classifying learning outcomes in terms of their complexity, the taxonomy has been used to design assessment tasks, and adopted as an approach to the identification of higher order learning. Primarily a cognitive task-approach strategy, SOLO is not dissimilar from Bloom's (1956) approach in focusing on cognitive strategies—as opposed to skills development.

Why are learning progressions important for 21CS agenda?

The use of developmental continua, or learning progressions, has become increasingly evident over the past 20 years. They are used in many countries to provide guidelines for teachers and students to understand the processes of learning in particular domains by identifying appropriate competency levels and reasonable aspirations for students at different grade or age levels. Not only do they underlie curricula, but they are used as frameworks to locate assessment tasks that sample student skills at increasingly difficult levels of complexity or sophistication. Setting standards and developing assessment tasks, however, requires evidence of what can reasonably be achieved at these levels.

Popham (2007) provides a specific example:

"A learning progression is a carefully sequenced set of building blocks that students must master en route to mastering a more distant curricular aim. These building blocks consist of subskills and bodies of enabling knowledge. To illustrate, if a curricular aim calls for students to become skilled writers of persuasive essays, a learning progression for this aim might include a subskill that requires students to be able to craft supporting arguments for a given position. To master this subskill, students might need bodies of knowledge that enable them to understand certain spelling and punctuation rules or to use specific vocabulary—for example, *sound*, *valid*, and *justifiable*—associated with argumentation. The complete learning progression for a persuasive writing skill might include a half dozen subskills." (p. 83)

Although such evidence is available for traditional learning domains, we do not have similar evidence around what is reasonable for 21CS. This, of course, has implications for curriculum, pedagogy, and assessment of 21CS. Standards and curricula are often prioritized in educational documentation, but it is the students' progress toward the standards and meeting the learning goals that is important. While learning standards describe what students should have learned by a specific stage in their education, learning progressions provide a way of thinking about how students can meet those standards, with appropriate instruction. Heritage (2008) refers to “sequence,” “continuity,” and “coherence” as implicit in the notion of learning progression. In other words, empirical learning progressions which are based on evidence describe the sequence of successively more complex processes and their demonstrations that contribute to mastering and applying a particular skill. In this way, learning progressions reflect typical trajectories of specified learning domains that describe how skills or concepts might be demonstrated, both in their early forms and in increasingly advanced forms.

Furthermore, in order to design and develop assessment tools, it is necessary to identify the behaviors that relate to the learning domain of interest. Similarly, teachers need to be able to identify behaviors that indicate differing levels of skill or knowledge if they are to teach at the appropriate levels of challenge. This means that they need to have access to descriptions of how skills progress over time so that they can design classroom tasks that are within the zone of proximal development for their students. The zone of proximal development is a concept created by Vygotsky (1978) that indicates when a skill can be used with the help of another person but not yet independently so it indicates a “readiness to learn.” Children need to be taught at a level of difficulty where they are moving from the familiar to the unfamiliar, or the known to the unknown, from guided learning to independence. Using this pedagogical idea, teachers can support, or “scaffold” the learning of their students.

The identification of a general sequence in learning comes from multiple sources. From what we observe in early skills development, in children around us and from our own experience, we hypothesize a sequence in development or learning. In order to confirm that what we observe is a sequence true of most, we gather data from large and varied populations. Then, we confirm or adjust our understanding of the sequence. A scientific approach includes the conscious process of hypothesizing the phenomenon, followed by collection of data to identify the degree of fit to the theory. In practice, many learning progressions have emerged from a more historical approach due to common knowledge about learning sequences. Where particular learning domains have not been highly valued, as with 21CS, we lack some of this common knowledge, and the scientific tradition in confirming its validity has not been widely followed. Quite apart from developing learning progressions, however, is the challenge of designing and developing assessments of 21CS that are able to capture the social and cognitive capacities that underlie constructs such as collaboration, critical thinking, and problem solving.

CHALLENGE 3: DESIGNING APPROPRIATE AND AUTHENTIC ASSESSMENT OF 21CS

Assessment in formal education systems is primarily undertaken through written work. Although there has been some movement toward diverse forms of assessment, exemplified by open-ended activities such as projects, presentations, and portfolios, the majority of classroom assessment and national or large-scale examinations remains relatively narrow in scope and form. A concern for 21CS is the degree to which traditional forms of assessment are adequate for the capture and reporting of students' proficiencies. With curricula typically the first focus of an education reform, assessment tends to lag behind. The risk is therefore lack of alignment between curriculum and assessment, both at classroom and national levels (Care, 2018b), leading to potential confusion throughout the system about what is actually valued.

The overriding question that guides assessment is “what do we want to know?” The answer to this question directs all subsidiary questions—concerning what is assessed, how it is assessed, and how it is reported and used. What do teachers, governments, non-government organizations, and researchers each want or need to know? Putting aside the matter of certification, assessment can provide information to inform improvement. The latter can be experienced directly by students in their current classrooms, or indirectly through improvements emanating from system responses. Typically, assessments are categorized across classroom, national, and international levels. And these different destinations for student assessment results influence the style and mode of the assessment.

Major functions of assessment are:

- to inform teaching and learning at the individual student level;
- to provide information from large-scale assessment programs for use at system evaluation, accountability, and policy level; and

- to use data to “select out,” or qualify, students for further education, training, and development opportunities.

In the case of 21CS, none of these major functions are currently widely fulfilled. As discussed, our lack of knowledge of 21CS learning domains, and the very nature of those domains pose major challenges regarding how to integrate the skills into curriculum, how to teach through appropriate pedagogical strategies, and how to assess them.

The question “what do we want to know?” also directs us to think about the nature of what is to be assessed. The great challenge in measurement of most human social and cognitive capacities is that the underlying ability that predicts how each of us behave and can perform is itself unseen. This underlying ability, or construct, is what we need to understand. Because we cannot measure the generic ability directly, constructs such as problem solving, critical thinking, collaboration, and communication, are measured through tools that are designed to capture indicators of these skills. The issue, however, is that the majority of current assessment practices rely on traditional formats, such as multiple choice, true/false, close-ended responses, and rating scales that limit what can be captured. These formats may be appropriate for evaluating rote knowledge, such as a basic numeracy question like “what is the area of the rectangle?” or perceptions, such as self-reporting on statements like “I am a good communicator.” These types of questions do not capture the processes that underlie the complex skillsets, which need to be stimulated and elicited, modeled and demonstrated, and applied to novel situations. One way to approach this challenge is to focus on the concept of authenticity—the degree to which assessment tasks resemble real-world tasks along a number of dimensions and provides evidence of what students can do (e.g., Mueller, 2005; Whitlock & Nanavati, 2013; Wiggins, 1989).

The premise for good assessment is that it captures valid indicators of the target construct. A “valid indicator” can be a behavior, a performance, or a product that is believed to “indicate” or demonstrate the underlying competency that we believe exists. To stimulate the behaviors from which these indicators can be captured, the assessment design must mirror the real-life demands of a situation that would provoke those behaviours (Care et al., 2016). This authenticity in terms of the stimulus situation does not guarantee validity of the assessment for the particular purpose for which it is being used, but does contribute to it. Since many 21CS are seen only through behaviors, as opposed to something that can be written, the challenge is both to capture the behavior and accurately interpret it as demonstrating the 21CS of interest. The direct link between competence and indicator that can be seen in a written task that is designed to assess writing ability for example, is not possible in this more challenging 21CS environment. Pellegrino, DiBello, and Goldman (2016) speak to this point: “an assessment is a tool designed to observe students' behavior and produce data that can be used to draw reasonable inferences about what students know” (p. 5), although in the case of 21CS, it must be extended to ‘what students can do, say and make.’ Gulikers, Bastiaens, and Kirschner (2004) take the position that level of authenticity is determined by degree of similarity to the criterion situation.

Specifically, Gulikers et al. (2004) five dimensions of authentic assessment are:

1. An authentic task presents as a set of activities that emulate professional practice;
2. The physical context reflects the way the competencies will be applied in professional practice;
3. The social processes (if these are relevant) reflect those applied in the real situation;
4. The product or performance mirrors a real life one, permits inferences about the underlying construct, includes multiple indicators, and is available to others for review; and
5. Criteria identify what is valued, and standards indicate levels of expected performance.

These dimensions are the ideal, and tasks can vary in degree of authenticity within and across these five dimensions. In terms of an education context, this needs to be mediated by the fact that early acquisition and demonstration of skills might not look strongly similar to the mature skills that need to be brought to bear to a real-world scenario. Assessment tasks need to be designed that can capture the earlier manifestations as well as the mature skills—hence the need to have a well-hypothesized learning trajectory, or progression, of the skill.

The descriptions of problem solving, critical thinking and collaboration⁶ earlier in this paper provide some guidelines (Table 4) to design of tasks for 21CS such as these.

⁶ For collaboration guidelines include suggestions modified from Child & Shaw, 2016

Table 4. Assessment strategies for problem solving, critical thinking, and collaboration

21CS	Recommended task strategies
Problem solving	<ul style="list-style-type: none"> • Tasks that require students to <i>take different perspectives</i> on an issue than the one presented; this is in contrast to where concepts or problems tend to be conceptualized from just one angle. • Tasks that require students to <i>identify the types of information</i> needed to bring to a particular problem scenario; students need to understand the “problem space” and what objects, individuals, or concepts are relevant to that space. • Tasks that challenge students to explore <i>multiple routes</i> to problem solution; this requires students to hypothesize, to check, to iterate, and to review.
Critical thinking	<ul style="list-style-type: none"> • Tasks that require students to identify <i>similarity of structural features</i>. Students are stimulated to organize the information they have, classify or categorize it, and evaluate common characteristics. • Tasks that require <i>representing</i> concepts or arguments through alternate media such as diagrams and text. Students need to understand the concepts or arguments in order to conceive of different ways of communicating that information. • Tasks that require students to <i>identify missing or alternative</i> components and to provide rationale. This prompts the student to think logically as well as hypothesize associations between components within a model.
Collaboration	<ul style="list-style-type: none"> • <i>Complexity and lack of definition</i> - tasks should be designed so that they are open-ended with more than one plausible solution. They should be created such that a single member of a group is unable to complete independently, in order to ensure that group members must share the cognitive and social load. • <i>Group dynamics prompt negotiation</i> - group characteristics, such as the size and composition of a group (Dillenbourg, 1999) may influence how students interact and engage in the collaborative process. • <i>Group motivation</i> - the task and group dynamic should be motivating. This is related to social interdependence, where group members may be more likely to be accountable, productive, and collaborative when group members want to work together.

These task types can relatively easily be integrated into assessments of key learning areas, whether mathematics, geography, or graphic design, for example. An important aspect of such use, however, is how to reward innovative responses that demonstrate skills proficiency. To do so would require application of some of the scoring mechanisms used for creativity, or scoring according to rubrics that reflect valuing of divergent responses. Since display of these skills may consist of exploratory behaviors rather than focus only on correct solutions, either open-ended response items or very well-structured and sophisticated tasks are most suited to their capture, recording, coding, and scoring. Evaluating the quality of responses to open-ended questions is the most technically difficult challenge. Analyses of OECD large-scale data have been exploring the characteristics of student problem solving responses in order to identify viable alternatives to correct solutions as measures of achievement in, for example, problem solving (e.g., Ramalingam & Adams, 2018). Another technical challenge lies in identification of individual competencies in group-based tasks, since individual demonstration may be dependent on the level of competence of other group members and the quality of the interaction. Although in a research context such elements can be controlled for, use in a real classroom climate makes evaluating individual performance difficult. Equally, attempts to attribute group level evaluations can lead to inaccurate evaluation of the individual. Solutions include the design of tasks that isolate particular types of contributions/activities to particular individuals, and rotation of these individuals across the different types of contributions/activities.

The vast majority of large-scale tests, at national, regional, and international levels tend to use formats that are easy to code and score at large-scale, and in turn, these formats are not those that provide for capture of non-

routine skills, such as those we are seeking. Nusche (2016) presented information drawn from an OECD review of testing formats used in the national assessments of 27 countries. The most common formats were multiple choice, closed format short-answer questions, and open-ended writing tasks and calculations—with very few that relied on performance or orally-based tasks, in the main due to the exigencies and economies of scoring. Of course, the vast majority of these national assessments are targeting literacy, numeracy, and key learning domain areas—all of which have traditionally and predominantly used closed-format responses effectively. As the world turns to assessment of 21CS, the issue not only of item format, but also how to capture processes, confronts us.

Rather than restricting the discussion to the theoretical, we present a small number of tools designed to assess the three competencies already discussed in this paper—problem solving, critical thinking, and collaboration—to illustrate the concept of authenticity. Based on the work undertaken by Care and Kim (2018) and Kim and Care (2017), we examine the degree of authenticity of tools using Gulikers et al.'s (2004) general framework, keeping in mind the challenges to 21CS assessment that we have identified. Table 5 shows the assessment tools along the dimensions of authenticity. For a more extensive discussion, please see Care and Kim (2018) and Kim and Care (2017).

Problem solving is perhaps the most thoroughly researched of those generalizable skills that are explicitly valued in our education systems. The four sets of processes: exploring and understanding; representing and formulating; planning and executing; and monitoring and reflecting represent a very strong theoretical and practical approach to skills. To this extent, for problem solving, the first challenge—lack of familiarity with 21CS—is not as big of a problem as for other 21CS. Although much of the research has focussed on problem solving in mathematics and sciences, most recent large-scale assessment of problem solving has extended its reach, for example with PISA 2012 problem solving assessment tasks (OECD, 2013b). Through everyday situations, across a wide range of contexts, the PISA 2012 Problem Solving computer-delivered assessment is designed to assess “an individual’s capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious” (OECD, 2013b, p. 4).

One item from the PISA 2012 field trial, MP3 Player, is examined in terms of the five dimensions of authenticity. This item is an interactive problem situation, in which students are able to explore, control, hypothesize, test, and analyze within a simulated environment. Specifically, students are told that they need to find out how to work an MP3 Player by interacting with the device. Then, they are asked to decide whether statements (e.g., “you have to set the volume before you can set the bass level”) about the MP3 player are true or false. These statements are intended to scaffold the students in exploring the system, so that they can perform the next task (e.g., “Set the MP3 player to Rock, Volume 4, Bass 2. Do this using as few click as possible. There is no RESET button.”) In the third item, students are provided four screens and have to choose which screen shows the MP3 player working properly. Finally, students are asked to describe how they could change the way the MP3 player works, so that that there is no need to have a particular button.

Table 5 summarizes the degree of authenticity in each of the five dimensions. In general, the assessment task resembles a situation that students could encounter in the real-world—one where they would have to figure out how to work an unfamiliar equipment or machine—and provides opportunities for unrestricted exploration, as one might have in similar situations. However, some aspects of the task, such as the assessment result (e.g., multiple choice and true/false) may not necessarily mirror a situation one might encounter in the real-world. Notably, quite apart from identifying final solutions, the benefit of the online platform in PISA 2012 (and also in PISA 2015 and 2018) is that the online data capture provides a repository of all actions taken by the student such that some of the cognitive processes can be analyzed in order to understand better how individuals go about problem solving.

Table 5. Examining 21CS Assessment Tools across five dimensions of authenticity

Dimensions of Authenticity ⁷	Description	PISA 2012 Problem Solving Assessment ⁸ “MP3 Player”	SimScientists Grassland-Food Web ⁹	PISA 2015 Collaborative Problem-Solving Assessment ¹⁰ “The Aquarium”
Assessment Task	Tasks resemble those carried out in the real-world with respect to the integration of knowledge, skills, and attitudes; and task complexity, relevance, and ownership.	Tasks require student to figure out how the MP3 player works. The underlying processes of exploring and understanding, planning and executing, representing and formulating, and monitoring and reflecting are needed to solve the problem.	Student is guided through a simulated environment and learns about the food web. Processes such as drawing warranted conclusions, evaluating information, and integrating knowledge are needed to complete the task.	Tasks require the student and a computer agent to work together to find the optimal conditions for fish living in an aquarium. The processes of exploring and understanding, representing and formulating, planning and executing, consensus-building, and decision-making are needed to solve the problem.
Physical or Virtual Context	Reflects the way competencies will be used in professional practice. Number and kinds of resources and amount of time should also resemble those available in criterion situation.	Tasks are computer-based and allow the student to interact within the simulated environment and explore the various components of the MP3 player. This is scaffolded with statements, but the amount of time for exploration is unrestricted.	Tasks are computer-based and allow the student to interact within the simulated environment. Some aspects are animated to show dynamic relationships among structures, mechanisms, and behaviors within systems. There is scaffolding and individualized coaching when necessary, until the student completes the task correctly. There is no time restriction as this is intended to be a learning experience.	Tasks are computer-based and allow the student and the computer agent to each manipulate three variables within the problem space and “chat” to collaborate on solving the problem. This unit has a time constraint of 15 minutes.
Social Context	Social processes (if relevant) resemble the social processes in an equivalent situation in reality.	The interaction with the problem space is done individually.	The task is completed individually.	The student interacts with a computer agent that simulates various scenarios and is scaffolded such that if the 'correct' response is not given, the agent provides multiple prompts to help direct the conversation. The student has to engage in social processes necessary to collaborate on completing the task.

⁷ Adapted from Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86.

⁸ Based on MP3 Player item from PISA 2012 filed trial in *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Retrieved from: https://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf

⁹ Based on the Grassland-Food Web example found in Ecosystems Try It on <http://simscientists.org/MiniSite/index.html>

¹⁰ Based on The Aquarium unit (p. 51-64) in *PISA 2017 Collaborative Problem Solving Framework*. Retrieved from: <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>

Assessment Result	Product or performance mirrors real-life scenario; permits making valid inferences about the underlying construct; includes multiple indicators; and is available to others for review.	Scores are based on 4 questions: true/false (no partial credit); interactive (partial credit depending on number of steps taken to achieve goal); multiple choice (no partial credit); and constructed response (expert scoring required).	Immediate feedback is generated by the computer for response that can be automatically scored and gradual coaching is conducted as students are guided through multiple trials to respond appropriately to the task.	Scores are based on 24 multiple choice chat response items and are given a score ranging from 0-2 points.
Criteria	Criteria should be set and made explicit and transparent to learners beforehand. Criteria should be related to a realistic outcome.	Results are summarized on a single composite problem-solving scale that are divided into six levels of proficiency, which characterize student performance. These are not made explicit beforehand.	Algorithms assess the level of a student response and tailored feedback is provided. The system takes into account not only the responses but also interaction with the simulated system, such as time taken to respond, variables manipulated, and the order and number of trials run in the system. These indicators are used to assess student's abilities. Automated scores and levels of coaching received are produced to report back to teacher and student to show what students know and are able to do. Example: "You have shown that you understand the role of producers but need to work on understanding the role of consumers."	Scores for each item are aggregated to provide an overall scale for CPS competency. This overall proficiency score is then mapped onto a four-level proficiency scale based on progressions.

The literature around critical thinking is extensive and has roots in both philosophy (with a focus on the qualities and characteristics of critical thinking, such as being inquisitive or open-minded) and psychology (with a focus on the behaviors or actions involved in thinking critically). For teaching, learning, and assessment purposes, identifying the behavioral outputs and markers along a learning progression is necessary, rather than focusing on the ideal qualities of a critical thinker. Critical thinking goes beyond acquiring and retaining information or having a set of thinking skills—it involves acting in novel situations to know what information to use and when to evaluate arguments, draw warranted conclusions, and make sound judgments (Halpern, 1998). The challenge with critical thinking, as with most complex skills, is the ability to transfer the (component) skills beyond the contexts of learning to novel situations (Willingham, 2007). Therefore, the tasks used to teach and assess critical thinking need to be true to this concept, and mirror, as much as possible, real-world thinking tasks and environments.

There are assessments of critical thinking, such as Educate INSIGHT Reasoning Skills assessment series for Grades 3-5, 6-8, and 9-12, which according to the developers "is specifically designed to measure the critical thinking skills" in analysis, evaluation, inference, induction, and deduction, as well as numeracy and quantitative reasoning (Insight Assessment, 2018). The questions are in a traditional multiple-choice format where students are asked to apply critical thinking skills to everyday topic areas. However, whether these items are able to assess and capture the complex nature of critical thinking skills, where students have to engage in the processes that are needed in real-life situations that are new and unpredictable, is a question. The format of the items may limit the capacity to measure both aspects of critical thinking, as well as its entirety (Soland, Hamilton, & Stetcher, 2013).

More recently, technology platforms have been used to assess critical thinking skills, which frequently occur within domain-specific knowledge, such as science. For example, SimScientists program, although not solely designed as an assessment tool, engages critical thinking processes. Embedded in Science, Technology, Engineering and Mathematics (STEM) environments, rich tasks take advantage of technology platforms to help students “reason with and about models of science systems” and develop and demonstrate inquiry skills. SimScientists computer-based modules are designed to selectively focus on integration of knowledge and application of science practices through a progression of tasks in authentic environments in which students are stimulated to explore, interrogate, and manipulate systems and cycles that are similar to those students may encounter in the real-world (Quellmalz, Buckley, Silbergliitt, Loveland, & Brenner, 2016).

All tasks:

- 1) are based on integrated knowledge about dynamic relationships among structures, mechanisms, and behaviors within science systems;
- 2) incorporate authentic, problem-driven inquiry practices;
- 3) provide scaffolding through individualized feedback and customized coaching toward mastery of task;
- 4) metacognitive self-assessment and reflection; and
- 5) comprise simulations that offer multiple opportunities for students to interact with science systems and respond (Quellmalz, Timms, & Buckley, 2009).

One example is Grassland-Food Web (WestEd, 2015), which provides information about the food web in grassland areas. The series of screens have both animated and non-animated features that allow a student to explore the system. It also includes items that ask students to respond “yes” or “no” to statements, such as “Only producers serve as food for other organisms.” If answers are not correct, immediate feedback is provided in the form of “You have not identified the correct role of a producer in the ecosystem. Please try again.” Another incorrect response results in scaffolded statements, so that the student learns from mistakes and responds correctly. The student cannot move onto the next task item until all responses are correct. Other items include making a food web diagram by drawing arrows, which also includes scaffolds for those who need additional support, open-ended questions that allow the student to compare his answer to the master answer and self-reflect. Table 5 examines authenticity of this task.

Compared to problem solving and critical thinking, the research around collaboration—working together to achieve shared goals—is relatively new in terms of how it develops and its underlying processes. Much of the research has focused on the foundations of collaboration, such as forming joint goals and joint attention during social interactions via non-verbal and verbal communications (Tomasello & Hamann, 2012), monitoring one’s behavior during a joint task (Warneken, Graefenhain, & Tomasello, 2012), and a shift from perspective as member of a group to a group identity.

Despite a focus on collaboration in recent years, there are very few assessments, if any, that respond to the three challenges. Some assessments of collaboration assess the skills within complex skillsets such as collaborative problem solving. One example is the PISA 2015 Collaborative Problem Solving (CPS) assessment, which defines collaborative problem solving as “the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills, and efforts to reach that solution” (OECD, 2017, p. 6).

One example unit is The Aquarium, which is examined in terms of the dimensions of authenticity (Table 5). Here, the student collaborates with a computer agent to find the optimal conditions for fish living in an aquarium by controlling variables within the simulated environment. In the first task, the student and the computer agent discuss how to solve the problem using pre-determined chat options. Then, through a series of five trials, the student and computer agent manipulate variables and determine which variables each can control, activate plans, and determine the best conditions for the fish to live in the aquarium. In the final task, the student and computer agent reflect on their collaborative work. Although the task resembles real-life situations in some aspects (e.g., the need to engage in social processes to solve a problem), much of the task does not mirror authentic scenarios. For example, the task items are a series of multiple choice questions. In fact, none of the task items requires the student to generate his or her own response, although it is possible to gain insight into the proficiency level of the student based on the response choices. In addition, the scoring of items based on credit, partial credit, and no

credit depending on the response is unrealistic in terms of the feedback received when collaborating in real-life tasks; however, the descriptions of the levels of proficiency that corresponds to the aggregated score provide clear standards and explanations that could be based on real-life situations.

Each of the three example assessments discussed above highlight different functions of assessment—and all through relatively “high-tech” environments for reasonably large-scale use. PISA 2012 Problem Solving and PISA 2015 Collaborative Problem-Solving assessments were designed to inform national education system accountability and benchmarking purposes, with little immediate benefit to the students who contributed the data. The SimScientists program stimulates development and learning in science environments through exploration of simulations, generating immediate learning outcomes for students. The online task platforms in each of the three tools are a move away from the traditional paper and pencil tests in capturing, at least in part, what students are able to do through interactions with others and within the task environments.

Education functions as a socialization mechanism. Consistent with Education For All, Education 2030 Agenda (UNESCO, 2015), and Sustainable Development Goals (United Nations, 2016) initiatives, 21st century education also functions as an integration mechanism, with children from relatively diverse backgrounds learning together. Unlike the 20th century role played by educational assessment to decide on social status and placement, assessment is idealized as a mechanism to promote learning for all, regardless of social background, and to some degree regardless of inherent cognitive capacities. The socialization mechanism has enormous influence on learning goals. In this 21st century, when 20th century learning outcomes are perceived as less useful in the workforce, the stimulus for change has found its identity in calls for students to be creative, to be able to solve problems, to act autonomously as well as in groups, and to use and produce digital artefacts. These 'process' style outcomes are very different in nature to knowing historical facts, periodic tables, or writing highly literary prose and poetry.

The role of learning progressions in assessment of 21CS

A learning progression can be defined from an empirical standpoint in which evidence is derived from behaviors which are interpreted as "indicating" the skill and at a particular level of competence. Test items, or performance tasks, are drafted, which provide the opportunity for these indicative behaviors to be demonstrated by respondents. The items are located on a progress map (Masters & Forster, 1997) from less to more complex, or 'difficult,' which co-locate individuals along a continuum that represents different levels of learning and/or skill. This continuum, once associated with descriptions of skill at increasing levels of difficulty, is the learning progression.

According to Heritage (2008) a learning progression on reading can be developed in the following steps:

- Content experts define subcomponents of effective reading strategies;
- These subcomponents are then arranged hierarchically based on their developmental progression, as informed by the literature; ideally, the order of this hierarchy should imply a linear sequence of development in which one has to achieve mastery of a lower level before one can progress to subsequent levels;
- Markers for each subcomponent are identified and linked with a learning goal;
- Observations of these markers are used as evidence of a student's location along the progression; and
- The markers are grouped into blocks to form a practical sequence for teaching and learning.

Notwithstanding the view that learning progressions constitute the most likely opportunity we have to describe gradual acquisition of complex skills, the realities of variations in patterns of acquisition mean that progressions aspired to for 21CS may need to be described in large, rather than small, grain size.

A more data-driven strategy incorporates the concepts discussed above with the measurement approaches espoused by item response theory (IRT). The first step is to define the construct to be measured as concretely and objectively as possible, and hypothesize increasing levels of complexity. Items are then drafted to target the behaviors across the hypothesized spectrum. For example, descriptors of numeracy subskills can include the

following in increasing level of skill: arranges collections of objects in terms of quantity, arranges numbers in order of size, counts forwards/backwards using whole numbers, and answers questions on quantity using numbers. Student response data from the items are then mathematically modeled to convert into continuous scales. The resulting item response model links the measured properties for both the individual and the task components. The consequence is that in addition to locating a person on a scale, the model allows the task components (or items) to be located on the same scale (e.g., Wilson et al., 2012). The rest of the process is conceptually driven; clusters of proximate items on the scale are interpreted by subject or skills experts. These drive the drafting of level descriptions to form a practical sequence for teaching and learning. The weakness of this approach is exactly the same as for all measurement—and concerns the degree to which the items actually target the skill of interest, and reflect its true acquisition pattern.

One thing to keep in mind is that learning progressions work best when the learning goal is well-understood and defined. How skills are defined depends on the researcher, theory, and purpose. A learning progression that covers multiple domains of learning can be problematic, as the sequencing within each domain becomes increasingly difficult to disentangle. For example, if items that populate a learning progression for mathematics include the need to comprehend the written word, the progression for mathematics becomes confounded by the skill progression of comprehension.

This challenge of developing a learning progression becomes more pronounced as the target skill increases in complexity. Related to the problem of multidimensionality in IRT, complex skills are less likely to conform to a measurement model that assumes a single latent trait. The IRT solution in this instance is to model a multidimensional measure where related latent traits are modelled separately (but still allowed to correlate if necessary). The consequence for the development of learning progressions is that a separate progression for each latent trait needs to be developed. For example, a latent construct such as collaborative problem solving can be modeled as being composed of cognitive and social dimensions (or even more dimensions, see Scoular, Care & Hesse, 2017), each with its own developmental progression. This means that there is considerable substantive research and development to be undertaken worldwide if nations are to be reasonably confident that they are talking about the same skills, and if they are to have reasonable expectations of student progress that relate to age and grade. This takes us to big questions: How long will it take to have the research base for learning progressions for multi-dimensional traits in 21CS? Do education systems need to wait for the research? Are there other ways to create measurement models? Can traditional criteria of validity and reliability be applied to 21CS or should they be relaxed or ignored?

Other assessment considerations

INFORMATION AND COMMUNICATION TECHNOLOGIES

There has been great interest in the facility offered by information and communication technologies to the world of assessment, and its affordances are very evident in the three examples outlined. Some of this interest is associated with economies of scale—through electronic capture of scannable forms or of input responses in online environments, and automated scoring. Some of the interest is due to the potential for more varied data capture, beyond correct and incorrect responses, to provision of information about individuals' thought processes. Collection of information about what a student might have referred to in order to address task demands, or how long making some decisions took, might reveal valuable information about ability level and where support is needed. This is precisely the manner of information that is useful when evaluating an individual's 21CS. As demonstrated through ATC21S, these technologies have also been useful in exploring the nature of these skills—the fine-grained data capture provides a wealth of information to help track how these skills develop, and how they manifest in different ways according to different situational requirements.

There is no doubt that the move to computer-based testing has enabled measurement of 21CS to a greater degree than might be possible in traditional paper and pencil modes. Of course, self-report measures that are often used for perceptions of resilience or self-efficacy and the like are equally amenable to both modes—these, however, do not capture skills, but merely beliefs about skills or characteristics. The variety of media that can be integrated into computer-based assessment in particular may provide a context of greater authenticity (Dragow & Mattern, 2006) for stimulus and capture of these skills than traditional methods. Computer-based assessment is naturally well suited for those 21CS that require ICT—such as Digital Reading (OECD, 2011) and Learning in Digital Networks (Wilson, Scalise & Gochyyev, 2018). Beyond the multimedia presentation of stimuli for tasks, computer-based assessment can provide for a variety of capture modes—apart from logfiles populated by mouse-clicks and drag and drop events, voice, eye movement, and chat can be recorded and used in automated scoring. These process data

offer rich information that can be analyzed to identify specific behaviours that can be linked with cognitive processes (e.g., Adams, Vista, Awwal, Scoular, Griffin, & Care, 2015; Shute & Becker, 2010). For example, the sequence of action and chat between two individuals collaborating online can provide information about collaboration dynamics including perspective taking, negotiation, and conflict resolution. Similarly, time elapsed between actions or interactions, and different sequences of actions, can be analyzed against outcome data for different parts of a task, to determine patterns that predict different performance levels. What is learned from patterns such as these is a deeper understanding of the complex cognitive or social processes approaches taken by students as they engage with tasks (Vista, Awwal, & Care, 2016; Vista, Care, & Awwal, 2017).

CULTURAL FACTORS IN MEASUREMENT OF 21CS

The diversity and complexity of skills that fall under the 21CS umbrella mean that consideration of cross-cultural differences is an imperative. It is largely accepted that learning domains such as numeracy and literacy (at least for many languages that rely on the Roman alphabet) are comparatively robust to cultural differences, or at least are assessed in a way that minimizes cultural and linguistic effects. Recognition of this is signaled by the current efforts of UNESCO's Global Alliance for Monitoring Learning¹¹ to respond to the Sustainable Development Goal 4.1 to generate a universal scale for each of these two constructs. For cognitive skills such as problem solving and critical thinking, it may be reasonable to assume that despite cultural differences, the cognitive processes contributing to these skills will be reasonably similar. For 21CS that rest more on interpersonal and intrapersonal characteristics, cultural differences will be of more concern in the context of aspirations toward universal learning progressions. For example, there is a large literature on Asian versus Western views of working and learning, often seen as cooperative versus individualistic approaches (e.g., Hofstede, 1986). There are also major differences within these two main views reflecting different cultures within regions (e.g., Liem, Nair, Bernardo, & Prasetya, 2008). We cannot therefore assume cultural uniformity of 21CS; empirical studies are needed as a foundation.

Domain-general skills, such as transversal competencies (Asia-Pacific Education Research Institutes Network, 2013) and global citizenship, are starting to be incorporated in major international large-scale assessments. However, for the latter, there are different approaches, some of which center around national versus global citizenship conceptualizations. This is summarized in Care and Beswick (2016): "The cultural neutrality claimed as characteristic of large-scale assessments does not obviate the fact that although there may be sufficient commonality between countries to measure the same construct, that construct will have different valence in each country by virtue of its unique cultural perspective and how this permeates its education system" (p. 943). Others focus just on some aspects of constructs - for example the PISA 2018 assessment of global competency (OECD, 2016) mainly samples its knowledge component and some self-report on attitudes, as opposed to the skills component.

From a different but complementary perspective, if the focus of assessment is *for learning* (i.e., formative or informing learning; Black & Wiliam, 2009), and the main goal at the systems level is *education for all*, then local contexts matter more than cross-national comparability in educational assessments. The extent to which cultural differences affect the measurement approach or bias the results depends on the construct being measured. There is robust evidence of cross-cultural differences, but also similarities, in studies of values for example (Sverko, 2006), and in implications of formative assessment practices for teacher-student dynamics (Cagasan, Care, Robertson & Luo, in press). Some 21CS are not just transversal in terms of learning domain but also in terms of geography, while other skills are very much affected by cultural contexts (e.g., communication skills and global citizenship).

The recommended assessment strategies and examples of tools and programs indicate what is possible in the realm of assessment of 21CS. Some of the technical challenges are outlined in the next section.

COUNTRY MODELS AND EXAMPLES

Taking the perspective that change in education goals must take shape through alignment of curriculum, pedagogy, and assessment, we explore the cases of three countries that are in the process of re-aligning their goals within their local system. As will be seen, the primary route taken to initiate the shift in learning goals is through curriculum reform, but approaches to assessment aligned with the new learning goals have also been noteworthy. Alignment is a criterion for a fully functional system, not only across the curriculum, pedagogy, and assessment triad, but also of assessment across classroom, examination, and national assessment levels.

¹¹ <http://gaml.uis.unesco.org/>

The case of the Philippines

The Republic of the Philippines' Department of Education (DepEd) initiated a major reform in 2013; this included structural reforms through adding years of schooling to the basic education sector, curriculum revision, formulation of an assessment framework, and focus on 21CS. However, the main effort in implementing “K to 12” has been on the curricular changes, with attention on the new mother tongue-based multi-lingual education delivery in the elementary years, the 'spiral' science approach in the junior secondary, and extension of the basic education system to Years 11 and 12. Although aspirations for nurturing "the holistically developed Filipino" (Okabe, 2013), who will have 21CS and who will be prepared for further education and employment are highly visible in formal DepEd communications, or 'memos' (Department of Education, 2015, 2016, 2017), the mechanisms for implementation of this set of new learning goals is not yet well-developed. What has been explored, however, is how to assess these skills—notably at the national level.

DepEd Order No. 8 (2015) defines *classroom assessment* as “an ongoing process of identifying, gathering, organizing, and interpreting quantitative and qualitative information about what learners know and can do” (p. 1). Specifically, the purposes of classroom assessment are:

- a. "To keep track of learners' progress in relation to **learning standards and in the development of 21st century skills**;
- b. To promote self-reflection and personal accountability among students about their own learning;
- c. To provide bases for the profiling of student performance on the learning competencies and standards of the curriculum."

The identification of 21CS as a priority highlights the ambitions of DepEd for its students. Similarly, the first statement of purpose for the *national assessments* (Department of Education, 2016) is: “Assessment is the process of measuring learners' progress in the attainment of learning standards and 21st century skills” (DepEd Order No. 55). DepEd Order No. 29 (2017) for *system assessment*, identifies the 21CS areas of global citizenship/civics education, through the regional large-scale assessment program, South East Asia Primary Learning Metrics; and global competence through OECD's PISA 2018 cycle. For the country's National Achievement Test, 21CS are again identified for its implementation at the end of Grade 6, 10, and 12: “This assessment covers core 21st century skills and the core...learning areas of Languages, Humanities, Communication, Mathematics, Science, Social Science, and Philosophy” (p. 16).

Notwithstanding the explicit focus on 21CS assessment outlined in formal orders, and identification of 21CS in curricular goals, there is little evidence that these competencies are prioritized in classroom practice. Although a great deal of high standard technical work has been committed to assessment of 21CS within content domains by DepEd's Bureau of Educational Assessment (BEA), it is not clear that this commitment has been paralleled by focus on 21CS by the Bureau of Curriculum Delivery, nor by ensuring that teachers are well-versed in the nature of these skills and how they might be enhanced or taught in the classroom. BEA completed a full audit of curriculum in certain subjects at selected grade levels in order to identify the opportunities for learning, teaching, and demonstration of 21CS. Following this, using a template approach to drafting test items, the unit finalized a set of items in which the same indicators of skills were elicited across three subjects, in order to build a robust multi-disciplinary assessment approach. These items were piloted and trialed, and are now included in the national achievement tests.

In order to roll out education reform, curriculum re-writing is just one step. Training teachers to focus on the new goals of the education system, and on understanding and application, rather than content coverage, is another. This poses a very real issue in classrooms where traditional instruction predominates with pedagogical practices focused on subject matter and structure. To equip students to responding to the non-routine, constructivist pedagogical practices which include emphasis on learning processes and discovery (Zuzovsky, 2013), need to be integrated within classroom teaching and learning. Associated with these pedagogies, “informal” formative assessment practices which reflect unstructured but responsive and dynamic teaching are best designed to stimulate development of 21CS. The reality is that many Philippine classrooms follow a common and relatively inflexible classroom structure (Griffin, Cagasan, Care, Vista, & Nava, 2016). This has not changed, as would be consistent with the K to 12 education reform learning goals.

The case of Australia

In the Australian curriculum, an emphasis on 21CS is reflected in the 'General Capabilities' (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2013). There are seven of these: Literacy, Numeracy, Information and Communication Technology (ICT) Capability, Critical and Creative Thinking, Personal and Social Capability, Ethical Understanding, and Intercultural Understanding (ACARA, 2013). They act as the realization of goals set out by the Melbourne Declaration on Educational Goals for Young Australians (2008) "that all young people in Australia should be supported to become successful learners, confident and creative individuals, and active and informed citizens." ACARA outlines two approaches to the integration of the general capabilities in teaching and learning. First, within learning area content descriptions, applications of specific general capabilities are provided (Figure 5). Second, 'learning continua' have been developed and are available, which describe the relevant knowledge, skills, behaviors, and dispositions at particular points of schooling for each of the capabilities. The Australian system provides robust and comprehensive resources for integrating the capabilities with the key learning areas of the curriculum.

Figure 5. Example of clickable guidelines and resources for integration of general capabilities within key learning areas of the Australian curriculum

Year 7 Content Descriptions

Geographical Knowledge and Understanding

Unit 1: Water in the world

Classification of [environmental resources](#) and the forms that water takes as a resource (ACHGK037 - [Scootle](#))

The way that flows of water connects places as it moves through the [environment](#) and the way this affects places (ACHGK038 - [Scootle](#))

The quantity and variability of Australia's water resources compared with other continents (ACHGK039 - [Scootle](#))

The nature of [water scarcity](#) and ways of overcoming it, including studies drawn from Australia and West Asia and/or North Africa (ACHGK040 - [Scootle](#))

Unit 1: Water in the world

Classification of [environmental resources](#) and the forms that water takes as a resource

Numeracy

Estimating and calculating with whole numbers

- Estimate and calculate

Recognising and using patterns and relationships

- Recognise and use patterns and relationships

Unit 1: Water in the world

Classification of [environmental resources](#) and the forms that water takes as a resource

Critical and Creative Thinking

Inquiring – identifying, exploring and organising information and ideas

- Identify and clarify information and ideas
- Organise and process information

Downloaded 10-May-2018 from <https://www.australiancurriculum.edu.au/f-10-curriculum/humanities-and-social-sciences/geography/?year=12377&strand=Geographical+Knowledge+and+Understanding&strand=Geographical+Inquiry+and+Skills&capability=ignore&capability=Literacy&capability=Numeracy&capability=Information+and+Communication+Technology+%28ICT%29+Capability&capability=Critical+and+Creative+Thinking&capability=Personal+and+Social+Capability&capability=Ethical+Understanding&capability=Intercultural+Understanding&priority=ignore&priority=Aboriginal+and+Torres+Strait+Islander+Histories+and+Cultures&priority=Asia+and+Australia%E2%80%99s+Engagement+with+Asia&priority=Sustainability&elaborations=true&elaborations=false&scotterms=false&isFirstPageLoad=false>

Teachers are expected to teach these “General Capabilities” within their subject specializations, but how each state supports teachers to do so varies. For example, the State of Victoria’s Department of Education and Training funds the provision of professional development opportunities for teachers to enhance their understanding on how to teach the general capabilities, through the Victorian Curriculum and Assessment Authority.¹² The training is currently offered for critical and creative thinking, ethical capability, and personal and social capability through online professional learning sessions, face-to-face workshops, conference days, and on-demand sessions at school the level. However, there are major differences in the extent to which schools access these services.

In terms of teaching and learning approaches, the State of Victoria expects each school to nominate how it will approach integration of the four general capabilities prioritized in the Victorian variant of the national curriculum (Critical and Creative Thinking, Personal and Social Capability, Ethical Understanding, and Intercultural Understanding). Schools are formally required to note that they are complying with the policy, and must report explicitly to parents on the general capabilities every two years. Schools may opt to integrate the capabilities within current key learning domains, offer electives sampling them, or include them in “experience” units. Some schools have taken the path of identifying flagship subjects to highlight particular capabilities. For example, Melbourne High School, a selective-intake senior secondary state school in inner Melbourne, has taken a specific curricular approach to developing the capabilities. The school has designed a philosophy unit that takes Critical and Creative Thinking and Ethical Understanding as cornerstones—an approach designed not only to enhance the capabilities, but also to promote philosophy as a learning domain. Additionally, all subject teachers are required to identify how they are integrating the general capabilities within their subjects. Melbourne High School leadership sees its approach as providing an opportunity to diversify the school’s reputation from its very high achieving “academic” character to a more humanist identity.

In terms of assessment, there is variation across states but little clear guidance. The State of New South Wales’ Department of Education, for example, provides an article which identifies the SOLO Taxonomy (Biggs & Collis, 1982) and its five levels of increasing complexity in learning outcomes as a useful approach. The degree to which this can usefully be applied to capabilities such as Intercultural Understanding or Personal and Social Capability is not clear. West Australia’s School Curriculum and Standards Authority states that teachers are expected to teach and assess general capabilities to the extent that they are incorporated into the key learning areas, but the templates provided for assessment of these areas are not similarly provided for the general capabilities.

The case of Kenya

Consistent with a shift in education globally toward quality education and development of skills in young people, the Kenyan Ministry of Education, Science and Technology (MOEST) initiated a major reform in 2014. This constituted an effort to address issues of access, quality, relevance, and equity as identified by the Constitution of Kenya and Kenya Vision 2030 (Ministry of Education, Science and Technology, 2015). The Second Medium Term Plan of Kenya Vision 2030 prioritized curriculum reform toward a competency-based curriculum to equip all learners with the 21st century competencies and qualifications that can promote national values and inspire individual innovation and life-long learning. The curriculum reform vision is “to enable every Kenyan to become an engaged, empowered, and ethical citizen. This will be achieved by providing every Kenyan learner with world class standards in the skills and knowledge that they deserve, and which they need in order to thrive in the 21st century. This shall be accomplished through the provision of excellent teaching, school environments and resources...pedagogical tools...” (Kenya Institute of Curriculum Development [KICD], 2017, p. 10). In order to achieve this vision and to “nurture every learner’s potential,” the new curriculum is based on values and theoretical perspectives on learning and development to identify seven core competencies that every learner in basic education can achieve. These core competencies are embedded in subject areas such as English, mathematics, science and technology, and include:

- Communication and collaboration
- Self-efficacy
- Critical thinking and problem solving
- Creativity and imagination
- Citizenship

¹² <http://www.vcaa.vic.edu.au/Pages/foundation10/viccurriculum/proflearning/complete-professional-learning.aspx>

- Digital literacy
- Learning to learn

Furthermore, specific learning outcomes are identified for each education level. For example, by the end of middle school (upper primary—Grades 4-6—and lower secondary—Grades 7-9), learners should be able to communicate effectively to different audiences; demonstrate social skills, and spiritual and moral values for peaceful co-existence; and apply digital literacy skills appropriately for communication and learning. Curriculum designs are in the process of being created¹³, and are intended to facilitate the implementation of the competency-based curriculum by providing guidance to teachers. For each grade level, strands (subject topic area), sub-strands, specific learning outcomes for the strand, suggested learning experiences, and key inquiry questions are identified. In addition, core competencies, values, suggested activities and assessment formats (e.g., observation, oral questions), and assessment rubrics are identified. However, most guidance remains around the teaching and assessment of subject areas, rather than a specific focus on the core competencies.

Although the curriculum is the current focus, establishing new pedagogical approaches to implement a competency-based curriculum and institutionalize formative assessment at all levels of basic education are considered part of the reform process. Now that the curriculum designs have been completed for pre-primary and lower primary, the MOEST is focusing on implementation and assessment of the new curriculum. There is acknowledgement that learning should be experiential, with an inquiry based approach, and that learners should engage in non-formal and informal activities to acquire core competencies and translate learning experiences into real life situations through age-appropriate projects and action research. However, it is not yet clear what these teaching approaches and strategies might look like in the classroom. Similarly, beyond the recognition that “too much focus on summative assessment should be avoided” and “a range of assessment focuses on the development of student learning outcomes, cross-curricular competencies, and literacy and numeracy should be adopted” (KICD, 2017, p. 26), there is little guidance on what assessments that are aligned with the new competency-based curriculum should entail. There are plans to redesign assessment to ensure that it is used as a tool for learning as well as for gathering information on whether learning outcomes have been achieved. Some suggestions include observational checklists to indicate mastery, homework, rubrics, anecdotal notes, and performance indicators. MOEST will be working to incorporate these suggestions into the current grading system, which relies on percentages, and cut off points that indicate competency level (e.g., competent, fairly competent, not yet competent), rather than on descriptive learning progressions that would identify behavioral indicators to inform more grounded understanding of student progress.

In each of these countries, we see a common approach of embedding skills in subject matter or key learning areas. There are stark differences across the countries in the amount of resources and guidance given. None of the countries provide explanations of why particular pedagogical approaches should be adopted. The Philippines is the stand-out case for assessment by virtue of its work in integrating 21CS into its national achievement tests, although this is not paralleled in classroom practice. These countries, which are in the forefront of a global shift to accommodate new learning goals, provide a picture of the complexity and the decisions that this shift denotes. All three countries reveal that reform begins with policies and proclamations with high aspirations but few details, limited resources, and little teacher training.

All of the characteristics outlined in Table 6 can be seen in some classrooms in these three countries, and probably in most countries around the world. The key issue for education systems adopting 21CS is the variable distribution of these practices across classrooms.

Table 6. Mainstream characteristics of 20th and 21st century assessments

Characteristics of 20 th century assessments	Characteristics of 21 st century assessments
Subject-based, often a narrow focus on mathematics, reading, and science	Skills- or competency-based, including cross-disciplinary skills such as communication, collaboration, and problem solving

¹³ <https://kicd.ac.ke/curriculum-reform/curriculum-design>

Questions are mostly multiple choice, and have one correct answer	Multiple types of questions, including open-ended and observation of behaviors; process is as important as outcome
Students all answer the same questions on tests	Students take subsets of a test (adaptive testing) and still be located on the same scale
Paper and pencil tests	Diverse presentation of tests and test items through games and complex tasks in a digital environment
Primarily measures knowledge	Measures a wide range of human competencies
Use analytical approaches based on classical test theory	Use modern analysis techniques, such as item response theory and structural equation modeling
Can be based on decades of research on how skills progress in domains such as reading and mathematics	Currently limited research on how skills progress

CONCLUSION

The noticeable shift in educational goals toward equipping students with a broad range of skills is being recognized through education and curricular reform efforts. However, the shift in educational goals has not yet translated into practice, and countries are struggling with how to fully implement a 21CS agenda that focuses on teaching and assessment that is aligned with the changing goals. This paper described the shift at the global level, and provided examples of how countries—unlike in terms of geographic regions, educational achievement (as measured by international large-scale assessments), and economic productivity—demonstrate remarkably similar curricular structures and learning goals.

In addition, this paper examined the educational shift toward a focus on 21CS and discussed the overarching challenge of aligning the components of the education system—namely, curriculum, assessment, and pedagogy—in light of this shift. Specifically, we identified what is necessary to incorporate a 21CS agenda within national education systems, which emphasizes the need for alignment of education’s three delivery systems of curriculum, pedagogy, and assessment, as well as the critical issue of having deep knowledge of the learning goals, especially when those goals are new. Within this context, three challenges to incorporating a 21CS agenda were highlighted.

Challenge 1: Understanding the nature of 21CS

Although there is research regarding the subskills relating to 21CS, there is a lack of knowledge and understanding of the *developmental nature* of these critical skills in the new learning goals. Three 21CS—collaboration, critical thinking, and problem solving—are described in terms of their components and processes. However, the interrelations among subskills and their developmental trajectories are unclear. This poses questions about how to differentiate between levels of skill development, and therefore, has implications for how to teach and assess these skills.

Challenge 2: Developing learning progressions of 21CS

Several possible approaches to this differentiating level of skills are described, one of which constitutes the use of learning progressions—pathways to acquisition and development of skills. These descriptions of how students typically achieve mastery of a particular learning domain—from basic to more sophisticated levels—form the foundation for aligning teaching and learning with the goals and standards in the curriculum. Although learning progressions are available for traditional learning domains, such as mathematics and science, a reasonable sequence of learning 21CS does not exist.

Challenge 3: Designing appropriate and authentic assessment of 21CS

For 21CS, the issue is the degree to which traditional forms of assessment can capture and report on students’ proficiencies. One way to approach this challenge is to examine the authenticity of assessment tasks. It is clear that

with the examples presented and analyzed using Gulikers et al.'s (2004) framework, there is a strong degree of authenticity—the tasks are administered within the artificial environment of assessment itself, rather than reflecting student efforts to complete the real-life tasks that these assessment tasks attempt to emulate.

Re-aligning the system: Examples of country reform efforts

Finally, the reform efforts of three countries—the Philippines, Australia, and Kenya—are described to demonstrate some of the accomplishments and challenges of implementing a learning goal shift. The similarities in learning aspirations of education systems globally is a promising sign that educators agree about the need for a new look at what is most valued in terms of student learning outcomes. These similarities signal the likelihood of major efforts to surmount some of the challenges identified in this paper. Although we have focused on assessment, it is essential to understand that concurrent review of pedagogical strategies—and methods of integration of skills into the curriculum—must occur for systems to experience success in early reform efforts.

REFERENCES

- Adams, R. J., Vista, A., Awwal, N., Scoular, C., Griffin, P., & Care, E. (2015). Automatic Coding Procedures. In P. Griffin & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills - Methods and Approaches* (pp. 115-132). Dordrecht: Springer.
- Adedeji, S. O., & Olaniyan, O. (2011). Improving the conditions of teachers and teaching in rural schools across African countries. Addis Ababa, Ethiopia: UNESCO International Institute for Capacity Building in Africa. Retrieved from <http://unesdoc.unesco.org/images/0021/002160/216062e.pdf>.
- Ahonen, A., Häkkinen, P., Pöysä-Tarhonen, J. (2018). Collaborative Problem Solving in Finnish Pre-Service Teacher Education: A Case Study. In E. Care, P. Griffin & M. Wilson (Eds.). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Cham: Springer.
- Alexander, P. A., & Murphy, P. K. (1998). The research base for APA's Learner-Centered Psychological Principles. In N. Lambert & B. L. McCombs (Eds.), *How students learn: Reforming schools through learner-centered education*. Washington, DC: American Psychological Association.
- Anderson, C. (2012). *Makers: The New Industrial Revolution*. New York: Penguin Random House.
- Anderson, L. W. & Krathwohl, D. R., et al. (Eds.) (2001) A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives. New York: Longman.
- APA Work Group of the Board of Educational Affairs (1997). *Learner-centered psychological principles: Guidelines for school reform and redesign*. Washington, DC: American Psychological Association.
- Asia-Pacific Education Research Institutes Network. (2013). Integrating transversal competencies in educational policy and practice *ERI-Net Regional Policy Study Series* (Vol. 1): Bangkok, Thailand.
- ATC21S. (2014). Framework for 21st century skills. Retrieved from <http://www.atc21s.org/>.
- Australian Curriculum, Assessment and Reporting Authority (ACARA) (2013). *General Capabilities in the Australian Curriculum*. Sydney: ACARA.
- Australian Council for Educational Research. (2016). *Linguistic quality assurance in large-scale learning assessments*. New Delhi: RMSA-TCA and the National Council of Educational Research and Training.
- Biggs, J., & Collis, K. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Dordrecht: Springer.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability* (formerly: *Journal of Personnel Evaluation in Education*), 21(1), 5.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*, by a committee of college and university examiners. Handbook I: Cognitive Domain. NY, NY: Longmans.
- Cagasan, L., Care, E., Robertson, P., & Luo, R. (in press). Capturing formative assessment practices in the Philippines through an observation tool. *Educational Assessment Journal*.
- Care, E. (2018a). Twenty-first century skills: From theory to action. In E. Care, P. Griffin & M. Wilson (Eds.). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Cham: Springer.
- Care, E. (2018b). Assessment in the Philippines' Formal Education System. In C. Magno and A. David (Eds.) *Philippine and Global Perspectives on Educational Assessment*, pp. 30-52. Manila: Philippine Educational Measurement and Evaluation Association.
- 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., & Beswick, B. (2016). Comparison and Countries. In D. Wyse, L. Hayward, & J. Pandya (Eds.), *The SAGE handbook of curriculum, pedagogy and assessment* (Vol. 2, pp. 928-945). London: Sage.
- Care, E., & Kim, H. (2018). Assessment of 21st century skills: The issue of authenticity. In E. Care, P. Griffin & M. Wilson (Eds.). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Cham: Springer.
- Care, E., & Luo, R. (2016). Assessment of transversal competencies: Policy and practice in the Asia-Pacific region. Bangkok, Thailand: UNESCO.
- Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, 29(4), 250-264.
- Care, E., Vista, A., & Kim, H. (2018). Assessment of transversal competencies: potential of resources in the Asian region. Paris and Bangkok: UNESCO.

- Child, S., & Shaw, S. (2016). Collaboration in the 21st century: Implications for assessment. *Research Matters*, 22 (Summer 2016), 17-22.
- Child Trends (2015). Key “soft skills” that foster youth workforce success: toward a consensus across fields. Washington DC: Child Trends.
- Dede, C. (2010). Comparing Frameworks for 21st Century Skills. In J. Bellanca and R. Brandt (Eds.), *21st Century Skills: Rethinking how students learn*. Bloomington, IN: Solution Tree Press, pp. 51–76.
- Delors, J., Al Mufti, I., Amagai, I. A., Carneiro, R., Chung, F., Geremek, B.,... Nanzhao, Z. (1996). *Learning: The treasure within*. Paris, France: UNESCO.
- Department of Education (2015). Policy Guidelines on Classroom Assessment for the K to 12 Basic Education Program. DepEd Order No. 8. Manila: Republic of the Philippines.
- Department of Education (2016). Policy Guidelines on the National Assessment of Student Learning for the K to 12 Basic Education Program. DepEd Order No. 55. Manila: Republic of the Philippines.
- Department of Education (2017). *Policy Guidelines on System Assessment in the K to 12 Basic Education Program*. DepEd Order No. 29. Manila: Republic of the Philippines.
- Dillenbourg, P. (1999). What do you mean by 'collaborative learning'? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 1-19). New York: Pergamon.
- Drasgow, F., & Mattern, K. (2006). New Tests and New Items: Opportunities and Issues. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based Testing and the Internet: Issues and Advances* (pp. 59-76). Chichester, UK: John Wiley and Sons Ltd.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values and goals. *Annual Review of Psychology*, 53, 109-132.
- Education Commission (2016). *The Learning Generation: Investing in Education for a Changing World*. The International Commission on Financing Global Education Opportunity. http://report.educationcommission.org/wp-content/uploads/2016/09/Learning_Generation_Full_Report.pdf. Accessed 8th Sep 2018.
- Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.
- Enyedy, N., & Stevens, R. (2014). Analyzing collaboration *The Cambridge handbook of the learning sciences, second edition* (pp. 191-212). Cambridge, UK: Cambridge University Press.
- Ercikan, K., & Oliveri, M. E. (2016). In search of validity evidence in support of the interpretation and use of assessments of complex constructs: Discussion of research on assessing 21st century skills. *Applied Measurement in Education*, 29(4), 310-318.
- European Centre for the Development of Vocational Training (CEDEFOP) (2012). *Curriculum Reform in Europe: The Impact of Learning Outcomes*. Luxembourg: Publications Office of the European Union.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39(3), 193.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-6.
- Griffin, P., Cagasan, L., Care, E., Vista, A., & Nava, F. (2016). Formative assessment in the Philippines. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge*. Springer: Dordrecht.
- Griffin, P., & Care, E. (2014). Test Construction. In P. Griffin (Ed.), *Assessment for Teaching* (pp. 156-173). Cambridge: Cambridge University Press.
- Greenstein, L. (2012). *Assessing 21st century skills: A guide to evaluating mastery and authentic learning*. Thousand Oaks, CA: Corwin.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52, 67-86.
- Halpern, D. (1998) Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449-455.
- Heritage, M. (2008). *Learning Progressions: Supporting Instruction and Formative Assessment*. University of California, LA: National Center For Research On Evaluation, Standards, And Student Testing (CRESST). <http://www.k12.wa.us/assessment/ClassroomAssessmentIntegration/pubdocs/FASTLearningProgressions.pdf>.
- 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*. Dordrecht: Springer.
- Hoffman, R. R. (1998). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams, W. Faulkner & J. Fleck (Eds.), *Exploring expertise* (pp. 81–100). New York, NY: Macmillan.
- Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of Intercultural Relations*, 10, 301-320.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674-689.
- Kenya Institute of Curriculum Development. (2017). *Basic education curriculum framework*. Retrieved from <https://kicd.ac.ke/curriculum-reform/basic-education-curriculum-framework/>.

- Kim, H., & Care, E. (2017). Considering authenticity of skills assessments. *Journal of Educational Assessment in Africa*, 12, 165-175.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, 96(4), 674-689.
- Lai, E. R., & Viering, M. (2012). *Assessing 21st century skills: Integrating research findings*. Retrieved from http://images.pearsonassessments.com/images/tmrs/Assessing_21st_Century_Skills_NCME.pdf.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice*, 32(3), 131-137.
- Liem, A. D., Nair, E., Bernardo, A. B. I., & Prasetya, P. H. (2008). The influence of culture on students' classroom social interactions. In D. M. McInerney and A. D. Liem (Eds.) *Teaching and Learning: International Best Practice*, pp. 377-404. IAP: Charlotte, NC.
- Looney, J. (2011). Alignment in complex education systems: Achieving balance and coherence. *OECD Education Working Paper*, No. 64.
- Masters, G. N., & Forster, M. (1997). *Developmental Assessment*. Victoria, Australia: Australian Council for Educational Research.
- Mayer, R. E. (2013). Problem solving. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*. New York, NY: Oxford University Press.
- Monk, D. H. (2007). Recruiting and retaining high-quality teachers in rural areas. *The Future of Children*, 17, 155-174. <https://files.eric.ed.gov/fulltext/EJ795884.pdf>.
- Mueller, J. (2005). The authentic assessment toolbox: Enhancing student learning through online faculty development. *Journal of Online Learning and Teaching*, 1, 1-7.
- 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.
- Nusche, D. (2016). Student assessment and its relationship with curriculum, teaching and learning in the twenty-first century. In Wyse, D., Hayward, L., & Pandya, J. (2016) (Eds.). *The SAGE Handbook of Curriculum, Pedagogy and Assessment*. London: Sage.
- OECD. (2011). PISA 2009 Results: Students on Line: Digital Technologies and Performance (Volume VI). Paris: OECD.
- OECD (2013a). *Synergies for better learning: An international perspective on evaluation and assessment*. Paris: OECD. <http://www.oecd.org/edu/school/synergies-for-better-learning.htm>.
- OECD (2013b), PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy, OECD Publishing.
- OECD (2014). PISA 2012 Results: Creative Problem Solving. Paris: OECD.
- OECD (2015), Skills for Social Progress: The Power of Social and Emotional Skills, OECD Skills Studies, OECD Publishing. <http://dx.doi.org/10.1787/9789264226159-en>.
- OECD (2016). Global competency for an inclusive world, OECD Publishing.
- OECD. (2017). *PISA 2017 Collaborative Problem Solving Framework*. Retrieved from: <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>
- Okabe, M. (2013) 'Where does Philippine education go? The "K to 12" program and reform of Philippine Basic Education', Institute of Developing Economies Discussion Paper No 425. Available from <http://www.ide.go.jp/English/Publish/Download/Dp/pdf/425.pdf> (accessed 15 March 2015).
- O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2004). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, 10, 361-373.
- Partnership for 21st Century Learning. (2015). *P21 framework definitions*. Retrieved from http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf.
- Pellegrino, J. W., & Hilton, M. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. Washington, D.C: National Research Council.
- Pellegrino, J. W., DiBello, L. V., & Goldman, S. R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. *Educational Psychologist*, 0(0), 1-23.
- Pitman, T., & Broomhall, S. (2009). Australian universities, generic skills and lifelong learning. *International Journal of Lifelong Education*, 28(4), 439-458.
- Popham, W. J. (2007). All about accountability: The lowdown on learning progressions. *The Prepared Graduate*, 64(7), 83-84.
- Quellmalz, E. S., Timms, M. J., & Buckley, B. C. (2009). Using science simulations to support powerful formative assessments of complex science learning. WestEd. Retrieved from http://simscientists.org/downloads/Quellmalz_Formative_Assessment.pdf
- Quellmalz, E. S., Silbergliitt, M. D., Buckley, B. C., Loveland, M. T., & Brenner, G. T. (2016). Simulations for Supporting and Assessing Science Literacy. In Y. Rosen, S. Ferrara, & M. Mosharrarf (Eds.) *Handbook of Technology Tools for Real-World Skill Development*. IGI Global, pp. 191-229.

- Ramalingam, D., & Adams, R. (2018). How can use of data from computer-delivered assessments improve measurement? In E. Care, P. Griffin & M. Wilson (Eds.). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Cham: Springer.
- Reimers, F. M., & Chung, C. K. (2016) (Eds.). *Teaching and Learning for the Twenty-First Century: Educational Goals, Policies and Curricula from Six Nations*. Cambridge, Mass.: Harvard Education Press.
- Robinson, K., & Aronica, L. (2015). *Creative Schools: Revolutionizing Education from the Ground Up*. Penguin.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 69-97). Berlin: Springer.
- Rushkoff, D. (2013). *Present Shock*. NY: Penguin.
- Schleicher, A. (Ed.) (2012), *Preparing Teachers and Developing School Leaders for the 21st Century: Lessons from around the World*, OECD Publishing. <http://dx.doi.org/10.1787/9789264174559-en>.
- Schmidt, E., & Cohen, J. (2013). *The New Digital Age: Reshaping the Future of People, Nations and Business*. Hodder & Stoughton.
- Schmidt, E., & Cohen, J. (2013). *The New Digital Age: Reshaping the Future of People, Nations and Business*. Hodder & Stoughton.
- Scoular, C., & Care, E. (2018). Teaching 21st century skills: Implications at system levels in Australia. In E. Care, P. Griffin & M. Wilson (Eds.). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Dordrecht: Springer.
- Scoular, C., Care, E., & Hesse, F. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement*, 54 (1), 12-35.
- Shute, V., & Becker, B. J. (2010). *Innovative assessment for the 21st century*: Springer.
- Siarova, H., Sternadel, D., & Mašidlauskaitė, R. (2017). Assessment practices for 21st century learning: review of evidence. *NESET II report*, Luxembourg: Publications Office of the European Union, doi: 10.2766/71491.
- Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st Century Competencies: Guidance for Educators*. Global Cities Education Network, Asia Society and Rand Corporation. <http://asiasociety.org/files/gcen-measuring21skills.pdf> (Accessed 12 October 2016).
- Sternberg, R. J. (1986). *Critical thinking: Its nature, measurement, and improvement*. Washington, DC.: National Institute of Education.
- Stolk, Y., Kaplan, I., & Szwarc, J. (2016). Review of the strengths and difficulties questionnaire translated into languages spoken by children and adolescents of refugee background. *International Journal of Methods in Psychiatric Research*, 26(4).
- Sverko, B. (2006). The Structure and Hierarchy of Values Viewed Cross-Nationally. In D. E. Super, B. Sverko and C. M. Super (Eds.), *Life Roles, Values, and Careers: International Findings of the Work Importance Study*. Jossey-Bass.
- Tomasello, M., & Hamann, K. (2012). Collaboration in young children. *Quarterly Journal of Experimental Psychology*, 65(1), 1–12.
- United Nations General Assembly (2015). *Transforming our World: the 2030 Agenda for Sustainable Development*. A/RES/70/1. United Nations.
- UNESCO (2015a). *Transversal competencies in education policy and practice (Phase I): Regional synthesis report*. Paris, France: UNESCO. Retrieved August 14, 2015 from <http://unesdoc.unesco.org/images/0023/002319/231907E.pdf>
- UNESCO (2015b). *Sustainable Development Goals: 17 Goals to Transform our World*. <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>. Paris, France: UNESCO
- UNESCO (2016a). *School and teaching practices for twenty-first century challenges: Lessons from the Asia-Pacific region (Phase II): Regional synthesis report*. Paris, France: UNESCO. Retrieved October 17, 2016 from <http://unesdoc.unesco.org/images/0024/002440/244022E.pdf>.
- UNESCO (2016b). *Preparing and supporting teachers in the Asia-Pacific to meet the challenges of twenty-first century learning (Phase III): Regional synthesis report*. Paris, France: UNESCO. Retrieved June 5, 2017 from <http://unesdoc.unesco.org/images/0024/002468/246852E.pdf>.
- UNESCO (2017). *Education for Sustainable Development Goals: Learning Objectives*. Paris: UNESCO.
- United Nations (2016). *The Sustainable Development Goals Report*. <http://unstats.un.org/sdgs/report/2016/The%20Sustainable%20Development%20Goals%20Report%202016.pdf>.
- Vista, A., Awwal, N., & Care, E. (2016). Sequential actions as markers of behavioural and cognitive processes: Extracting empirical pathways from data streams of complex tasks. *Computers and Education*, 92, 15-36.
- Vista, A., Care, E., & Awwal, N. (2017). Visualising and examining sequential actions as behavioural paths that can be interpreted as markers of complex behaviours. *Computers in Human Behavior*. doi:10.1016/j.chb.2017.01.027.
- Vygotsky, L. S. (1978). *Mind and society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Warneken, F., Graefenhain, M., & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children's understanding of joint intentions in collaborative activities. *Developmental Science*, 15(1), 54-61.

- WestEd (2015). *SimScientists WestEd Ecosystems Try It*. Retrieved from <http://simscientists.org/MiniSite/index.html>.
- Whitlock, B., & Nanavati, J. (2013). A systematic approach to performative and authentic assessment. *Reference Services Review*, *41*, 32-48.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessments. *Phi Delta Kappan*, *70*, 703-713.
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design* (Expanded Second Edition). Alexandria, Va.: ASCD.
- Willingham, D. T. (2007). Critical thinking: Why is it so hard to teach? *American Educator*, *31*, 8-19.
- Wilson, M., Beja, I., Scalise, K., Templin, J., Wiliam, D., & Torres Irribarra, D. (2012). Perspectives on methodological issues. In P. Griffin, B. McGaw, & E. Care (Eds.). *Assessment and Teaching of 21st century Skills*. Dordrecht: Springer.
- Wilson, M., Gochyyev, P., & Scalise, K. (2016). Assessment of learning in digital interactive social networks: A learning analytics approach. *Online Learning*, *20*(2), 97–119.
- Wilson, M., Scalise, K., & Gochyyev, P. (2018). Learning in digital networks as a modern approach to ICT Literacy. In E. Care, P. Griffin, & M. Wilson (Eds.) *Assessment and Teaching of 21st century Skills: Research and Applications*. Cham: Springer.
- World Economic Forum (2016). *New Vision for Education: Fostering Social and Emotional Learning through Technology*. Geneva: World Economic Forum.
- World Bank. (2018). *World Development Report 2018: Learning to Realize Education’s Promise*. Washington, DC: World Bank.
- Wyse, D., Hayward, L., & Pandya, J. (2016) (Eds.). *The SAGE Handbook of Curriculum, Pedagogy and Assessment*. London: Sage.
- Zuzovsky, R., (2013). What works where? The relationship between instructional variables and schools’ mean scores in mathematics and science in low-, medium- and high-achieving countries. *Large-scale Assessments in Education*, *1*(2).

