SKILLS FOR LIFE

Measuring 21st Century Skills in Latin America and the Caribbean

IDB Inter-American Development Bank

Author

Maria Eugenia Oviedo

Master in Public Policy
Harvard Kennedy School

Advisors

Christopher Avery

Roy E. Larsen Professor of Public Policy
Harvard Kennedy School

Janina Matuszeski

Lecturer in Public Policy
Harvard Kennedy School

The Harvard Kennedy School at Harvard University seeks to improve public policy and leadership through cutting-edge research, the teaching of outstanding students, and direct interaction with practitioners. As a part of the Master in Public Policy, graduate students complete a Policy Analysis Exercise (PAE). The PAE is a capstone project in which students examine and develop solutions to a policy or management challenge presented by a client organization.

Abstract

Measuring 21st century skills is one of the most important tasks of education and training systems. Measurement is what allows program managers to evaluate the effectiveness of a particular education program, diagnose the needs of individual students, or assess their development over time. Still, countries in Latin America and the Caribbean—and around the world—face challenges in this area. Using data from interviews, case studies, and an in-depth literature review, this policy brief evaluates the advantages and disadvantages of different types of assessment tools for measuring 21st century skills, discusses challenges of applying them in less-developed countries, and offers recommendations for stakeholders seeking to measure 21st century skills among children and youth, by utilizing existing instruments or adapting and testing them in local contexts.

1. Introduction

In today's fast-changing and globalized economy, students' 21st century skills, such as collaborative problem-solving and communication skills, are more important than ever (Wolvin & Lim, 2022; Mateo Diaz et al., 2019, 2022a, 2022b). For education and training systems, it is critical to accurately and reliably measure these transversal skills to foster their development. However, in comparison to the assessment of content-based knowledge, assessment of 21st century skills is in its infancy (Care et al., 2018). In Latin America, few education systems systematically assess these skills among their student populations, leaving youth without a way to demonstrate these skills in the labor market. In this introduction, we will review the importance of 21st century skills and the challenges in measuring these skills.

The Importance of 21st Century Skills

Latin America's youth are growing up in a fast-changing world. Rapid technological changes are generating new markets and changing the way we work and interact with others. Automation, artificial intelligence, and robotics offer the potential to increase economic output and productivity, but also threaten to displace millions of workers in the region—putting at risk approximately 63% of jobs across Latin America and the Caribbean (Weller et al., 2019). Climate change is endangering livelihoods and markets—with the economic costs of climate change estimated to represent between 1.5% and 5% of the region's GDP by 2050 (United Nations, 2015). Most recently, the COVID-19 pandemic, an unprecedented health crisis, has exacerbated the region's deep-rooted challenges, exposing how urgent it is to invest in more resilient societies.

The challenges of today—and the unknowns of tomorrow—have redefined the skills that will be most valuable in our societies. In recent years, the return to noncognitive skills has risen as the return to traditional cognitive skills has remained stable (Deming, 2017; Edin et al., 2017). Our rapidly changing labor markets, in which more people will move in and out of career fields throughout their lives, require critical thinking skills, problem-solving competencies, and learnability. In an age marked by globalization and increased migratory flows, communication skills and openness are ever more necessary to thrive in increasingly diverse societies.

Are our education systems preparing our youth with the skills they need? Several sources indicate that our education and training systems are not up to par. Curricula often do not reflect the needs of the labor market, and many teachers still utilize archaic pedagogies, overly emphasizing passive forms of learning and failing to exploit the types of active learning most conducive to 21st century skills development (Bruns & Luque, 2015; Mateo Diaz et al., 2022b). It is no surprise that in international assessments of cognitive and non-cognitive skills, Latin American students fall behind the rest of the world. In the 2015 PISA examination of "collaborative problem-solving," six of the seven participating Latin American countries (the exception was Chile) scored in the bottom 20% of countries (Bos et al., 2017).

Investing in 21st century skills is critical. The Inter-American Development Bank defines 21st century skills, or transversal skills, as those that are: essential to human development, necessary for navigating healthy and productive lives, applicable in multiple contexts and transferrable from one field of life to another, and not specific to any particular job, field or discipline (Mateo Díaz et al., 2020). They refer to a broad set of skills that include advanced cognitive skills such as critical thinking, executive function skills such as self-regulation, and an array of socioemotional skills such as adaptability, motivation, and empathy.

A vast global literature suggests that 21st century skills are **strongly associated with a host of positive life outcomes** from educational attainment and productivity to mental and physical health and general wellbeing. In the academic realm, socioemotional competency positively predicts academic achievement, emotions and attitudes (including reduced learning anxiety) and educational attainment, in part by enabling students to set and work towards their academic goals and more adeptly overcome difficulties (Wang et al., 2019; Portela-Pino et al., 2021). Interventions to foster socioemotional skills development have been shown to enhance educational outcomes in studies from Zambia, India, Switzerland and others (Sorrenti et al., 2020; Ashraf et al., 2020; Edmonds et al., 2020). Socioemotional skills are also predictive of employability, wages, and productivity across a range of occupations. In Malawi and Ghana, higher non-cognitive skills among female farmers lead to improved agricultural productive decisions, such as the adoption of profitable cash crops and greater technical efficiency in production (Ali et al., 2020; Montalvao et al., 2017). In India and Colombia, studies focused on the garment production and retail sectors, respectively, find that on-the-job training that focuses on improving workers' "soft" skills, such as communication and time management, have large positive effects on worker productivity (Adhvaryu et al., 2018; Prada et al., 2019). Finally, higher levels of these skills predict a range of health-related outcomes: reduced drug and alcohol use, teenage pregnancy rates, mental difficulties, and others (Savelyev & Tan, 2019).

Challenges of Measuring 21st Century Skills

While the international community has largely reached a consensus on the importance of 21st century skills, much remains to be done to ensure our education systems are adequately teaching and assessing 21st century skills among learners (Care et al., 2018; Galloway et al., 2017). **Measuring 21st century skills is one of the most important tasks of education and training systems. Measurement is what allows program managers to evaluate the effectiveness of a particular program, diagnose the needs of an individual student, or assess their development over time.**

Still, countries around the world—including in Latin America—face significant challenges in the field of 21st century skills measurement (USAID, 2017; Kim & Care, 2020). In public education systems, assessments are almost entirely geared towards measuring subject-specific cognitive skills, such as reading, writing, and mathematics. Few, if any, systems include assessments of transversal skills. Even internationally, large-scale examinations have only recently begun to expand beyond these traditional subject-specific skills, leaving countries without internationally comparable data on transversal skills. **A critical challenge of measuring 21st century skills stems from the fact that most assessment instruments have been developed and tested for use in Western, high-income countries, which creates challenges for their applicability and the interpretability of their results in other contexts.**

Additionally, because of insufficient and inadequate ways to measure 21st century skills in Latin America, youth are often left with few ways to demonstrate these skills—which are often unobservable in the hiring process—to potential employers.

To address this issue, this brief answers the following questions: What assessment tools exist for measuring 21st century skills? What are some challenges regarding applying them in developing countries? How can policymakers make the best use of existing instruments? We answer these questions using an in-depth literature review, interviews with key stakeholders in the private, public and academic sectors (*N* = 12), and case studies (*N* = 3). Chapter 2 begins with a review of different types of 21st century skills assessment instruments and their pros and cons. Chapter 3 reviews common challenges to implementing these assessments in low- and middle-income countries and discusses possible approaches to addressing them. Chapter 4 presents three case studies to show how performance-based assessment instruments have been used around the world and to illustrate common challenges. Chapter 5 offers different options that policymakers and/or organizations can pursue to best invest in 21st century skills measurement, and also offers considerations for implementation. Finally, Annex B contains a short, non-exhaustive compilation of instruments.

2. Understanding Different Types of Skills Assessments

This chapter offers a review of three main types of 21st century skills assessment instruments and a discussion on the advantages and disadvantages of each.

There are three main types of 21st century skills assessment instruments

There are three broad types of skills measurement instruments (Galloway et al., 2017; Duckworth & Yaeger 2015). These are:

- Instruments based on self-rating or self-reporting;
- Instruments based on ratings of others; and
- Performance-based assessments, including tasks, simulations, and games

Instruments based on self-rating

Self-reporting and self-ratings usually take the form of a questionnaire that asks the respondent to rate themselves on several measures or indicate the extent to which they agree or disagree with several statements, usually following a 5-point scale (Galloway et al., 2017). For example, in the Big Five Inventory (BFI) instrument, one of the most widely used assessments of personality traits, a student might be asked how true statements such as 'worries a lot' or 'tends to be lazy,' are for them. Other types of questions sometimes included in these questionnaires are situational judgment tests, which ask respondents to indicate how they would act in a given situation (Galloway et al., 2017).

Instruments based on ratings of others

The second type of instrument involves ratings by others, including teachers, parents, observers, and peers. Ratings by others can take the form of questionnaires like the ones described above; a teacher, for example, rates a child based on their interactions with that student over time. This instrument might take the form of more complex exercises, such as observation exercises aided by an observation checklist.

Instruments based on performance

More recently developed instruments are performance-based measures. In the realm of traditional academic assessments, performance assessments are defined as those where students, rather than selecting from provided answer choices, must "construct an answer, produce a product, or perform an activity" (Darling-Hammond & Adamson, 2010). In the field of 21st century skills measurement, they take the form of any scenario that has been designed to elicit meaningful differences in particular behaviors (Duckworth & Yaeger, 2015). These scenarios may take the form of in-person or online simulations, online tasks, and video game-based tests. Another type of measure, usually used for formative purposes, is portfolio-based assessment, whereby teachers utilize rubrics to evaluate a set of student products (ATS2020, 2020).

No perfect measure: All types of instruments have limitations and advantages

Instruments based on self-rating

These instruments are quick, easy to implement, and—if delivered via paper—require no technological infrastructure for their application. At the same, however, they suffer from different sources of bias. As Duckworth and Yaeger (2015) explain, there is an opportunity for bias at virtually every stage of the test-taking process: from the moment a student reads and interprets an item, to the point when he or she selects and finalizes one of the available choices.

For instance, responses may suffer from reference bias if frames of reference differ systematically across respondents. As an example, respondents might compare themselves to their immediate social reference group, reducing the objectivity of their responses (Galloway et al., 2017). Additionally, if social norms for judging behavior are different between contexts, then responses are likely to vary. Studies have found, for example, that students attending higher-achieving schools may score themselves lower on measures such as self-control compared to students at other schools (Goldman, 2006, as cited in Duckworth & Yaeger, 2015). Students might simply have different interpretations of the pragmatic meaning of a question item based on cultural differences or individual-level idiosyncrasies. Another critical drawback is the effect of social desirability bias, or the tendency in most cultures to rate oneself in a way that is consistent with what is perceived as 'desirable' by others—which often means rating oneself highly on 'positive' attributes (Kreitchmann et al., 2019). Finally, other forms of bias involve memory bias—imperfect recollections of one's own behaviors or actions—and acquiescent responding, or the preference for the 'positive' side of the rating scale regardless of item content (Weijters et al., 2013).

Psychometricians have developed several strategies to address some of these issues, such as anchoring vignettes, forced-choice methods, and situational judgment tests. To address the issue of respondents (perhaps of different cultures) interpreting questions in different ways, anchoring vignettes present hypothetical situations and people that illustrate various skills levels. Respondents are asked to rate the vignettes on the same scale used for their self-reporting; then, by comparing the respondent's self-assessment to their assessment of the hypothetical person, researchers can 'recode' a respondent's self-assessment (King, n.d.). To address the issue of social desirability bias or outright 'faking,' forced-choice methods ask respondents to select which of two (or more) options best describes them, but make it unclear which option is the most 'desirable' one (Galloway et al., 2017). Several studies show that both forced-choice methods and anchoring vignettes can improve cross-country validity in self-reports (Weiss & Roberts, 2018; Avvisati et al., 2019).

Instruments based on ratings by others

Instruments based on reporting by others, particularly teachers, have the benefit of providing a "non-egocentric" perspective and the point of view of someone who interacts with other same-age students over many years (Duckworth & Yaeger, 2015). However, they suffer from many of the biases present in self-report measures (e.g., reference bias, memory bias) as well as unique challenges. For instance, teachers' reports at end-of-year assessments can be driven primarily by first impressions, downplaying progress throughout the year (Duckworth & Yaeger, 2015). Additionally, 21st century skills—particularly socioemotional skills—are manifested somewhat differently in different contexts; thus, parents, teachers, and observers' responses are influenced by their unique frames of reference. Reports by teachers, then, will be driven by observations of a child that are restricted to a school setting and that do not consider how a child may behave outside of school (Duckworth & Yaeger, 2015). This is why responses by teachers and parents can diverge significantly (Major et al., 2015). The same discrepancies can be found between teacher-report and student-report-based instruments. The figure below by Duckworth & Yaeger (2015) illustrates the process by which students and teachers respond to questionnaire items and why responses can differ. Finally, respondents can also be vulnerable to social desirability bias. In particular, if the use of these assessments is meant for high-stakes purposes, the possibility of manipulation (inflation of scores) by teachers, parents, or observers cannot be discounted.

Figure 1. The process by which students and teachers respond to questionnaire items

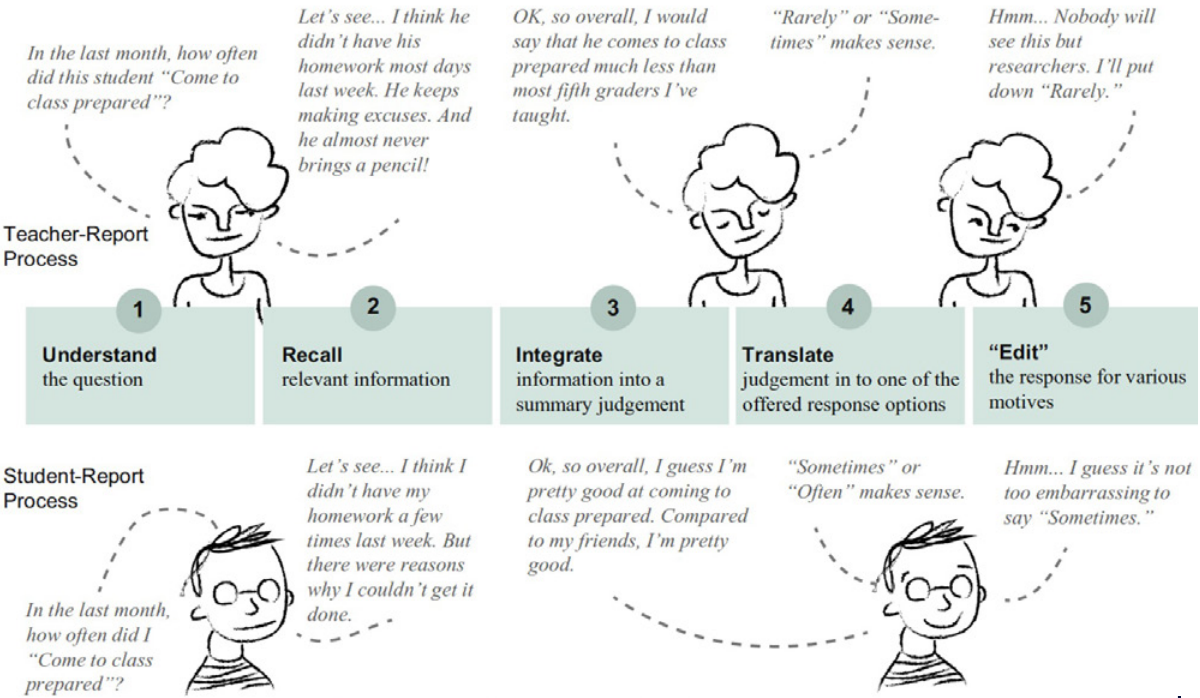

FIGURE 1. *The process by which students and teachers respond to questionnaire items.*

Source: Duckworth & Yaeger (2015).

Instruments based on performance

Performance-based assessments, which include online tasks, simulations, and games designed to elicit and observe particular behaviors among students, are other types of assessment instruments for measuring 21st century skills. Their greatest advantage is that they do not rely upon the subjective judgments of students or teachers, reducing the effects of social desirability bias and other effects that threaten the validity and reliability of instruments based on reporting or observation. By assessing behavior at a moment in time, these measures can also be more sensitive to subtle changes in behavior than questionnaires. By placing students in identical constructed scenarios or situations, they largely eliminate the effects of confounding differences in reference points (for instance, "time spent doing homework" varies for students who get assigned more or less homework) (Duckworth & Yaeger, 2015).

Game-based assessments have been one of the fastest-growing types of performance-based assessments over the past 10 years, both in education and human resources (Kim & Ifenthaler, 2019; Oranje et al., 2019). During gameplay, players produce a rich stream of performance data while they perform complex tasks, drawing on the skills or competencies that game developers want to assess (Shute & Ventura, 2013). The evidence needed to assess these skills is provided by the players' interactions during the game itself and is contrasted with a typical or expected outcome in a singular activity. For this reason, game-based instruments are sometimes called 'stealth assessments': the assessment is directly and 'invisibly' woven into a digital environment without disrupting gameplay (Shute & Ventura 2013; DeRosier et al., 2012).

Figure 2. User Interface in game-based assessment Hall of Heroes, produced by Centervention

Source: Centervention (n.d.).

Game-based assessments offer several advantages. First, they offer the possibility to assess 'hard-to-measure skills' by providing immersive, interactive environments where test-takers can show these skills rather than report on them (Oranje et al., 2019). Second, their sophisticated mechanics and interactive environments mean they are highly engaging to children, helping to produce a 'state of flow'—a state of heightened engagement and concentration that can improve the validity of assessments (Shute et al., 2016; Oranje et al., 2019). This heightened engagement can also help to reduce test anxiety that is commonly associated with conventional assessment methods, which also helps to improve the accuracy of results (Ren, 2019). Additionally, given their dynamic nature, games could be built to measure competencies over time, or to adjust the difficulty of the tasks in response to a players' current level of skill (at sub-skill or overall levels) (Shute et al., 2016). Finally, from an administrative perspective, once an assessment is developed, the time spent on training for using the system is minimal. Data analysis is usually built to be automated, which represents an advantage over traditional assessments (DeRosier et al., 2012).

The advantages of games and other performance-based assessments must be considered alongside their limitations. Experts argue that performance in a game-based assessment depends not only on students' differing skills, but also on other aspects that in theory should not impact results. These include familiarity with the game setting or with technology in general, the language used, and cultural aspects embedded in the game. Other threats to validity include how a particular player reacts to the game mechanics, the storyline, or other challenges of the game itself (Sandberg, 2016). It is for these reasons that game-based assessments have typically been utilized for formative purposes (for providing teachers and students feedback on student performance) rather than for summative purposes or high-stakes purposes (making decisions about grade promotion, certification, etc.) (Oranje et al., 2019).

Table 1 summarizes the discussion above on the advantages and disadvantages of these three assessment types, while Annex B contains a short compilation of these different instruments.

Table 1. OECD rubrics on creativity (domain-general, comprehensive, and class-friendly)

Type of Instrument	Advantages	Disadvantages
Instruments based on self-report	Several instruments are publicly available and free to use, thereby reducing costs Can be administered via paper, eliminating the need for technological infrastructure Are prevalent and well-studied (Galloway et al., 2017).	Results can be influenced by respondent bias, for instance: reference bias (frames of reference differ across respondents), social desirability bias (providing inaccurate answers that seem desirable), memory bias (imperfect recollections), etc. (Kreitchmann et al., 2019).
Instruments based on ratings of others	Can provide a "non-egocentric" perspective (Duckworth & Yaeger, 2015). Can provide the point of view of someone who interacts with other same-age students over many years (Duckworth & Yaeger, 2015).	Responses are vulnerable to many of the same biases as instruments based on self-report Responses may vary depending on whom the instrument is applied (teachers, parents, observers), since each observes the child in different contexts (Major et al., 2015).
Performance-based instruments	Do not rely upon the subjective judgments of individuals Create highly engaging environments for children and youth, which can improve test score validity and reduce test anxiety (Oranje et al., 2019) Little time spent on training test administrators	Performance can be influenced by factors unrelated to skill levels, such as familiarity with technology, hand-eye coordination, physiological state (e.g., hunger, fatigue) (Sandberg, 2016) Require technological infrastructure such as Internet, digital devices, etc.

Instruments based on self-rating are the most prevalent type, while performance-based measures are the fastest-growing

Despite their limitations, instruments based on self-report are the most prevalent and well-studied of all types of instruments. An in-depth analysis by USAID that analyzed instruments measuring soft skills and life skills among youth found that most of the tools utilized in youth programs around the world are self-report questionnaires (Galloway et al., 2017). Of a 74-item list of instruments they selected based on a set of quality criteria, all but nine instruments were based on self-report. Among their 10 highest-rated instruments (rated on criteria such as evidence of reliability and validity, ease of administration, and evidence of international use), all but one are based on self-report (see Table 2). Still, because of the known psychometric limitations of these tools, Galloway et al. (2017) describe the predominance of these tools as a 'limitation' in the literature.

Table 2. Ten Top-Rated Instruments for Measuring Soft Skills among Youth (Galloway et al., 2017)

Assessment	Type of Assessment	Description
California Healthy Kids Survey, Social and Emotional Health Module	Self-report questionnaire	This 21-item questionnaire is used by school districts in California to measure key soft skills among students 10-19.
Chinese Positive Youth Development Scale (CPYDS)	Self-report questionnaire	The CPYD addresses several priority skills for youth development, including positive self-concept, self-control, and goal orientation. It has been used as part of monitoring and evaluation for the Hong Kong-based Project P.A.T.H.S.
SENNA 1.0 & SENNA 2.0 (two instruments)	Self-report questionnaire	The SENNA instrument was developed to assess socioemotional skills in Brazil in group populations and has been used for 5th, 10th, and 12th graders. A second edition, released in 2016, adds items to measure self-efficacy, making the tool more complete.
Child and Adolescent Wellness Scale (CAWS)	Self-report questionnaire	CAWS is a self-report instrument that assesses social and emotional competencies. It has been validated for use among adolescents ages 12-18 and usage of this instrument has been recorded in the US, Japan, South Korea, Taiwan & Thailand.
The Big Five Inventory (BFI)	Self-report questionnaire	The Big Five Inventory (BFI) is one of the most widely used instruments for identifying personality factors: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism.

Anchored BFI	Self-report questionnaire	The Anchored BFI tool uses the same instrument as the BFI tool, but also includes anchoring vignettes (or scenarios) as well as situational judgment tests.
Jamaica Youth Survey	Self-report questionnaire	The Big Five Inventory (BFI) is one of the most widely used instruments for identifying personality factors: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism.
Anchored BFI	Self-report questionnaire	The Jamaica Youth Survey was developed as part of a World Bank-funded project to evaluate programs for promoting well-being among teenagers (12-18) in Jamaica. It addresses key skills such as positive self-concept, higher-order thinking, and social skills.
Responses to Stress Questionnaire (RSQ)	Self-report questionnaire	The RSQ measures coping and involuntary stress responses; participants rate how often they use each coping method or experience each type of involuntary stress response on a 1 to 4 scale. It has been used internationally with respondents over age 9.
KnackApp	Self-report questionnaire	KnackApp is a game-based assessment of soft skills or 'knacks' including positive attitude, responsibility, and higher-order thinking. The game uses machine learning and AI to analyze gameplay data that is used to score players across different dimensions.

Source: Adapted from Galloway, T, Lippman, L., Burke, H., Diener, O., and Gates, S. (2017). Measuring Soft Skills & Life Skills in International Youth Development Programs: A Review and Inventory of Tools. U.S. Agency for International Development.

3. Assessment in Low- and Middle-Income Countries: Challenges and Potential Solutions

Having reviewed the different types of 21st century skills assessment instruments, this chapter will review common challenges to implementing these assessments in low- and middle-income countries, and discuss possible approaches to addressing these challenges. These findings draw heavily from interviews held with twelve (12) stakeholders, which included: (a) assessment developers from both the private sector and academia in the U.S and Latin America; and (b) institutional users who have utilized skills assessment instruments among youth. This chapter focuses on three, commonly cited challenges: (1) poor fit of existing instruments in a new context, (2) technological infrastructure, and (3) different levels of prior exposure to technologies and games.

Challenge 1: Poor fit of existing instruments in a new context

Most skills assessment instruments have been developed and tested in Western, developed countries. When these same instruments are applied to developing or non-Western contexts, this can lead to multiple issues that compromise the validity of the instrument for such contexts. In all, issues of language and literacy, cultural norms, and colloquialisms can all compromise results for an instrument (Duckworth & Yaeger, 2015). In the case of instruments based on self-reporting or reports by others, because measures are not anchored on "objective" phenomena, users can have different interpretations of the meaning of measures such as 'perseverance' and 'self-discipline, and thus respond to test items in different ways. On a more basic level, the language itself may be confusing if it is presented to second-language speakers, or if it is presented to individuals with low levels of literacy. These issues can also appear in performance-based assessments (such as games) containing little or no text, if the storyline, interface, or other aspects of the game are not a good cultural fit for the context where the game is played.

Interviews conducted as part of this project exemplified many of these issues. Researcher Sarah Pouezevara from RTI International, who deployed a U.S.-made self-assessment questionnaire of soft skills in the Philippines, theorized that a test item, such as 'I tend to leave a mess in my room' may elicit different responses because "Americans might have a bigger tolerance for mess than Filipinos or vice versa" (S. Pouezevara, personal communication, 2021). Similarly, the director of a Haiti-based youth training non-profit said that he discontinued his use of a commonly-used U.S.-developed skills assessment because test-takers struggled with cultural references, and because the test "assumed familiarity with the language and colloquial phrases" that youth in his program did not understand. These issues can happen even with tests containing little or no text. In the case of KnackApp, a game-based soft skills assessment that only includes one line of text in the instructions, users in Colombia mentioned that having the instructions in English was a barrier (J. Gaviria Gómez, personal communication, 2021).

Approaches

How have practitioners and assessment developers approached the issue of contextual fit when deploying instruments in non-Western contexts?

Approach 1. Adapt and validate an existing instrument for use in a particular context
Most self-report instruments undergo a rigorous adaptation, testing, and validation process when utilized across different contexts. For instance, the global humanitarian agency World Vision utilizes the Search Institute's Developmental Assets Profile (DAP) assessment of socioemotional skills in its youth development work. The organization has adapted and created local language versions of the test for 15 different countries using a standardized process that involves:

(1) focus groups with local youth to confirm the cultural relevance of targeted concepts;
(2) a translation exercise conducted in collaboration with local "youth experts";
(3) pilot testing with 100 youth, followed by adjustments to the tool; and
(4) field testing with 300 youth (USAID, 2017).

In a more extensive adaptation process, the Education Development Center created an improved version of the widely-used Big Five Inventory (BFI) assessment by developing two test item types to be applied alongside the BFI—anchoring vignettes and situational judgment tests—to create the Anchored BFI tool. They then implemented the tool in the Philippines and Rwanda through a process that involved translation into the local languages, cognitive labs, pilots, and revisions to the tool. The addition of anchoring vignettes and situational judgment tests were found to improve reliability (so that the test would be as reliable in the two countries as in the U.S.), and reduce measurement differences between the countries, among other improvements (Education Development Center, 2016).

An instrument can sometimes be adapted for use across a region. In 2009, the Costa Rican Ministry of Education and the Omar Dengo Foundation collaborated with the Assessment and Teaching of 21st Century Skills ATC21STM global consortium to build a Latin American version of the task-based, computerized assessment of 21st century skills. A research team translated the assessment into standard Latin American Spanish and created new content. In a portion of the assessment titled Webspiration, which is devoted to poetry, researchers selected 3 poems in Spanish to replace those in the original English-language task, consulting with Spanish literature experts to select poems that could be used throughout the region (Care et al., 2018). Other parts of the assessment were also localized such that the test could be used in other Latin American countries (see Case 1).

Approach 2. Use an instrument requiring minimal or no adaptation

Opting for instruments requiring minimal adaptation is easiest in the case of game-based or simulation-based assessments that have little or no text and where this text can easily be translated by the developer to other languages at low cost. For instance, the game-based test KnackApp, which brands itself as a "language agnostic" tool and uses only 1 line of text in the instructions, will soon be available in 20 languages (G. Halfteck, personal communication, 2021). Similarly, Roblox's game-based assessment of soft skills, used worldwide in McKinsey & Co.'s standardized hiring process, has been translated into a handful of languages (including two forms of Spanish) at the request of local McKinsey offices; they wanted to make sure that differences in English proficiency levels did not affect test scores (J. Buckley, personal communication, 2021). However, beyond a simple translation, developers did not make any changes to the game itself.

Finally, it is sometimes possible to utilize a game-based assessment in new contexts without making any adaptations to the instrument when it is applied among students proficient in the language of the test. The game-based assessment of socioemotional skills Hall of Heroes, which is produced by Centervention and available only in English, was tested by researchers in schools in the U.S., Malaysia, and South Africa as part of a study to compare socioemotional skills across cultures. In Malaysia and South Africa, the researchers chose English-instruction schools and students who had received at least 8 years of instruction in English; they utilized the same version of the test across all countries (Irava et al., 2019).

Approach 3. Develop a new instrument, ideally with the built-in ability for customization

Other organizations have opted to invest in the development of their instruments, often to use them at a regional or global level. An example is seen in the case of the test on computational thinking that is currently being developed by the Universidad de la República in Uruguay in collaboration with university researchers in Mexico and Paraguay. The researchers, with support from a contracted software developer, are building a video game-based assessment of computational thinking skills that could be used throughout the region. In explaining why they decided to create their own instrument instead of repurposing an existing one, the researchers mentioned three factors: the unique educational priorities of the region; the need for an instrument that could be made available at no cost to other countries; and the need for an instrument that could be adaptable to countries' realities (E. Angeriz, personal communication, 2021). While the game's premise and tasks will be the same throughout the region, the researchers are designing the game so that it is easy to contextualize to individual countries, e.g., local idioms/dialects and cultural references such as the names of local topography. The game will be validated through trials in different countries, starting in Montevideo, Uruguay.

In another example, RTI International invested between 2016 and 2018 in the development and testing of a game-based assessment of 21st century skills titled Home Alone. The game was built to be used in RTI's client countries throughout the world, but was also built to be customizable. For instance, it was built with a translation interface that allows developers to change the language of the text, and with a functionality that can allow designers to replace images in the game if this is needed to make it more contextually appropriate. As their researcher explained, a developer could "build the characters in their three or four different poses, and then you can swap them out so that they are wearing different clothes, for example" (S. Pouezevara, personal communication, 2021).

Challenge 2 - Technological Infrastructure

When implementing digital assessments, technological issues are some of the most commonly cited. Poor Internet connections are a major barrier for completing Internet-enabled assessments, both in schools and in homes where users engage in assessments on their own devices. A Haiti-based NGO leader described issues when asking its hiring candidates to take a game-based assessment from home: "We had people in the middle of a game run out of electricity, run out of battery, lose Internet, [and] they can't upload their results."

Another issue arises when the devices available for test-taking are not compatible with the test. This was seen in Chile during the implementation of the SIMCE TIC test of digital abilities, which was produced by the Catholic University of Chile for the Ministry of Education. Many school computers were low-quality or poorly maintained, and all ran on different operating systems, which meant not all computers were appropriate for use during the test. This led the team to conduct technical checks on the functional computers before the test and "embargo" the computers from use until the day of the test (F. Cortés, personal communication, 2021).

Lastly, users sometimes do not have a smartphone that can support a digital assessment. In India and Colombia, a major issue in the use of the KnackApp soft skills assessments was that users did not possess an appropriate smartphone (or a smartphone at all) to take the test from home.

Approaches:

Approaches to Internet connectivity issues

- **Build or select a test that has offline playability and 'asynchronous scoring.'** A game can be played offline such that data is stored locally and can later be synced to a central database (when the device is connected to the Internet) for data analysis and scoring. However, this approach may not be appropriate if the assessment used is adaptive; that is, where the questions shown on the test depend on a student's answers to prior questions. This requires the device to be connected to the scoring system in real time (J. Buckley, personal communication, 2021).

Approaches to device compatibility issues

- **Test an assessment instrument** on multiple types of devices during pilots and trials

- **Select or develop an assessment** that can run on different types of devices and operating systems, and does not require the latest technologies (i.e., can be used on devices that are a few years old) (G. Halfteck, personal communication, 2021).

- In the case of school-based assessments, one approach is to **provide the devices** when the school does not have sufficient and/or appropriate ones. In the example of SIMCE TIC in Chile, because many schools lacked computers (or adequate computers), the Ministry of Education deployed hundreds of notebook devices (F. Cortés, personal communication, 2021).

Approaches to users' lack of smart devices or Internet connection

- In cases where users are asked to take a test at home but lack the connectivity or necessary device, three organizations interviewed for this project mentioned that they a**sked the youth to come into their offices to use their Internet and/or use devices provided by the entity**.

Challenge 3 - Different Levels of Prior Exposure to Technologies & Games

Another challenge to using digital assessments of 21st century skills is the effect that different levels of exposure to technology have on performance. Studies from the United States have found that prior experience with gaming technologies can influence performance in game-based assessments (Ren, 2019; Mayer, 2018). In low- and middle-income contexts, where disparities in access to technologies are particularly significant, this can have a particularly strong effect on the accuracy of results. In Chile, performance in the SIMCE TIC 2013 exam was found to be influenced by students' prior experience with computers, as reported in demographic questionnaires (F. Gomez, personal communication, 2021). Similarly, the effects of prior exposure to video games were one of the reasons why researchers at RTI cautioned against the use of their game-based assessment for high-stakes purposes (S. Pouezevara, personal communication, 2021).

A lack of familiarity with games can also cause users to feel skeptical about the use of these tools for high-stakes purposes. In Colombia, a private nonprofit using a game-based tool for hiring mentioned the case of a woman who said that, because she had never played video games, she expected she would do very poorly and wanted to quit the process. Program staff mention that they have had to educate test-takers about the test and the science behind it to build confidence in the test. "We have to debunk the myth that we can only hire using traditional assessments," they said (J. Gaviria Gómez, personal communication, 2021).

Approaches

- **Statistically control** for the effects of gaming experience or prior exposure to technology as confounding variables when analyzing results (Ren, 2019).

- **Level the playing field through warm-up tutorials**. One way to make results less reliant on familiarity with video games is to introduce a warm-up period before beginning data collection (Pouezevara et al., 2019). For instance, Kim and Shute (2015) tested a game-based assessment that began with 10 minutes of warm-up, including an interactive tutorial to learn the mechanics of the game and a tutorial video about the characters in the game.

- **Utilize a game that can adapt to skill levels**. It is ideal to ensure that each student's gaming experience is adequately matched to their skill level. If a player's skill is too great for the task, they might become bored; if it's too low for the task, they might become anxious or frustrated (Oranje et al., 2019). Some approaches include allowing players to choose their path and levels of challenge during gameplay or utilizing an adaptive game that naturally adjusts based on performance level (Kim & Shute 2015). More recently, games are being developed that can utilize facial tracking to detect the mood of the player (Verma et al., 2019).

4. Case Studies

I use three case studies to showcase how performance-based assessment instruments have been used around the world, and to illustrate and exemplify the challenges that are described in the preceding chapter relating to the use of 21st century skills assessments in low- and middle-income countries. The first case study examines the use of an existing performance-based instrument applied without modifications. The second case study describes the application of an instrument after being modified for use in Spanish-speaking Latin America and the Caribbean. Lastly, the third case study describes a game-based instrument conceptualized, developed, and tested for use around the world. These case studies are based on interviews with key stakeholders as well as existing academic literature on these instruments.

CASE 1: Assessing Occupational Fit in India
KnackApp & the National Skills Development Corporation (NSDC)

In 2019, India's National Skills Development Corporation—part of the Ministry of Skill Development and Entrepreneurship—partnered with predictive talent analytics company KnackApp to provide vocational counseling to over 8,000 students attending the NSDC's vocational training centers.

KnackApp produces game-based applications that leverage behavioral science, artificial intelligence, and machine learning to assess users on several different skills and aptitudes. As a user plays one of KnackApp's games, which measure how people behave and perform across different situations, they produce gameplay data that are used to score the individual across different dimensions: cognitive abilities, personality traits, emotional and social skills, mindsets, and aptitudes. With these scores, KnackApp then builds models to predict job performance in a broad range of occupations.

The partnership between KnackApp and India's NSDC was forged as a way to tackle an important issue at the NSDC: a number of their trainees were quitting their first jobs within a few months of starting, often stating that they could not "cope" with the job or that it was not a "good match" for them. To address this problem, vocational counselors across the institution's skills training centers began to utilize KnackApp as a vocational guidance tool aimed at helping youth to better understand their talents and make an informed choice about which training track to enroll in. The NSDC is now collecting data to understand whether the training and employment outcomes of trainees who selected one of the career paths 'recommended' by KnackApp differed from those of trainees who 'went with their gut' and chose other paths. Some outcomes they are observing include whether trainees are passing their final certification assessments, obtaining employment, and staying employed. As their project coordinator put it, this work was done for research purposes rather than for a "prescriptive" aim (A. Kashyap, personal communication, 2021).

Program coordinators report having had a positive experience with KnackApp, but also point to important challenges. A key challenge is in finding alignment between the skills or 'knacks' that KnackApp measures and the skills that are part of India's robust National Qualifications Framework (NQF). Under the framework, each recognized job role in India is assigned a Qualification Pack that is comprised of a short set of National Occupation Standards, each of which specifies the standard of performance a person must have for a particular job function. Hence, program coordinators found it difficult to perform one-on-one mapping of the app's 'knacks' and the standards in each occupation, which complicated vocational counseling.

Another important insight regarding the potential uses of app results for credentialing was program coordinators' belief that using KnackApp would not be effective by itself, since Indian employers place great value on the skills certifications that are performed and issued by industry bodies and based on the NQF. Their coordinator noted that "employers don't want to look at the result of one app per se; they want to look at what happens when a candidate goes through the training and what the outcomes are" (A. Kashyap, personal communication, 2021). For that reason, the focus for now is on encouraging more Indian youth to make informed career choices and understanding KnackApp's effects on their choices.

Source: A. Kashyap, personal communication, 2021.

Figure 3. Users in rural India using KnackApp

Credit: Guy Halfteck, knackit.com.

CASE 2: Adapting an international instrument to the Latin American context Assessment and Teaching of 21st Century Skills (ATC21STM)

Assessment and Teaching of 21st Century Skills (ATC21STM) was created in 2008 as a global collaboration, spearheaded by the University of Melbourne, in collaboration with Cisco, Intel, Microsoft and six participating governments, and with the support of the Inter-American Development Bank. It assembled over 200 researchers from around the world to develop the ATC21S test, a formative, performance-based digital assessment of 21st century skills designed to be used globally. The test measures two domains: (1) Collaborative Problem-Solving (CPS) and (2) Learning in Digital Networks (LDN) - a domain that encompasses information literacy, digital competence, learning in networks, and technological awareness.

To develop the tool, this global team organized itself into working groups to first select and define the particular skills it would measure, develop a methodological approach, design the technological infrastructure, and construct individual tasks. Once "prototype tasks" were developed, they were tested iteratively between 2009-12 across all six participating countries: Australia, Finland, Netherlands, Singapore, USA, and Costa Rica. In Costa Rica, which was selected as the lead country for the ATC21S Latin American Chapter, the project was headed by the Ministry of Education and the Omar Dengo Foundation. Costa Rica's experience showcases the process of adapting and localizing an instrument for use in LAC and the challenges of implementing an online assessment in public schools in the region.

Costa Rica's design and the testing process began by localizing the prototype tasks for use throughout the Latin American region. Beyond the translation of all tasks to "standard" Spanish, two tasks developed to test LDN required additional work. These two tasks—the Arctic Trek and Webspiration (a poetry task)—both required students to explore external websites, but it became difficult to find Spanish-language web sources similar enough in difficulty to the English versions. Hence, for the Arctic Trek exercise, the Costa Rican team—in collaboration with science advisors—opted to search for a different theme for which online resources did exist in Spanish, ultimately settling on the Antarctic (Care et al., 2018). Once the new online resources were selected based on a set of criteria, the team modified the tasks and illustrations used in the test to correspond to the new theme. For the Webspiration task, the country research team searched for three poems in Spanish to replace those in the original task (Care et al., 2018). After consulting with Spanish literature experts, shortlisting a set of regionally representative poems, and conducting cognitive labs with children to finalize poem selections, the team then modified the text in the assessment tasks (questions) to match the new poems. Finally, because the original assessment asked students to visit the English-language website favoritepoem.org., which features "Americans saying poems they love," the LAC team decided to make their own videos, paying attention to cultural and ethnic diversity and showcasing speakers from different countries (Care et al., 2018).

With the assessment translated into Spanish and localized for use in LAC, the country team began its iterative testing process. The process, which involved six different stages, took place over a few years. The stages were: (1) concept checks for the CPS tasks; (2) cognitive laboratories for CPS tasks; (3) pilots for CPS tasks in four schools; (4) two trials for the CPS tasks involving 500 and 593 students, respectively, (5) cognitive labs for LDN tasks; and (6) another set of trials combining CPS and LDN tasks (with 776 students) (Care et al., 2018). In all trials, teachers received rigorous training regarding the administration guidelines. Feedback showed that students and teachers responded positively to the assessment, finding it challenging but fun (Care et al., 2018).

Use of results in Costa Rica: Results from the test were utilized to provide the government a baseline assessment of problem-solving and digital skills in its schools. Because researchers had tested the assessment in a sample comprised only of schools that were a part of the Ministry of Education's PRONIE digital learning initiative, this exercise was intended as one exploratory approach to assessing the impact of the PRONIE program. Magaly Zúñiga, director of Evaluation and Research at the Omar Dengo Foundation and an original member of the implementation team, says her organization also sees the ATC21S as a way to measure skills among Costa Rican students every few years. Although an attempt in 2018 to use the test as part of a quasi-experimental assessment of PRONIE was postponed, the team hopes to be able to conduct this kind of evaluation in the future.

Application in Buenos Aires: In 2018, the ATC21S instrument was also applied in the City of Buenos Aires as part of an effort to evaluate the effectiveness of the City's Digital Education Plan (PIED), which seeks to enrich learning and teaching through the use of digital devices. Researchers noted that there was an overlap between PIED's competency framework and the competencies targeted in the ATC21S tool (Grunstein et al., 2019). The evaluation involved 1,225 students across 46 schools that are part of a subset of the PIED program (called Plan Sarmiento, which focuses on equipping schools with digital technologies). Two weeks after the examination, school principals and teachers received performance reports that described findings and provided guidance for improvement (Grunstein et al., 2019).

Lessons Learned

Low Internet bandwidth cited as the greatest challenge: The research team in Costa Rica cites Internet bandwidth as their biggest hurdle. The research team chose a sample of schools with sufficient Internet bandwidth to support the online assessment, but in the final trials, 26% of students' scores were lost to connectivity issues and non-completion (Care et al., 2018). "At one point we determined that only eighteen students could take the test at a time, but sometimes, depending on the time of the day, it would be as low as ten," Zúñiga explained. However, she says these insights from the trials will prepare them to deploy the instrument in the future.

Technical assistance from the international team was critical: The translation and contextualization of the assessment to the Latin American context was made possible thanks to a close collaboration with the global ATC21S team for the data analysis and validation process. Additionally, it was critical to have technical support (a school psychologist and remote IT support) during the pilots and trials to navigate issues during implementation.

Use of results as a proxy for program effects: Researchers find that this assessment serves as a way to provide evidence that the Ministry of Education's investments in digital learning have effects on students' problem-solving and digital skills. In Costa Rica, they were able to do this after comparing the PRONIE program's learning standards and ACTS21 skills maps and finding a large amount of overlap between the two.

Sources: Care et al., 2018; Grunstein et al., 2019; M. Zúñiga, personal communication, 2021.

CASE 3: Developing a game-based assessment for use in global contexts Inconclusive results from RTI's "Home Alone"

From 2016 to 2018, researchers at RTI International developed and piloted a game-based assessment designed to measure problem-solving skills by gathering metrics on several sub-measures of problem-solving: task completion, time management, accepting instruction, problem and solution identification, and self-regulation. "Home Alone," an assessment designed for use around the world, is a game in which a player takes the role of an older sibling who needs to take care of two younger siblings while the parents are out: shopping for meal ingredients, cooking, and tending to the sibling's needs. This project highlights lessons and challenges in developing new performance-based assessments.

Development and Testing Process

Selecting assessment domains and developing tasks: The team began the development process by convening three expert panels to determine the assessment domain they would ultimately focus on (problem-solving), define the domain, and deconstruct the sub-domains that the game would be built to measure. Then, the team at RTI, in conjunction with a contracted game developer, conducted focus groups with a diverse group of 15-year-old students in Morocco to develop possible content scenarios for the game and gain their feedback on relevant problem-solving skills.

Pilot tests in Morocco and the Philippines: The testing phase began with an initial, 100-participant pilot in Morocco that tested game functionality and areas of improvement. In addition to playing the game, students took a math and reading assessment and completed a 30-item self-report questionnaire (i.e., academic mindset, executive functioning, task completion). Parents completed an identical questionnaire to indicate their evaluation of their children. In a second pilot in Manila, Philippines, children completed two skills assessment questionnaires (including the Raven's Standard Progressive Matrices) as one way to measure the validity of the game (i.e., by establishing the extent to which the game's measures correlated with measures from well-studied assessments). They also completed an assessment aimed at understanding their experience with videogames, to test whether prior exposure to videogames affected performance (Pouezevara et al., 2019).

The testing process revealed several implementation challenges. For instance, technological challenges in the Philippines pilot caused the game to freeze and forced many students to have to restart the game an average of 2.3 times. Additionally, the principal researcher notes that proctoring the exam was an "unexpected challenge" given students' natural propensity to want to work together during gameplay (S. Pouezevara, personal communication, 2021). As she explained: "I think the hardest part is having kids not talk to each other. Games are naturally social, especially nowadays with massive multiplayer games [...] so they all wanted to fit in small groups and have fun with the game."

Discussion

Despite these challenges, researchers conclude that the 'Home Alone' game works as designed: getting students to make decisions to complete tasks in the game. Participant feedback found that students found the game engaging and relevant to their lives. An encouraging finding is that there are patterns and variations in the game play data that suggest the data 'differentiate between students based on their game play' (Pouezevara et al., 2019). Researchers also found acceptable levels of internal reliability between most variables collected and found that math and literacy skills had low correlations with the game scores. However, they point out three major limitations. First, they do find low correlations with the other well-studied skills assessment questionnaires (Raven's). Second, the researchers believe there could have been interference from students' prior gaming experience. Third, they feel that they simply do not have enough data to claim there is sufficient evidence of reliability. Without a very large data set and many more pilots, "our work is not quite finished," Pouezevara explained. In all, the research team concludes that they "don't have sufficient evidence yet to say that the game is a valid measure of problem-solving skills" (Pouezevara et al., 2019). Hence, they recommend its use as a learning tool or a source for discussion, but discourage its use for 'high-stakes' purposes.

Lessons Learned

- **Involve youth in the design process:** Researchers credit the positive feedback they received from users on their user-centered design process with focus groups to brainstorm scenarios to include in the game.

- **Strategies for mitigating technological challenges:** This pilot study provided insights on how to minimize technological challenges: the importance of (1) building games so that they can be played offline as well as on the Internet, and can be re-started in case of crashes; and (2) conducting testing across a wide range of platforms and devices as technology changes.

- **Caution in the use of results:** The game requires large amounts of data and several iterations before it can be considered a valid measure of problem-solving skills. Without this, it should only be used for formative assessment purposes or learning.

Sources: Pouezevara et al., 2019; S. Pouezevara, personal communication, 2021.

Figure 4. Examples of tasks in RTI International's Home Alone game-based tool

Scenario:

- You are left home alone and must take care of 2 younger siblings.

Chores:

- Plan and make breakfast, lunch, and dinner
- Clear the table & wash dishes
- Clean the house
- Feed the animals (cat & chickens)

Avoid distractions:

- Neighbors who want to play
- Chasing chickens
- Playing with the cat

Source: Pouezevara (2019).

In this chapter, we presented three case studies to showcase how performance-based assessment instruments have been used around the world, and understand common challenges. Important takeaways include the intensive and iterative nature of the assessment development and testing process; the importance of establishing strategies to mitigate any technological issues; and the need to exercise caution in the interpretation and use of results. The next and final chapter explores options for policy action and considerations to keep in mind when implementing 21st century skills assessments.

5. Implications: What are the key takeaways?

This chapter presents possible pathways for action that policymakers could pursue when investing in 21st century skills measurement, highlighting short-, medium- and long-term options as well as the pros and cons of each. The chapter also discusses considerations for implementation, including possible criteria for the selection of existing instruments.

Options for Action

The following are four different pathways that policymakers could pursue in investing in 21st century skills measurement.

Short Term: Utilize Existing Instruments Requiring Minor or No Modifications

In the short term, policymakers could begin by utilizing assessment instruments that require minor or no modifications. This could be the most cost-effective way for them to gain insights and feedback on the assessment-taking process before investing time and resources in other forms of assessments.

Approach 1: Utilize instruments that are available in Spanish and have been validated for use in Latin America. It is likely that the majority of these will take the form of self-report-based assessments that have either been developed for the region or have been translated into Spanish and used in the region by NGOs. See Annex B for a short list of instruments developed in the Latin American and Caribbean region.

Examples

Emotional Regulation Scale (RE-MESACTS), developed by the Latin American technical cooperation body on transversal and socioemotional skills (MESACTS) and validated in Argentina, Chile, Colombia, El Salvador, Paraguay and Uruguay (Comité Internacional para el Desarrollo de los Pueblos, 2020).

Assessment and Teaching of 21st Century Skills (ATC21S): A web-based assessment of Collaborative Problem-Solving and Learning in Digital Networks that has been adapted for use in LAC and applied in Costa Rica and Buenos Aires (see Case 2).

Approach 2: Utilize instruments that require only minor modifications and do not require extensive testing. These could take the form of game-based assessments that are not language-intensive, and which have already been used around the world.

Examples

KnackApp - a game-based assessment designed to measure different skills and aptitudes that has been used in more than 100 countries and tested at a global scale; soon to be available in Spanish (see case 1).

Pros:

- Policymakers can utilize the assessments in the short term without requiring an extensive testing and validation period.
- Most assessments that are available in Spanish are based on self-report, a type of assessment that is relatively easy to implement.

Cons:

- Most assessments available in Spanish are based on self-report, a type of assessment that is known to suffer from several important biases.
- Most assessments that are immediately available for use are created for formative rather than summative purposes, which means their applicability for high-stakes purposes (such as skills certification) is unclear.

Medium-Term: Adapt and validate existing instruments for use in LAC

In the medium-term, policymakers can also select high-quality existing instruments, adapt them to the Latin American context, and conduct validation studies. Applying an existing instrument to a new setting requires testing for its validity and reliability in a particular context and a particular purpose. As shown earlier, this process usually entails item pre-pilot translation modification (including with youth focus groups), a pilot application and field tests, local validity assessment of results, and post-pilot focus groups.

This type of adaptation exercise can be conducted for all types of instruments. In the case of instruments based on reporting, adaptation can involve translation into Spanish, and reviewing individual test items for their appropriateness in particular contexts. This work could potentially be done in-house (with the right team) or outsourced to psychometric specialists at universities or assessment companies. In the case of performance-based assessments, two potential options are: (1) work with educational assessment companies to commission the adaptation and testing of existing performance-based assessments for LAC contexts, or (2) work with companies in predictive hiring analytics to explore the possibility of adapting their assessments to school contexts (school-aged test-takers) and the LAC context.

Examples

Performance-based assessments:

Roblox - Roblox (formerly Imbellus) is the creator of the McKinsey game of problem-solving skills used by McKinsey & Co. in its global hiring processes. In 2019 they utilized grant funds from philanthropic foundations to adapt their existing problem-solving tasks for use in U.S. schools, making the test shorter and the difficulty levels appropriate for typical U.S 11th grade students (Buckley & Kim, 2020; J. Buckley, personal communication, 2021).

Assessments based on self-report:

WorkLinks Skills and Values Assessment (WLSVA) is a free tool developed by World Learning in 2020 to measure constructs such as consciousness, self-efficacy, and interpersonal skills, and certain civic values. It measures individual- and group-level change over time. Its copyright under a Creative Commons license allows others to translate the tool for non-commercial use, as long as they credit World Learning and license their adaptation under identical terms. The developers also describe the full validation process they conducted in Algeria, to serve as guidance for other countries (World Learning, n.d.).

Pros:

- Adapting instruments that have already undergone a validation process is cheaper and faster than developing instruments 'from scratch' (Cook & Hatala, 2016).
- Because these instruments have already been tested in other contexts, standardized validation guidelines or processes already exist.

Cons:

- Compared to the option of developing the policymakers' assessment tools (see below), this option may involve high recurring fees to the educational assessment company that produces the instruments.

Medium-Term: Evaluate instruments that are still in experimental stages

In the medium term, policymakers can also invest in the evaluation of instruments that are still in their developmental stages. This includes: a) instruments developed by other organizations that may have been tested in some contexts but that still require more testing, or b) assessments created by educational assessment start-ups in Latin America or elsewhere in the world.

Pros:

- This approach would allow interested policymakers to be at the cutting edge of soft skills assessments, testing emerging methods in performance-based assessments for use in global contexts
- Creating partnerships with organizations in the education measurement space could help policymakers gain knowledge about this field

Cons:

- Validating instruments is an iterative process that usually requires testing in different countries/ contexts, which implies significant logistical and financial investments

Examples

RTI 'Home Alone' Game - After conducting pilots of this game-based assessment of problem-solving skills in Morocco and Philippines, researchers say they require "many more pilots" and data from hundreds, if not thousands of students. Researchers remain open to partnering with any organization interested in testing the game with large sample (S. Pouezevara, personal communication, 2021).

Long-Term: Invest in the development of new instruments

In the long term, policymakers can invest in the development of new instruments for measuring 21st century skills among youth. The countries would join a growing list of organizations that are investing in improving global understanding of 21st century skills measurement. Two possible approaches would be: a) partner with existing education assessment developers to create an assessment built specifically for the Latin American region; or b) partner with a university in the region or elsewhere to create this instrument.

Pros:

- Can create an instrument that can later be utilized throughout the LAC region without paying recurring fees to a third party
- Can create an instrument that is more in line with technical realities of developing countries (i.e., limited Internet connection, etc.).
- This approach would address the dearth of instruments for certain hard-to-measure skills that are of value to policymakers.

Cons:

- Developing a new instrument is a resource-intensive exercise that can take several years; in the case of game-based assessments, researchers describe an iterative cycle of design, testing, and revision (Ren, 2019). They also describe an iterative validation process that can consist of several rounds of trials.

6. Considerations for Implementation

Selecting Skills to Prioritize

Developing a strategy for measuring and certifying 21st century skills begins with prioritizing a much shorter list of skills. The following are questions to ask and considerations to keep in mind when selecting a skill to prioritize. For example, McGrath and Adler (2022) reviewed and selected ten life skills based on measurability, malleability, and meaningfulness.

Is it important? The construct that is selected should be: 1) empirically shown to be predictive of important educational and life outcomes, and 2) considered in-demand by employers.

Is it measurable? We know that certain skills are much harder to measure than others, especially complex skillsets that are composed of different constructs. This is partly the reason why there is an uneven availability of instruments across skills, with some skills having several tools available to measure them (e.g., problem-solving, self-control, positive self-concept) and others having few tools (e.g., civic skills, entrepreneurship, flexibility, empathy) (Galloway et al., 2017).

Is it relevant to the context? Practitioners and experts stress that the skills that are chosen for measurement should be an appropriate match to the competency frameworks, learning objectives, and overall 'theory of change' of the education or training system where the instruments will be used. This is especially important when instruments are utilized for program evaluation or accountability purposes.

Is it politically supported? When instruments are utilized within an education system, it is necessary that there is support around the construct that is being measured. One of the ways to achieve this is to gain consensus around the definition of a construct (e.g., global citizenship; collaborative problem-solving) and ensure the testing process is not politicized.

Selecting the Appropriate Instrument

It is important to keep the following characteristics in mind when selecting an instrument for direct use or adaptation:

Designed for the same purpose as intended use: We know that evidence of validity and reliability are characteristics of a measure concerning a particular end use (Cook & Hatala, 2016). A tool that is created and validated for group-level analysis, then, is not appropriate for diagnosing the skills of a particular individual, and an instrument intended for formative assessments should not be used for summative purposes. It is important to understand what the intended use of any given instrument is before utilizing it for high-stakes purposes on an individual.

Appropriate for the youth of the relevant age groups: It is also important to evaluate whether an instrument has been tested and validated for the same age group. It is possible, in an adaptation and contextualization process, to modify an instrument to make it suitable for a different age group than that for which it was created. This can be done by shortening or adding items to the tool, changing the wording of items, and tweaking the level of difficulty of the tasks. However, all adapted instruments require rigorous testing in the target context.

Evidence of validity and reliability: It is critical to verify that there is evidence of validity and reliability of the tool's results in the context where it was tested. There are common standards for what constitutes acceptable psychometric properties. For instance, values of 0.3 for predictive validity and 0.7 for Cronbach's alpha test of reliability are two commonly recognized thresholds (Galloway et al., 2017). It is important to keep in mind that 'validation' of an instrument is not an endpoint, but rather a process by which we collect validity evidence to evaluate the appropriateness of the interpretations and uses of a particular assessment's results (Cook & Hatala, 2016).

Feasible to implement in testing context: It is important to ascertain that it is feasible to successfully implement the instrument in the desired context. From a technical standpoint, certain tests that require a high-speed Internet connection or high-end smart devices may not be suitable for certain school locations or at-home applications among certain demographics. Low digital competencies (limited exposure to technologies) among some populations may also point to a need to find an instrument that can be deployed via other mediums, while low literacy rates may indicate the need to apply an instrument based on observation checklists or where answers are collected by enumerators.

Ease of administration: A final criterion that helps select between instruments is the ease of administration. This includes the degree (if any) of training that is required of administrating personnel —for instance, observers, enumerators, teachers, or other staff. The amount of time required to score results and analyze data should also be taken into account, particularly if selecting an instrument that does not have automated data analysis.

Appendix A: Interviewees & Interview Questions

INTERVIEWEES

We thank interview participants who shared their valuable insights.

Category A: 21st century assessment developers

- **Sarah Pouezevara,** Senior Research Education Analyst, RTI International
- **Jack Buckley,** Head of Assessment & Learning Sciences, Roblox
- **Guy Halfteck,** Founder & President, KnackApp
- **Flavio Cortés,** Director of MIDE UC, Pontificia Universidad Católica de Chile
- **Juan José Goyeneche,** Professor, Universidad de la República de Uruguay
- **Emiliano Pereiro**, Director of Computational Thinking, Plan Ceibal
- **Esther Angeriz**, Professor, Universidad de la República de Uruguay

Category B: 21st century assessment users

- **Akshay Kashyap,** former Deputy Head of Short-Term Training, India National Skills Development Corporation
- **John Gaviria Gómez,** Talent Recruiter & Patricia Vahos, Talent Manager, Comfama, Colombia
- **[Anonymous],** Co-Founder & CEO, Haiti-based NGO
- **Magaly Zúñiga Céspedes,** Director of Evaluation and Research, Fundación Omar Dengo, Costa Rica

INTERVIEW PROTOCOLS

These interview protocols reflect the common sets of questions asked to interviewees in each interviewee category. However, each interview conducted was different, asking questions about the unique experiences of each interviewee.

Category A: 21st century assessment developers

- What is the intended purpose of your organization's assessment instrument, and what uses is it validated for (e.g., individual vs. group-level analysis, summative vs. formative purposes)?

- What was the instrument development process like? What challenges did you encounter in the development process?

- What challenges did you encounter in the implementation process?

- What kind of pre-implementation training is required to use your tool?

- What capacities (e.g., human, technological) does an organization need to have in order to utilize your instrument?

- Can you tell me about the costs associated with the use of your instrument at medium scale?

- If an education or training entity in LAC wanted to use your instrument, what steps would that entity have to go through?

- Would it be OK to attribute your comments to you, and include your organization name, in the final report?

Category B: 21st century assessment users

- For what purpose do you use this 21st century skill assessment instrument?

- What do currently utilize results from this app for?

- What has your experience been like utilizing this instrument with your students/trainees?

- What challenges have you experienced in using this instrument? What lessons have you learned?

- Why did you decide to partner with a private company for skills measurement instead of developing your own instrument?

- What advice would you give a Latin American country (for example, a national skills training institute) seeking to utilize this instrument to measure 21st century skills among its students?

- Would it be OK to attribute your comments to you, and include your organization name, in the final report?

Apendix B: Short Compilation of Instruments

Selection of Assessments Based on Self-Reporting				
Type of Assessment	**Skills targeted**	**Name of Instrument**	**Developing entity/source**	**Description**
Self-report questionnaire	empathy, self-efficacy, self-awareness, persistence, emotional self-regulation, behavioral self-control, gratitude, zest, and optimism	California Healthy Kids Survey, Social and Emotional Health Module	Developed by WestEd for the California Department of Education	This 21-item questionnaire is used by school districts in California to measure soft skills among students aged 10-19 (California School Climate, Health and Learning Surveys, n.d.).
Self-report questionnaire	bonding, behavioral competence, social competence, emotional competence, moral competence, self-determination, self-efficacy, recognizing emotions, beliefs in the future, clear and positive identity, self-determination	Chinese Positive Youth Development Scale (CPYDS)	Daniel Shek, Chinese University of Hong Kong	The CPYD addresses several priority skills for youth development, including positive self-concept, self-control and goal orientation. It has been used as part of monitoring and evaluation for the Hong Kong-based Project P.A.T.H.S (Shek et al., 2007).
Self-report questionnaire	Conscientiousness; emotional stability; extraversion; agreeableness; open-mindedness; self-efficacy.	SENNA 1.0 & SENNA 2.0 (two instruments)	Ayrton Senna Institute	The SENNA instrument was developed to assess socioemotional skills in Brazil in group populations and has been used for 5th, 10th and 12th graders. A second edition, released in 2016, adds items to measure self-efficacy, making the tool more complete (Primi et al., 2016).

Type of Assessment	Skills targeted	Name of Instrument	Developing entity/source	Description
Self-report questionnaire	Adaptability, Connectedness, Conscientiousness, Emotional self-regulation, Empathy, Initiative, Mindfulness, Optimism, Self-efficacy, and Social competence.	Child and Adolescent Wellness Scale (CAWS)	Copeland, E. P., Nelson, R., & Traughber, M. C.	CAWS is a self-report instrument that assesses social and emotional competencies. It has been validated for use among adolescents ages 12-18 and used in the US, Japan, South Korea, Taiwan & Thailand (Copeland et al., 2011).
Self-report questionnaire	openness, conscientiousness, extraversion, agreeableness, and neuroticism (or its opposite, emotional stability)	The Big Five Inventory (BFI)	John, O. P., Donahue, E. M., & Kentle, R. L.	The Big Five Inventory (BFI) is one of the most widely used instruments for identifying personality factors: openness to experience, conscientiousness, agreeableness, and neuroticism (Berkeley Personality Lab, n.d.)
Self-report questionnaire	openness, conscientiousness, extraversion, agreeableness, and neuroticism (or its opposite, emotional stability)	Anchored BFI	Education Development Center (EDC)	The Anchored BFI tool uses the same instrument as the BFI tool, but also includes anchoring vignettes (or scenarios) as well as situational judgment tests (Education Development Center, 2016).
Self-report questionnaire	Positive sense of self; self-control; decision-making skills; moral system of belief; and prosocial connectedness	Jamaica Youth Survey	World Bank	The Jamaica Youth Survey was developed as part of a World Bank-funded project to evaluate programs for promoting well-being among teenagers (12-18) in Jamaica. It addresses key skills such as positive self-concept, higher order thinking and social skills (Gardner et al., 2016).

Type of Assessment	Skills targeted	Name of Instrument	Developing entity/source	Description
Self-report questionnaire	self-control, positive self-concept, higher-order thinking skills, communication	Responses to Stress Questionnaire (RSQ)	Bruce Compass, Vanderbilt Stress and Coping Lab	The RSQ measures coping and involuntary stress responses; participants rate how often they use each coping method or experience each type of involuntary stress response on a 1 to 4 scale. It has been used internationally with respondents over age 9 (Stress and Coping Research Lab, n.d.).
Self-report questionnaire		Laureate Professional Assessment	Laureate Education	The LPA is a free, online assessment exclusively for students in the Laureate network. As a "situational judgment test," it uses real-world workplace scenarios with multiple-choice responses. Assessment-takers answer with what they are most likely and least likely to do (Laureate Education, n.d.).

Source: Adapted from Galloway, T, Lippman, L., Burke, H., Diener, O., and Gates, S. (2017). Measuring Soft Skills & Life Skills in International Youth Development Programs: A Review and Inventory of Tools. Washington, DC: USAID's YouthPower Implementation IDIQ- Task Order 1, YouthPower Action.

Selection of Performance-Based Assessments					
Type	**Skills Targeted** (Author terminology)	**Name of Instrument**	**Developing Entity**	**Description**	**Type**
Cognitive Test	Civic knowledge, attitude, engagement, behavior	International Civic and Citizenship Education Study (ICCS)	International Association for the Evaluation of Educational Achievement (IEA).	ICCS reports on students' knowledge and understanding of concepts and issues related to civics and citizenship, as well as their beliefs, attitudes, and behaviors concerning this domain (IEA, n.d.).	International assessment
Cognitive test + self-reported questionnaire on attitudes questionnaire	"Global competence" (knowledge, skills)	PISA 2018 Global Competence Study (OECD)	Organisation for Economic Co-operation and Development (OECD)	The PISA 2018 Global Competence assessment has two components: 1) a cognitive test focused on "global understanding", a combination of knowledge and skills required to solve problems related to global and intercultural issues; and 2) a set of self-report questionnaire items (OECD, 2018).	International assessment
Performance-based (tasks)	Collaborative Problem-Solving	PISA 2015 Collaborative Problem-Solving Computer-Based Assessment; (OECD)	OECD	CPS skills in PISA 2015 were assessed via individual computer-based CPS tasks, which required the individual to collaborate with computer agents in simulated real-life problem scenarios (OECD, 2017; Stadler et al., 2019).	International assessment

Type	Skills Targeted (Author terminology)	Name of Instrument	Developing Entity	Description	Type
Performance-based (tasks)	Collaborative Problem-Solving	Assessment and Teaching of 21st Century Skills (ATC21S);	University of Melbourne; Cisco; Intel; Microsoft	Pair-based set of tasks done in an asymmetric digital environment where players interact with each other to collaboratively solve a problem (ATC21S, n.d.).	Group-based
Performance-based (tasks)	Complex problem-solving	MicroDYN		A set of tasks done in a digital environment where players manipulate multiple variables to achieve certain targets (Greiff & Funke, 2008).	Individual
Performance-based (game)	Dependability; learning skills; problem solving; self-regulation	"Home Alone" Game (piloted in Morocco & Philippines in 2019)	RTI	The player takes the role of an older sibling who needs to take care of two younger siblings. As the player completes tasks in the games, variables are collected in the background (Pouezevara, 2019).	Individual
Performance-based (game embedded into lessons)	Content-based (science) critical thinking	SimScientists	WestEd	SimScientists is a set of projects by WestEd that create simulation and game-based science assessments embedded into curricular lessons (WestEd, n.d.)..	Individual

Type	Skills Targeted (Author terminology)	Name of Instrument	Developing Entity	Description	Type
Performance-based (game)	Multiple skills (cognitive abilities, emotional intelligence, logical thinking, problem-solving, decision-making)	KnackApp	KnackApp	Knack offers three games: Meta Maze, Balloon Brigade, and Wasabi Waiter. Players are assessed on aspects such as cognitive abilities, emotional intelligence, logical thinking, problem-solving, among others. It has been used in over 120 countries (Knack.it, 2021).	Individual
Performance-based (game)	Communication, Cooperation, Emotion Regulation, Empathy, Impulse Control, and Social Initiation.	Zoo U	Centervention	Zoo U is a game-based socioemotional skills assessment tool. Zoo U presents children with a series of six virtual simulations specifically crafted to elicit the target social skills (Centervention, n.d.; DeRosier & Thomas, 2017).	Individual
Performance-based (game)	Communication, Cooperation, Emotion Regulation, Empathy, Impulse Control, and Social Initiation.	Hall of Heroes	Centervention	Hall of Heroes is an online point-and-click adventure game where players navigate a virtual school-like world, engaging with other characters to solve social problem-solving tasks (DeRosier & Thomas, 2019).	Individual

Selection of Assessments Developed in Latin America and the Caribbean						
Type	**Skills Targeted** (Author terminology)	**Name of Instrument**	**Developing Entity**	**Description**	**Countries**	**Type**
Self-report questionnaire	Emotional regulation (comprised of positive emotional expression; strategies for self-regulation; regulation of negative emotions; & identifying emotions).	"RE-MESACTS" Emotional Regulation Scale	Mesa Regional de Cooperación Técnica sobre Competencias Transversales y Socioemocionales	The RE-MESACTS is a scale that measures emotional regulation in youth. The instrument was developed and validated in 2018 in six countries based on data from two pilots (Comité Internacional para el Desarrollo de los Pueblos, 2020).	Argentina, Chile, Colombia, El Salvador, Paraguay y Uruguay (pilot & validation)	Individual
Performance-based (tasks)	Digital skills	SIMCE TIC	Pontificia Universidad Católica de Chile, on behalf of the Ministry of Education	SIMCE TIC is a computer-based assessment of ICT skills along three dimensions: information, communication & ethics, and social impact. It was applied in 2013 and 2016 (Ministerio de Educación de Chile, n.d.).	Chile	National Assessment
Performance-based and knowledge-based (multiple-choice)	Digital skills	TICómetro	Universidad Nacional Autónoma de México (UNAM)	The TICómetro® is a diagnostic evaluation tool created in 2012 to gauge ICT skills among new university students at UNAM (UNAM, n.d.).	Mexico	Group-based assessment

Type	Skills Targeted (Author terminology)	Name of Instrument	Developing Entity	Description	Countries	Type
Performance-based	Computational thinking	Computational thinking measurement instrument	Universidad de la República (Uruguay); Universidad Nacional Autónoma de México (UNAM); Fundación Omar Dengo (Costa Rica); Paraguay Educa (Paraguay). Socioemocionales (MESACTS)	This is a game-based assessment of computational thinking that is currently being developed by a group of researchers from three Latin American countries. It is built for use among students 10 to 17 years old, and older students (university-aged students) in Mexico (Goyeneche & Pereiro, 2019).	Uruguay; México; Costa Rica; Paraguay	Individual
Self-report questionnaire	Conscientiousness; emotional stability; extraversion; agreeableness; open-mindedness; self-efficacy.	SENNA 1.0 & 2.0	Ayrton Senna Institute	The SENNA instrument was developed to assess socioemotional skills in Brazil in group populations and has been used for 5th, 10th and 12th graders. A second edition, released in 2016, adds items to measure self-efficacy, making the tool more complete (Primi et al., 2016).	Brazil	National Assessment

References

Adhvaryu, A., Kala, N., & Nyshadham, A. (2018). The skills to pay the bills: Returns to on-the-job skills training (NBER Working Paper No. 24313). National Bureau of Economic Research.
https://www.nber.org/system/files/working_papers/w24313/w24313.pdf

Ali, D. A., Bowen, D., & Deininger, K. (2020). Personality traits, technology adoption, and technical efficiency: Evidence from smallholder rice farms in Ghana. *The Journal of Development Studies, 56*(7), 1330-1348.
https://doi.org/10.1080/00220388.2019.1666978

Ashraf, N., Bau, N., Low, C., & McGinn, K. (2020). Negotiating a better future: How interpersonal skills facilitate intergenerational investment. *The Quarterly Journal of Economics 135*(2): 1095-1151.
https://doi.org/10.1093/qje/qjz039

ATS2020. (2020). *ePortfolio: What is an ePortfolio?*
https://resources.ats2020.eu/eportfolio

Avvisati, F., Le Donné, N., & Paccagnella, M. (2019). A meeting report: Cross-cultural comparability of questionnaire measures in large-scale international surveys. *Measurement Instruments for the Social Sciences, 1*(8), 1-10. https://doi.org/10.1186/s42409-019-0010-z

Bos, M. S., Moffa, N. D., Vegas, E., & Zoido, P. Latin America and the Caribbean in PISA 2015: How well can students work together to solve problems? Inter-American Development Bank.
https://publications.iadb.org/publications/english/document/Latin-America-and-the-Caribbean-in-PISA-2015-How-Well-Can-Students-Work-Together-to-Solve-Problems.pdf

Bosch, M., Pages, C., & Ripani, L. (2018). *The future of work in Latin America and the Caribbean: A great opportunity for the region?*
Inter-American Development Bank.
https://publications.iadb.org/publications/english/document/The_Future_of_Work_in_Latin_America_and_the_Caribbean_A_Great_Opportunity_for_the_Region_print_version.pdf

Bruns, B., & Luque, J. (2015). *Great teachers: How to raise student learning in Latin America and the Caribbean.* The World Bank.
https://openknowledge.worldbank.org/handle/10986/20488

Buckley, J., & Kim, Z. F. (2020, October 16). *Preparing for life after COVID with simulation-based assessment.* Inter-American Development Bank (Enfoque Educación blog series).
https://blogs.iadb.org/educacion/en/simulationbasedassessment/

Care, E., Griffin, P., & Wilson, M. (2018). *Assessment and teaching of 21st century skills: Research and applications (Educational Assessment in an Information Age).* Springer International Publishing. https://www.springer.com/gp/book/9783319653662

Cook, D. A., & Hatala, R. (2016). Validation of educational assessments: A primer for simulation and beyond. A*dvances in Simulation 1*(31), 1-12. https://doi.org/10.1186/s41077-016-0033-y

Darling-Hammond, L., & Adamson, f. (2010). Beyond basic skills: The Role of performance assessment in achieving 21st century standards of learning. Stanford University, Stanford Center for Opportunity in Education.
https://edpolicy.stanford.edu/sites/default/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning-executive-summary_0.pdf

Deming, D. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics, 132*(4): 1593-1640. https://doi.org/10.1093/qje/qjx022

DeRosier, M., Craig, A. B., & Sánchez, R. P. (2012). Zoo U: A stealth approach to social skills assessment in schools. *Advances in Human-Computer Interaction,* 654791, 1-7.
https://doi.org/10.1155/2012/654791

Duckworth, A. L., & Yeager, D.S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237-251.
https://doi.org/10.3102/0013189X15584327

Edin, P., Fredriksson, P., Nybom, M., & Öckert, B. (2017). *The rising return to non-cognitive skills (IZA Discussion Paper Series No. 10914).* IZA Institute of Labor Economics.
https://www.iza.org/publications/dp/10914/the-rising-return-to-non-cognitive-skill

Edmonds, E. V., Feigenberg, B., & Leight, J. (2020). *Advancing the agency of adolescent girls (NBER Working Paper No. 27513).* National Bureau of Economic Research. https://www.nber.org/papers/w27513

Education Development Center (2016). *Measuring youth's soft skills across cultures: Evidence from the Philippines and Rwanda.* Education Development Center, Professional Examination Services (ProExam), and Akilah Institute for Women.
https://www.edc.org/sites/default/files/uploads/Anchored-BFI-report.pdf

Galloway, T, Lippman, L., Burke, H., Diener, O., & Gates, S. (2017). *Measuring Soft Skills & Life Skills in International Youth Development Programs: A Review and Inventory of Tools. U.S.* Agency for International Development (USAID).
https://www.youthpower.org/resources/measuring-soft-life-skills-international-youth-development-programs-review-and-inventory-tools

Goldman, S. (2006). Self-discipline predicts academic performance among low-achieving adolescents. *Res: A Journal of Undergraduate Research, 2*(1), 84–97.

Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach (Educational assessment in an information age).* Springer.
https://www.springer.com/gp/book/9789401793940

Grunstein, V., Finoli, M., & Schurmann, S. (2019). *Espacio de resolución colaborativa de problemas en línea descripción de la propuesta y resultados de la implementación en la ciudad.* Ministerio de Educación de la Ciudad de Buenos Aires.
https://www.buenosaires.gob.ar/sites/gcaba/files/2019_espacio_de_resolucion_colaborativa_.pdf

Irava, V., Pathak, A., DeRosier, M., & Chatterjee Singh, N. (2019). Game-based socio-emotional skills assessment: A comparison across three cultures. *Journal of Educational Technology Systems, 48*(1), 51-71. https://doi.org/10.1177/0047239519854042

Kim, H., & Care, E. (2020). *Capturing 21st century skills: Analysis of assessments in selected sub-Saharan African countries.* Brookings Institution and United Nations Educational Scientific and Cultural Organization (UNESCO).
https://www.brookings.edu/wp-content/uploads/2020/09/UNESCO-Kim-Care_Africa-ministudy-2020.pdf

Kim, Y. J. & Ifenthaler, D. (2019). Game-based assessment: The past 10 years and moving forward. In Kim, Y. J, & Ifenthaler, D. (Eds.), *Game-ased assessment revisited* (pp. 3-13). Springer.
https://doi.org/10.1007/978-3-030-15569-8

Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education, 87,* 340–356.
https://doi.org/10.1016/j.compedu.2015.07.009

King, G. (n.d.). *Anchoring Vignettes Overview.* Gary King.
https://gking.harvard.edu/vign

Kreitchmann, R. S., Abad, F. J., Ponsoda, V., & Nieto, M. D., & Morillo, D. (2019). Controlling for response biases in self-report scales: forced-choice vs. psychometric modeling of Likert items. *Frontiers in Psychology, 15*(10), 1-12.
https://doi.org/10.3389/fpsyg.2019.02309

Laajaj, R., & Macours, K. (2017). *Measuring skills in developing countries (Policy Research Working Paper Series No. 8000).* World Bank Group. https://openknowledge.worldbank.org/handle/10986/26250

Major, S. O., Seabra Santos, M. J., & Martin, R. P. (2015). Are we talking about the same child? Parent–teacher ratings of preschoolers' social–emotional behaviors. *Psychology in the Schools,* 52(8), 789–799. https://doi.org/10.1002/pits.21855

Mayer, I. (2018). Assessment of teams in a digital game environment. *Simulation & Gaming, 49*(6), 602-619. https://doi.org/10.1177/1046878118770831

Mateo Díaz, M. M., Buenadicha Sánchez, C., Bustelo, Mç, Duryea, S., Heredero, E., Rubio-Codina, M., Rucci, G., Becerra, L. (2019). *21st century skills: Transversal skills development in Latin America and the Caribbean.* Inter-American Development Bank. https://publications.iadb.org/publications/english/document/21st_Century_Skills_en_en.pdf

Mateo Diaz, M., Lim, J. R., Cardenas-Navia, I., & Elzey, K. (2022a). A world of transformation: Moving from degrees to skills-based alternative credentials. Inter-American Development Bank. https://doi.org/10.18235/0004299

Mateo Diaz, M., Lim, J. R., Pellicer, C., López, E., Rodríguez, H., López, R., Magro Mazo, C., Vásquez Guerra, A., Quesada Alvarado, A., Brooks-Young, S., Álvarez, X., Ramos, Y., Rivas, A., Barrenechea, I., Brazão, V., Ndebele, V., Nathan, D., Groot, B. (2022b). The power of curriculum to transform education: How education systems incorporate 21st century skills to prepare students for today's challenges (M. Mateo-Diaz., & J. R. Lim, Eds.). Inter-American Development Bank. https://doi.org/10.18235/0004360

Mateo Diaz, M., Rucci, G., Amaral, N., Arias Ortiz, E., Becerra, L., Bustelo, M., Cabrol, M., Castro, J., Caycedo, J., Duryea, S., Groot, B., Heredero, E., Hincapie, D., Magendzo, A., Navarro, J. C., Novella, R., Rieble-Aubourg, S., Rubio-Codina, M., Scartascini, C., & Vezza, E. (2019). Future is now: Transversal skills in Latin America and the Caribbean in the 21st century. Inter-American Development Bank. https://doi.org/10.18235/0001950

McGrath, R. E., & Adler, A. (2022). Skills for life: A review of life skills and their measurability, malleability, and meaningfulness. Inter-American Development Bank. https://doi.org/10.18235/0004414

Montalvao, J., Frese, M., Goldstein, M., & Kilic, T. (2017). *Soft skills for hard constraints: Evidence from high-achieving female farmers. Policy Research Working Paper No. 8095.* World Bank. https://openknowledge.worldbank.org/handle/10986/27295?locale-attribute=fr

Oranje, A., Mislevy, B., Bauer, M. I., & Jackson, G. T. (2019). Summative game-based assessment. In D. Ifenthaler & Y, J. Kim. (Eds.), *Game-Based Assessment Revisited* (pp. 67-83). Springer. https://www.springer.com/gp/book/9783030155681

Portela-Pino, I., Alvariñas-Villaverde, M., & Pino-Juste, M. (2021). Socio-emotional skills as predictors of performance of students: Differences by gender. *Sustainability, 13*(9), 4807. https://doi.org/10.3390/su13094807

Pouezevara, S., Powers, S., Moore, G., Strigel, C., McKnight, K. (2019). *Assessing Soft Skills in Youth Through Digital Games [Conference paper].* International Conference of Education, Research and Innovation. Seville, Spain. https://doi.org/10.21125/iceri.2019.0774

Pouezevara, S. (November 2019). Assessing Soft Skills in Youth Through Digital Games. International Conference of Education, Research and Innovation. Seville, Spain. https://ierc-publicfiles.s3.amazonaws.com/public/resources/Pouezevara%20et%20al%20Game%20Based%20Assessment%20ICERI19.pdf

Prada, M. F., Rucci, G., & Urzúa, S. (2019). *Training, Soft Skills and Productivity: Evidence from a Field Experiment in Retail (Working Paper Series No. IDB-WP-1015).* Inter-American Development Bank. https://publications.iadb.org/publications/english/document/Training_Soft_Skills_and_Productivity_Evidence_from_a_Field_Experiment_in_Retail_en_en.pdf

Ren, X. (2019). Stealth assessment embedded in game-based learning to measure soft skills: A critical review. In D. Ifenthaler & Y, J. Kim. (Eds.), *Game-Based Assessment Revisited* (pp. 67-83). Springer. https://link.springer.com/chapter/10.1007/978-3-030-15569-8_4

Sandberg, H. (2016). *The good and the bad of game-based assessment.* Focus on ETS R&D, 1. https://www.ets.org/research/policy_research_reports/focus_on_rd/issue1

Savelyev, P., & Tan, K. T. K. (2019). Socioemotional skills, education and health related outcomes of high-ability individuals. *American Journal of Health Economics, 5*(2), 250-280. https://econpapers.repec.org/article/tpramjhec/v_3a5_3ay_3a2019_3ai_3a2_3ap_3a250-280.htm

Shute, V., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior, 63*(2016), 106-117. https://doi.org/10.1016/j.chb.2016.05.047

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games (The John D. and Catherine T. Macarthur Foundation reports on digital media and learning).* The MIT Press. https://mitpress.mit.edu/books/stealth-assessment

Sorrenti, G., Zölitz, U., Ribeaud, D., & Eisner, M. (2020). *The causal impact of socio-emotional skills training on educational success (IZA Institute of Labor Economics Discussion Paper No. 13087).* https://ssrn.com/abstract=3562877

United Nations. (2015). *The economics of climate change in Latin America and the Caribbean: The Paradoxes and challenges of sustainable development.* United Nations Economic Commission for Latin America and the Caribbean. https://repositorio.cepal.org/handle/11362/37311

USAID (2017). M*easuring youth competencies across contexts: lessons from implementers on how to adapt soft skills measurement tools.* U.S. Agency for International Development (USAID). https://www.youthpower.org/resources/technical-brief-measuring-youth-competencies-across-contexts-lessons-implementers-how-adapt-soft-skills-measurement-tools

Wang, Y., Yang, Z., Zhang, Y., Wang, F., Liu, T., & Xin, T. (2019). Socio-emotional competency positively predicts academic achievement, academic emotions and attitudes. *Frontiers in Psychology, 10,* 1-10. https://doi.org/10.3389/fpsyg.2019.01282

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: an integrative model. *Psychological Methods, 18*(3), 320–334. https://doi.org/10.1037/a0032121

Weiss, S., & Roberts, R. D. (2018). Using Anchoring Vignettes to Adjust Self-Reported Personality: A Comparison Between Countries. F*rontiers in Psychology, 9,* 325, 1-17. https://doi.org/10.3389/fpsyg.2018.00325

Weller, J., Gontero, S., Campbell, S. (2019). *Cambio tecnológico y empleo: una perspectiva latinoamericana: Riesgos de la sustitución tecnológica del trabajo humano y desafíos de la generación de nuevos puestos de trabajo (Macroeconomía del desarrollo Series No. 201).* Comisión Económica para América Latina y el Caribe (CEPAL). https://www.cepal.org/es/publicaciones/44637-cambio-tecnologico-empleo-perspectiva-latinoamericana-riesgos-la-sustitucion

World Learning (n.d.). *WorkLinks Skills and Values Assessment (WLSVA).* World Learning. https://www.worldlearning.org/what-we-do/wlsva-toolkit/#toolkit-form

Wolvin, A., & Lim, J. R. (2022). Skills for life: Listening. Inter-American Development Bank. https://doi.org/10.18235/0004351

Annex References

Assessment and Teaching of 21st Century Skills (n.d.). About the project. ACT21S. http://www.atc21s.org

Berkeley Personality Lab (n.d.). The Big Five Inventory: Frequently asked questions. Berkeley Personality Lab. https://www.ocf.berkeley.edu/~johnlab/bfi.htm

Centervention (n.d.). Zoo U: An Evidence-based SEL Game for Students in Grades 2-4. Centervention. https://www.centervention.com/zoo-u-sel-game/

California School Climate, Health and Learning Surveys (n.d.). Using CalSCHLS to assess social-emotional learning and health. https://data.calschls.org/resources/CalSCHLS_AssessSELH.pdf

Comité Internacional para el Desarrollo de los Pueblos (2020). Escala de regulación emocional RE-MESACTS: resultados del proceso de validación psicométrica en seis países de América Latina. Mesa Regional de Cooperación Técnica sobre Competencias Técnicas y Socioemocionales. https://www.mesacts.com/wp-content/uploads/2020/03/Escala-de-Regulación-Emocional-RE-MESACTS_publicación-oficial-febrero-2020.pdf

Copeland, E. P, Nelson, R. B. & Traughber, M. C. (2010). Wellness dimensions relate to happiness in children and adolescents. Advances in School Mental Health Promotion, 3(4), 25–37. https://www.tandfonline.com/doi/abs/10.1080/1754730X.2010.9715689

DeRosier, M. E., & Thomas, J. M. (2018). Establishing the criterion validity of Zoo U's game-based social emotional skills assessment for school-based outcomes. Journal of Applied Developmental Psychology, 55, 52–61. https://www.sciencedirect.com/science/article/abs/pii/S0193397316301514

DeRosier, M. E, & Thomas, J. M. (2019). Hall of heroes: A digital game for social skills training with young adolescents. International Journal of Computer Games Technology, 2019, 1–12. https://www.hindawi.com/journals/ijcgt/2019/6981698/

Education Development Center (2016). Measuring youth's soft skills across cultures: evidence from the Philippines and Rwanda. Education Development Center, Professional Examination Services (ProExam), and Akilah Institute for Women. https://www.edc.org/sites/default/files/uploads/Anchored-BFI-report.pdf

Gardner, J. M., Williams, K. R., Guerra, N. G, & Walker, I. (2011). The Jamaica Youth Survey: Assessing core competencies and risk for aggression among Jamaican youth. Caribbean Quarterly, 57(1), 35–53. https://cpb-us-w2.wpmucdn.com/sites.udel.edu/dist/0/659/files/2013/05/Meeks-Garnder-et-al.-2011-Jamaica-Youth-Survey.pdf

Goyeneche, J. J., & Pereiro, E. (2019). ¿Cómo evaluamos una habilidad cognitiva en construcción? Pensamiento computacional en la región. Fundación Ceibal Revista +Aprendizajes 2(3), 24-27. https://issuu.com/bibliotecadigital-esramada/docs/revista__aprendizajes_3

Greiff, S., & Funke, J. (2008). Measuring complex problem-solving: The MicroDYN approach. University of Heidelberg. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.145.2135&rep=rep1&type=pdf

International Association for the Evaluation of Educational Achievement (n.d.). "International Civic and Citizenship Education Study." International Association for the Evaluation of Educational Achievement. https://www.iea.nl/studies/iea/iccs

Knack.it. (2021). Knack. https://knackapp.com

Laureate Education (n.d.). How the LPA works. Laureate Professional Assessment. http://lpa.laureate.net/the-assessment/how-the-lpa-works.html

Ministerio de Educación de Chile (n.d.). Simce TIC: Qué evalúa. http://www.enlaces.cl/sobre-enlaces/quienes-somos/

Organisation for Economic Co-operation and Development (2018). Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework. https://www.oecd.org/pisa/Handbook-PISA-2018-Global-Competence.pdf

Organisation for Economic Co-operation and Development (2017). PISA collaborative problem-solving framework. https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Pouezevara, S. (November 2019). Assessing Soft Skills in Youth Through Digital Games. International Conference of Education, Research and Innovation. Seville, Spain. https://ierc-publicfiles.s3.amazonaws.com/public/resources/Pouezevara%20et%20al%20Game%20Based%20Assessment%20ICERI19.pdf

Primi, R., Santos, D., John, O. P, & De Fruyt, F. (2016). Development of an inventory assessing social and emotional skills in Brazilian youth. European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 32(1), 5-16. https://core.ac.uk/download/55717672.pdf

Shek, D. T. L., Siew, A. M. H, & Lee, T. Y. (2007). The Chinese Positive Youth Development Scale: A validation study. Research on Social Work Practice, 17(3), 380–391. https://journals.sagepub.com/doi/abs/10.1177/1049731506296196?journalCode=rswa

Stadler, M., Herborn, K., Mustafi , K., & Greiff, S. (2019). Journal of Intelligence, 7(3), 1-15. https://www.mdpi.com/2079-3200/7/3/15/htm

Stress and Coping Research Lab (n.d.). Responses to Stress Questionnaire (RSQ). Vanderbilt University. https://my.vanderbilt.edu/stressandcoping/rsq/

Universidad Nacional Autónoma de México. (n.d.). Informes TICómetro. UNAM. https://educatic.unam.mx/publicaciones/informes-ticometro.html

WestEd (n.d.). What is SimScientists? WestEd. http://simscientists.org/home/index.php

SKILLS FOR LIFE

21st Century Skills is an initiative led by the Inter-American Development Bank (IDB) that brings together public and private sector stakeholders. The initiative strengthens learning ecosystems to equip Latin American and Caribbean citizens with transversal skills.

https://clic-skills.iadb.org/en/
skills21@iadb.org

Inter-American Development Bank