RAND CORPORATION

Research Report

JONATHAN SCHWEIG, GARRETT BAKER, LAURA S. HAMILTON, BRIAN M. STECHER

Building a Repository of Assessments of Interpersonal, Intrapersonal, and Higher-Order Cognitive Competencies

There is growing consensus among policymakers, educators, business leaders, and the general public that the skills and education that are needed to successfully participate in the workplace are changing as technology advances and global economies become increasingly interconnected (National Center on Education and the Economy, 2007; Saavedra and Opfer, 2012). It is not sufficient for job seekers to have only basic numeracy and literacy skills (Levy and Murnane, 2005). Instead, employers seek applicants who possess a variety of competencies, including communication and collaboration competencies, critical thinking, creativity, interpersonal competencies, and self-management (National Center on Education and the Economy, 2007). Furthermore, employers anticipate that the importance of these competencies will continue to increase for new workforce entrants in the future (Casner-Lotto and Barrington, 2006). Many of these same competencies are perceived to be important for a wide range of life outcomes beyond the workplace, including individual goal fulfillment (Nagaoka et al., 2015), the promotion of healthy behaviors, and a decrease in criminal activity (Heckman, Pinto, and Savelyev, 2013; Almlund et al., 2011). These competencies are variously referred to in the literature as 21st-century competencies, social-emotional competencies, or deeper learning competencies (e.g., Partnership for 21st Century Skills, 2008; National Research Council, 2012; William and Flora Hewlett Foundation, undated).

There is also a growing consensus that teachers and schools should attend to the development of skills and competencies that are expected to better prepare students for work and careers. Newly adopted content standards, including the Next Generation Science Standards and the Common Core State Standards for Math and English Language Arts, as well as several state-adopted

standards emphasize critical thinking, problem-solving, and communication (Kyllonen, 2012), and assessments aligned to these standards have been developed and are currently used in several states (Kyllonen, 2012). Additionally, the Every Student Succeeds Act (ESSA) of 2015 features language that expressly broadens the definition of *student success* beyond traditional achievement measures (Melnick, Cook-Harvey, and Darling-Hammond, 2017; Ferguson, 2016; Gayl, 2017) to include student engagement, educator engagement, student access to and completion of advanced coursework, post-secondary readiness, and school climate and safety. This expanded definition creates opportunities to incorporate instruction on complex cognitive, interpersonal, and intrapersonal competencies into school curricula. In fact, ESSA provides several funding streams that can be used to support instruction that focuses on students' interpersonal and intrapersonal competencies (Grant et al., 2017). Several states, including Massachusetts, have developed plans for implementing social and emotional competency-based curricula across the state (Massachusetts Department of Elementary and Secondary Education, undated).

Even as enthusiasm for incorporating these competencies into schools and classrooms has increased, there are still several issues that adversely affect widespread adoption in state or district policy. First, although there might be a consensus about the importance of the competencies themselves, there is much less agreement about how best to measure these competencies. Many of these competencies— including collaboration, self-control, and self-regulated learning—are not amenable to familiar, efficient testing approaches, such as multiple-choice testing, that are traditionally used for measuring academic achievement (Kyllonen, 2012). At the same time, assessments of these competencies have proliferated—there are hundreds of protocols, surveys, questionnaires, and other instruments for measuring these skills and competencies (Coryn et al., 2009; Soland, Hamilton, and Stecher, 2014). However, these existing assessments are often developed in a fragmented style, with specific organizations or consortia creating assessments of the specific domains salient to their work, with little (if any) coordination across these diverse efforts (Yuan, Stecher, and Hamilton, 2015). Many of these assessments exist in various stages of development, with limited information about their technical quality or their suitability for use in educational contexts. The information that does exist often is not easily accessible to practitioners, which makes it difficult, perhaps impossible, for interested individuals to appraise the validity, reliability, and comparability of indicators for inferences about student or school improvement (Hough, Kalogrides, and Loeb, 2016; Stecher and Hamilton, 2014). As a result, there is a systematic lack of access to simple, efficient, high-quality assessments, and that often hinders efforts to promote these competencies through either classroom-based assessment or large-scale indicator or accountability systems.

If educators increase their focus on these skills through the adoption of standards, curricula, or instructional practices, they will need a way to assess the impact of this instruction and to identify students' strengths and weaknesses in these competencies. In addition, district and state education leaders who want to promote these competencies could benefit from a way to monitor students' progress. Assessments of these competencies are not only likely to be useful for classroom teachers,[1] they also could be used in school improvement plans or in accountability systems. In fact, some districts, including the CORE districts in California (Core Districts, undated), are using measures of interpersonal and intrapersonal competencies in school reporting systems (West, 2016). However, no state has yet included such measures in statewide school accountability systems in their ESSA plans (Blad, 2017), and there are important questions about whether appropriate measures of these competencies exist for use in high-stakes school accountability (Blad, 2017; Duckworth and Yeager, 2015).

One way to address some of the conceptual and practical challenges related to assessment is to establish a database of assessments that features both descriptive information about the assessments and available assessment materials. Although compendia of assessments have been developed (as we will discuss later), no system yet has allowed education practitioners and researchers to identify such assessments

that are appropriate for use with students in kindergarten through 12th grade (K–12) or to search for measures that have the features that they need for their particular contexts and purposes (Yuan, Stecher, and Hamilton, 2015). The RAND Assessment repository was designed to address this gap.

The repository was designed to benefit three distinct user groups: practitioners, researchers, and policymakers.[2] It enables practitioners to explore what assessments are available and to obtain key information about what they are designed to measure, how they operate, what demands they place on students and teachers, and what kinds of uses their scores support. For researchers, the repository is a place to identify assessments—both operational and under development—related to a given construct, to quickly scan the evidence about their reliability and validity, and to contact the developers for more information. For policymakers, the repository offers a way to see whether assessments exist to support various policy options. As we explored approaches to designing and populating the repository, we focused on approaches that could, to the extent possible, address the needs of each of these groups simultaneously without requiring the material to be customized to each group. In this report, we briefly summarize the approach that we carried out to develop the repository, including the processes we used to search for, code, and summarize information about assessments.[3] We also describe some challenges that will need to be addressed to promote effective use of these assessments and improve policy and practice.

Repository Development

Our approach to developing the repository is similar to that of a "systematic review," a methodology developed to provide researchers and policymakers with an explicit, rigorous approach to synthesize and collate research results across a large number of published studies (Gough, Oliver, and Thomas, 2017). Although systematic reviews generally focus on the critical synthesis of research findings rather than the collection of assessments and instruments, our approach uses the same basic logic and stages (Figure 1). In the remainder of this section, we briefly describe each of these stages in more detail.

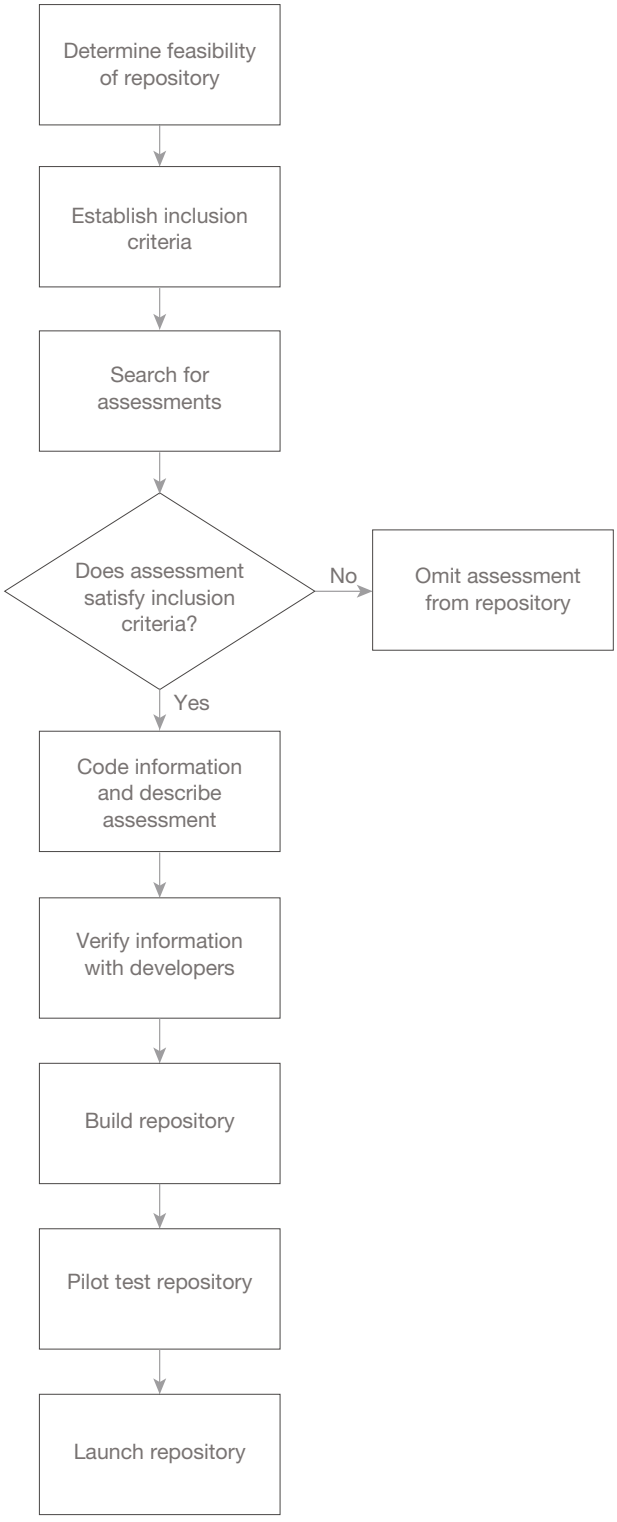
Determine Feasibility of Repository

The first step in building the RAND repository was to conduct a feasibility study (Yuan, Stecher, and Hamilton, 2015). The objectives of this study were to determine the processes that would be needed to collect, review, and code assessments for an online, interactive database; to identify a preliminary set of user-interface functions that would be helpful for practitioners, researchers, and policymakers; and to identify potential challenges that would arise in repository development.

To accomplish these objectives, researchers first conducted a search for relevant assessments. Assessments were identified from an existing list collected as a part of a prior project on 21st-century competencies (Soland, Hamilton, and Stecher, 2013). Additionally, the team searched databases of educational and psychological assessments, such as the American Psychological Association (APA) PsycTESTS® database, and reviewed projects funded by the Institute of Education Sciences and the National Science Foundation that might be developing assessments of hard-to-measure competencies. Researchers also relied on other repository websites and databases of research articles and reports (for full details, see Yuan, Stecher, and Hamilton, 2015). Based on this phase of the feasibility study, the research team concluded that

The repository was designed to benefit three distinct user groups: practitioners, researchers, and policymakers.

Figure 1. Flowchart Illustrating Repository Development Approach



there were enough assessments available that could be sufficiently described and cataloged to make a repository a useful resource.

The feasibility study also explored some practical issues around developing an online repository. Yuan, Stecher, and Hamilton (2015) reviewed 12 existing repository websites and compared and contrasted their features. Additionally, the research team conducted interviews with potential users (including teachers, school administrators, researchers, and leaders of nonprofit education organizations) to collect feedback about users' needs for a repository and how they might use it if it existed. This work (the website review and interviews) resulted in creation of a set of key features that would be included in the repository design, including a flexible search function, the ability to compare and contrast assessments, a consistent design for presenting information, and provision of links to related research (Yuan, Stecher, and Hamilton, 2015). As a part of this phase of the research, a small-scale pilot version of the repository was developed (Yuan, Stecher, and Hamilton, 2015). This pilot served as the prototype for the fully developed version.

Establish Inclusion Criteria

After determining that an online, interactive repository was feasible, we developed a conceptual framework that guided decisions about the kinds of assessments to include in the repository. The criteria for determining inclusion are summarized in Table 1, which describes information about the competencies that are measured, the purposes of the instrument, the unit of analysis, age range, and geography.

As an organizing tool, we used the competency classifications developed by the National Research Council (2012), which characterizes the relevant competencies as cognitive (e.g., critical thinking, problem-solving, creativity, information literacy), interpersonal (e.g., communication, collaboration, empathy, trust, conflict resolution, leadership), or intrapersonal (flexibility, adaptability, curiosity, perseverance, grit, integrity, self-regulation). Future upgrades of the repository could expand its scope to include some of the excluded categories, such as assessments that were developed internationally or that focus on school or classroom climate.

Search for Assessments

We searched for assessments using a three-part process. We began by searching compilations of instruments that had been created by other organizations. These included 12 separate compendia (see Table 2), such as comprehensive, large-scale reports produced by Mathematica (Atkins-Burnett et al., 2012), REL Southeast (Fredricks et al., 2011), and Child Trends (2014).

Our second step was to search for assessments on major online databases, including PsycNet and the Buros Center for Testing, as well as websites maintained by the Stanford Center for Assessment, Learning and Equity (SCALE) and the Center on Great Teachers and Leaders at the American Institutes for Research (AIR), which maintains a comprehensive guide to evaluation products for use in educational settings.

The third and final step in the search process was to conduct outreach with experts in the field. We located email contact information for as many assessment developers as possible ($N = 173$), searching developer websites and pulling author contact information from peer-reviewed journal articles. In an email to each developer, we provided a description of the project objectives and notification of our intent to include the assessment in the online repository. Each developer was invited to review the assessment summary and provide feedback and updated information. Additionally, developers were asked to provide updated web links for accessing measures. We received feedback on 33 assessment summaries. Ten developers replied to confirm that the information we had used was accurate and up to date. An additional seven developers requested that their assessments be excluded from the repository. In some cases, developers were able to provide related assessments that we had not located in our search process.This process was particularly helpful in identifying new assessments that might have been missed by compendia created even just a few years ago, or lesser-known assessments that did not appear in the compendia we reviewed.

Table 1. Inclusion Criteria for the Repository

Criterion	Inclusion Criteria	Examples of Excluded Assessments
Competencies	Assessments must relate directly to intrapersonal, interpersonal, or higher-order cognitive competencies.	Assessments that are academically focused (e.g., end-of-course assessments, Scholastic Aptitude Tests, Advanced Placement exams) and do not produce indicators about a student's intrapersonal, interpersonal, or higher-order cognitive competencies that are separate from indicators of academic achievement in a subject area, such as mathematics, are excluded.
Purpose	Assessments must not be intended for use exclusively in clinical settings and should not treat problem behaviors as providing evidence regarding competencies. In addition, they must not exist only as curriculum-embedded assessments. Assessments must be appropriate for use in school settings with the general student population.	We exclude those assessments that focus exclusively on problem behaviors, are intended to diagnose social or psychological disorders, or are intended to be used to refer students to special services (including special education programs). We also exclude assessments that were designed to be used exclusively in the context of a particular curriculum, such as scoring rubrics designed to appraise a student's work on a six-week literacy or science project.
Unit of analysis	Assessments must focus on measuring an individual student's competencies (even if score reports contain classroom or school aggregates), and not school or classroom features.	Assessments of aggregate phenomena, such as school and classroom climate are excluded, as are surveys that ask respondents to evaluate programs, interventions, or program providers.
Development process	Assessments must be developed by assessment experts, consortia, or researchers.	Assessments developed by individual teachers that were intended for use only in a specific class or school context are excluded. We also exclude local modifications of existing instruments—for example, in cases where specific schools or school districts made small modifications to existing instruments and repackaged them.[a]
Age range	Assessments must be applicable to school-age students (ages 5–18), meaning assessments that are designed for age ranges that include K–12 (e.g., measures for prekindergarten through age 5 or ages 16–adult).	Assessments that are exclusively for use with preschool-age children, postsecondary-age students, or adults are excluded.
Geography	Assessments must be suitable for use with students in the United States. All included assessments were developed in the United States and/or validated in international settings where the dominant language is English, including students from the United States, Canada, Australia, or the United Kingdom.	Assessments that were developed internationally and validated exclusively on populations from countries where the dominant language is NOT English, and for which there was no evidence of suitability for use with school-aged populations in the United States, are excluded.

[a] We included CORE district assessments because they are widely used.

Code Information and Describe Assessment

We were able to collect information on 271 assessments. After sorting the assessments identified in the search into those that met all inclusion criteria those that did not, we described the 213 selected assessments in a database and coded information about each assessment.[4] We coded information on 24 dimensions (Table 3), including the name of the assessment, its purpose, information about how the assessment can be obtained, and available reliability and validity evidence. In some cases, this information was included in the sources that were used in the assessment search. However, when the information was not included in that initial search, additional information was obtained through a web search (using Google Scholar, PsycNet, ERIC, and other databases). Finally, our collaboration with the team that worked on a similar guide under the auspices of the Assessment Work Group (AWG) convened by the Collaborative for Academic, Social, and Emotional Learning (CASEL) identified additional information that we reviewed.[5]

It should be noted that some (approximately 40) of the assessments meeting our inclusion criteria are

Table 2. Sources Searched

Category	Source	Title
Previously published compilations of instruments	Atkins-Burnett et al., 2012	*Landscape Analysis of Non-Cognitive Measures*
	Fredricks et al., 2011	*Measuring Student Engagement in Upper Elementary Through High School: A Description Of 21 Instruments*
	National Collaboration for Youth, 2012	*A Shared Vision for Youth: Common Outcomes and Indicators*
	Reeves, Venator, and Howard, 2014	*The Character Factor: Measures and Impact of Drive and Prudence*
	Wilson-Ahlstrom et al., 2014	*From Soft Skills to Hard Data: Measuring Youth Program Outcomes*
	Child Trends, 2014	*Measuring Elementary School Students' Social and Emotional Skills*
	Haggerty, Elgin, and Woolley, 2011	*Social-Emotional Learning Assessment Measures for Middle School Youth*
	SEL Solutions at American Institutes for Research, 2010	*Are You Ready to Assess Social and Emotional Development?*
	Building Equitable Learning Environment Network[a]	Survey Measurement Menu
	Galloway et al., 2017	*Measuring Soft Skills and Life Skills in International Youth Development Programs*
	Noltemeyer et al., 2015	*Mental Health, Social-Emotional and Behavioral Screening and Evaluation Compendium*
	Soland, Hamilton, and Stecher, 2013	*Measuring 21st Century Competencies: Guidance for Educators*
	Elliott et al., 2018	"Development and Initial Validation of a Social Emotional Learning Assessment for Universal Screening"
	Stillman et al., 2018	*Measuring Noncognitive Variables: Improving Admissions, Success and Retention for Underrepresented Students*
	Sedlacek, 2017	"Strengthening Social Emotional Learning with Student, Teacher, and Schoolwide Assessments"
	Grant et al., 2017	*Social and Emotional Learning Interventions Under the Every Student Succeeds Act: Evidence Review*
World Wide Web	SCALE Performance Assessment Project, undated	
	Center on Great Teachers and Leaders at AIR, 2013	
	Buros Center for Testing, undated	
	American Psychological Association PsycNET, undated[b]	
	Education Endowment Foundation, 2017	
Outreach	Correspondence and referrals from experts in the field	
	List of measures compiled by CASEL's AWG[a]	

[a] This list was provided to the authors and is not available to the public.

[b] PsycNET search terms: interpersonal competence OR intrapersonal OR socioemotional OR critical thinking. We excluded "Adulthood (18 years and up)" from the search, as well as PsycTESTS Classification 7300 "Physical Health/Illness Related Assessment."

Table 3. Information Coded for All Assessments Meeting Inclusion Criteria

Dimension	Details
Name	• Assessment name
Purpose	• Summary of the purpose of the instrument, as described by the developer or vendor, including specific constructs measured and intended uses if provided
Main constructs measured	• Whether the instrument measures cognitive, intrapersonal, or interpersonal competencies. Main constructs are also coded using the coding system developed by the Taxonomy Project at Harvard's Ecological Approaches to Social Emotional Learning (EASEL) Lab (Jones, undated)
Applicable grade levels	• Ages and/or education levels of individuals whom the measures intended to assess
Publication year for most recent version	• Year in which most recent version of assessment was published
Year originally developed	• Year instrument was first developed
Notes	• Any miscellaneous but relevant items of information related to the assessment and/or its development (e.g., whether the assessment was designed as a part of a specific intervention or is a modification or update of another assessment)
Related measures	• Other assessments that are related (e.g., short forms, parent, student, and teacher forms, etc.)
Respondent	• Person who is primarily responsible for completing the instrument (i.e., the student/child or another reporter, such as a teacher or parent)
Method of administration	• How assessment is administered (pencil/paper, digital/online, interview, etc.)
Number of items and item format	• Whether items are selected-response, performance-assessment tasks, extended-response, etc.
Administration time	• Approximate administration time in a typical setting (estimated by developers)
Available languages	• Languages (other than English) in which assessment is available, according to the developers
Fee for use	• Whether users must pay a fee to access and use assessment. We used three categories: 1. "No" indicates no fee, the assessment is free and publicly available 2. "Yes" indicates a fee charged by developer to access the instrument 3. "N/A" indicates no applicable fee but that there might not be a way to access the instrument without contacting developers directly, or that access is limited in some way (e.g., the instrument is published in a journal that requires a subscription for access).
Credentials required for administration	• Whether users are required to have specific credentials to purchase or administer assessments (e.g., a requirement for "Qualification Level B" to purchase or administer assessments). Qualification Level B refers to an individual with at least one of the following: • A master's degree in psychology, education, occupational therapy, social work, counseling, or a field closely related to the intended use of the assessment, and formal training in the ethical administration, scoring, and interpretation of clinical assessments • Certification by or full active membership in a professional organization that requires training and experience in the relevant area of assessment • A degree or license to practice in the health care or allied health care field • Formal, supervised mental health, speech/language, occupational therapy, social work, counseling, and/or educational training specific to assessing children, or in infant and child development, and formal training in the ethical administration, scoring, and interpretation of clinical assessments
Overall score reporting	• Whether an overall (composite) score is reported in the score report
Subscore reporting	• Names of subscores (if any) that are reported
Scoring procedures	• How scores are obtained (e.g., hand-scored, computer-scored, scored online)
Interpretive information	• Whether instrument developer provides any information to assist in score interpretation
Populations for which technical quality evidence has been collected	• Demographic information about the sample used to collect evidence about score reliability and evidence about validity of inferences based on scores
Reliability evidence	• Any information about the precision or generalizability of scores (e.g., internal consistency or test-retest reliability)
Validity evidence	• Any information about sources of validity evidence, including evidence based on content, evidence based on response processes, evidence based on internal structure, and evidence based on relationships with other variables
Obtaining a copy of the measure	• Stable web link for obtaining a copy of the assessment (this is sometimes a digital object identifier link directly to an article if there is no affiliated website)
References	• Related article(s) containing validity/reliability information for the assessment

proprietary. In many of these cases, we did not have permission to include detailed information about the assessment in the repository. Nevertheless, we included publicly available information about these assessments so that users are aware of them and can contact the developers if they would like additional information or are interested in acquiring the instrument.[6]

Two specific codes in Table 3 merit further discussion. First, we used the National Research Council competency classifications (e.g., cognitive, interpersonal, and intrapersonal) to describe the main constructs measured. Although it is possible to search the repository using these three classifications, we acknowledge that these broad descriptors might not always be helpful for practitioners who are searching for assessments that were designed to appraise or measure specific competencies. To address this limitation, all assessments were also coded by a team of researchers from the EASEL Lab, led by Stephanie Jones of the Harvard Graduate School of Education, using codes for social-emotional learning domains that classify domains into approximately 20 categories.[7] This finer-grained classification will serve as a crosswalk for finding similar assessments in the repository. Users can also enter a competency into the search box to identify all assessments that use that term in their descriptions.

Second, we relied on the descriptions of reliability and the sources of validity evidence outlined in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, and Joint Committee on Standards for Educational and Psychological Testing, 2014) to frame the reliability and validity evidence that we included, coded, and described in our database. We classified validity evidence into four categories: *evidence based on test content* (evidence about the relationship between the content of a test and the construct it is intended to measure), *evidence based on response processes* (the relationship between the construct and the nature of the response or performance engaged in by the test taker); *evidence based on internal structure* (the degree to which relationships among items and test components are consistent with the proposed constructs), and *evidence based on relations with other variables* (whether test scores relate to external variables in predicted way). Where possible, we relied on publicly available full-text psychometric studies that were authored by the instrument developers as the primary source of information on score reliability and validity. We located these studies using Google Scholar and the PsychTEST database by searching for the full name of the instrument, reviewing the resulting abstracts, and excluding all results where the language of publication was not English.

In some cases, a psychometric study conducted by the instrument developer exists, but is not publicly available (approximately 40 cases). In these cases, we used Google Scholar to search for articles and reports citing the psychometric study (using the "cited by" search feature and the search terms "Full Name of Instrument" validity OR reliability) to determine whether the relevant information had been cited and reported in a secondary source (such as an instrument review, a dissertation, published test review, handbook entry, encyclopedia entry, or a supplemental psychometric investigation).

In other cases (approximately 25), no information about a psychometric study conducted by the instrument developer could be located, in which case we used an alternative search process. First, we located the article, publication, or website that was the original source of the instrument and determined whether it cited any psychometric studies referenced in the original source. This could include studies conducted by other researchers or older studies conducted by the instrument developer. We attempted to track down any studies so cited or referenced to obtain the relevant technical information. If we were unable to locate information on reliability or validity, we noted that in the database. Ultimately, we were unable to locate this information for 36 instruments.

The final step was to contact the developer of each assessment, share our coded summaries, and request corrections or additions. In addition to addressing factual errors, this process provided developers an opportunity to apprise us of any new evidence of validity or reliability that had been gathered and to inform us about related assessments that we might not have identified in our original search.

Pilot Development and Testing

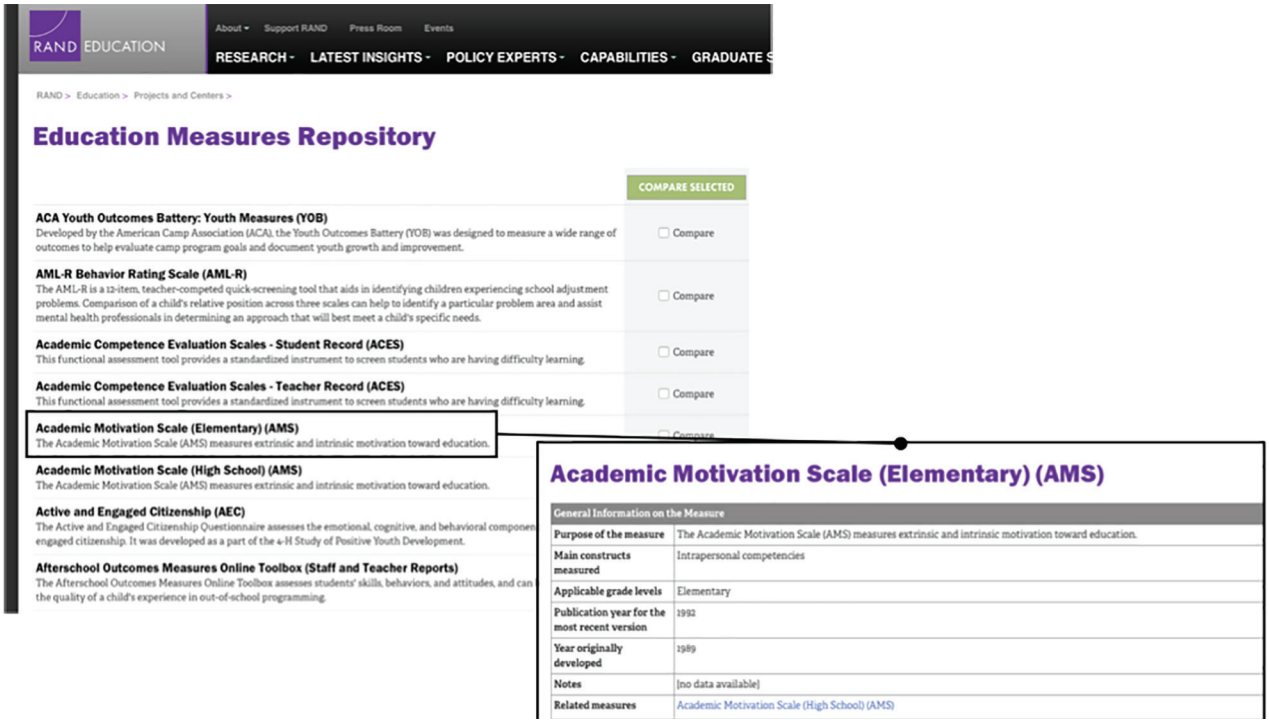
A major goal was to make the repository as user-friendly as possible. A key part of the development process involved ongoing discussions among researchers and RAND's web design team about the repository design. Pilot testing took place over a six-month period with five test users. These test users were education researchers, website developers, and measurement experts. Key features of the repository, including the search engine and presentation of data in a table format, were closely scrutinized. We also paid careful attention to how we organized and labeled data. For example, the "related measures" tag was added to alleviate potential confusion regarding various versions of an assessment (e.g., student, teacher, and parent versions of a survey) and to make it easier for users to ensure they had all desired data on all the assessments—if they wanted only the "short form," for example.

After coding all of the information about the assessments that satisfied the inclusion criteria and entering this information into the database, we began building a pilot version of the repository (Figure 2) using the findings from the feasibility study (Yuan, Stecher, and Hamilton, 2015) as guidance. In addition to providing the information on the 24 coded dimensions in Table 3, the repository is searchable and can be filtered by keyword to facilitate the user's ability to locate assessments of interest. There are multiple search options, including options to narrow the search iteratively. Search results can be sorted by any of the coded dimensions in Table 3. In addition, users can select and compare up to three assessments at a time in a side-by-side display. Additionally, all information about assessments is displayed in the same template to make it easier to locate the information of interest to a user (Yuan, Stecher, and Hamilton, 2015).

We also assembled an advisory panel of practitioners and measurement experts to use the pilot version of the repository and provide feedback and suggestions for improvement. The project team and RAND's web design team reviewed and synthesized this feedback and revised the website accordingly.

Along with the online repository, we have created a supplemental user guide that explains both the features and usability of the online platform. The guide also provides details about terminology and the data contained in the repository. We are collaborating

Figure 2. Prototype of the Repository



Along with the online repository, we have created a supplemental user guide that explains both the features and usability of the online platform.

with the AWG to develop guidance to help practitioners and other users interpret the information about technical quality and make informed decisions about what measures to adopt and how to use them to inform decisionmaking.

Key Challenges

We encountered a number of challenges during the development of the repository. The first was the need to make the repository user-friendly: Information can be easily accessed and used by stakeholders with differing levels of familiarity with assessment. We approached this challenge by employing an iterative design process that closely involved researchers and members of RAND's web design team in the development and refinement of key features of the repository, including the search engine and the tabular displays of information in the assessment summaries. We also paid careful attention to how we organized and labeled data (for example, the "related measures" tag, as discussed earlier). In another instance, based on feedback from the design process, the "Fee" column tags were reworked to clarify that a measure labeled as "No fee" was available—in cases where the measure was not accessible, an "N/A" tag was used.

Second, we encountered a number of assessments with very limited descriptive information. In some cases, the measures were dated and the web link was dead or nonexistent. In other cases, a measure was created for the purpose of research, and although a psychometric study and details about the measure existed, a physical copy of the measure was not attainable—the creators likely did not intend for the measure to be used by others outside their study. In addition, as previously described, some measures are proprietary, and even in the cases where information on these proprietary measures was available, we did not have permission to use detailed information in the repository. We decided to include these measures in the repository even when we could not find all the desired information. This decision stems from the fact that information about the measure, particularly its psychometric properties, might still be relevant to various stakeholders in the research field. In addition, it is not implausible that the creators of the measure could be contacted by others who wish to inquire about acquiring a copy of the measure.

A third challenge is encouraging appropriate use and discouraging uses that are not warranted based on the available evidence. This is of particular concern because the repository was designed to make information about assessments available to a wide range of potential users. We cannot ensure that measures will not be misused, but we would argue that making information about the measures and their technical quality more widely available will ultimately lead to improved practices and will reduce the risks of misuse. The practitioner guide that we have are developing will include guidance to promote appropriate use, particularly for practitioners who might not have training in how to interpret information about the technical quality of assessments.

A final, ongoing challenge is that the repository will retain its utility only if it is frequently updated. Research and development efforts in measurement of these complex skills are widespread, including efforts to improve existing measures and to develop new ones. For example, a recent AWG design challenge (McKown, Read, and Bookman, 2017) resulted in submission of 20 ideas for new assessments of intra- or interpersonal competencies, some of which might become available for use over the next few years.[8]

We plan to maintain this repository as a tool for researchers, practitioners, and policymakers. In addition, we hope to update it periodically to ensure that it reflects new developments in the measurement of interpersonal, intrapersonal, and higher-order

cognitive competencies. In this way, the repository can become a key element in disseminating promising new assessments to researchers, policymakers, and practitioners.

Conclusion

Educators, policymakers, and others have expressed growing interest in promoting and assessing students' interpersonal, intrapersonal, and complex cognitive skills. This has generated a clear need for a trustworthy source of information about assessments and their characteristics, along with guidance on appropriate use of those assessments. This report describes our approach to creating a resource that can address the needs of practitioners, policymakers, and researchers who are interested in capturing a broader range of information about students' competencies than is available through commonly used standardized tests.

Although we are hopeful that the repository will benefit practitioners, policymakers, and researchers, the need to keep the scope reasonable required us to exclude assessments that might be of interest to other audiences. We hope that this effort will lead to other efforts to develop similar resources for expanded age groups (e.g., pre-K, postsecondary) and other domains (e.g., school climate and culture). Assessments and data can promote better practice, policy, and research, but only if users have the access and information they need to incorporate these resources into their work effectively.

Notes

[1] We use the terms *assessment* and *measure* interchangeably in this report. One reason for not settling on a single term is because some of the instruments in the repository are referred to as assessments while others are referred to as measures.

[2] We received feedback from colleagues and advisers that the word "repository" would not be understood by many of our intended users. After much deliberation, we decided to call the online web-based tool "the RAND Assessment Finder." However, *repository* is the term we have used with our funders and within the development team, and we decided to use it in this report for simplicity.

[3] The Assessment Work Group (AWG), which was convened by the Collaborative for Academic, Social, and Emotional Learning (CASEL), is leading an effort to produce a related resource that will provide information about assessments to practitioners. The AWG tool is provides more-detailed guidance for practitioners and includes fewer assessments than the RAND repository. These efforts are intended to be complementary, and the RAND and AWG teams have collaborated on both projects. At the time of this writing, the specific scope and features of the AWG tool had not been finalized.

[4] This is the number of assessments in the repository as of May 2018.

[5] As already mentioned, the AWG was convened by CASEL to produce a related resource that will provide information about assessments to practitioners.

[6] The fact that some assessments were proprietary means that it was not always possible to obtain information about some coded information—particularly information about score reliability and evidence of validity. Where necessary, we indicated that "no information was available in the references reviewed" in the relevant repository fields.

[7] The EASEL codes contained six domains, 22 categories, and 178 subcategories.

[8] At the time of this writing, two additional design challenges were planned.

References

Almlund, Mathilde, Angela Lee Duckworth, James J. Heckman, and Tim D. Kautz, *Personality Psychology and Economics*, Cambridge, Mass.: National Bureau of Economic Research, NBER Working Paper No. w16822, 2011.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, and Joint Committee on Standards for Educational and Psychological Testing, *Standards for Educational and Psychological Testing*, Washington, D.C., 2014.

American Psychological Association PsycNet, search page, undated. As of July 24, 2018: http://psycnet.apa.org/

Atkins-Burnett, Sally, Camila Fernández, Lauren Akers, Jessica Jacobson, and Claire Smither-Wulsin, *Landscape Analysis of Non-Cognitive Measures*, Seattle, Wash.: Mathematica Policy Research, 2012.

Blad, Evie, "No State Will Measure Social-Emotional Learning Under ESSA. Will That Slow Its Momentum?" *Education Week*, Vol. 37, No. 8, October 11, 2017, pp. 1–9. As of July 23, 2018: http://www.edweek.org/ew/articles/2017/10/04/no-state-will-measure-social-emotional-learning-under.html

Buros Center for Testing, "Test Reviews and Information," webpage, undated. As of July 24, 2018: http://buros.org/test-reviews-information

Casner-Lotto, Jill, and Linda Barrington, *Are They Really Ready to Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century US Workforce*, Washington, D.C.: Partnership for 21st Century Skills, 2006.

Center on Great Teachers and Leaders at American Institutes for Research, "Guide to Evaluation Projects," webpage, 2013. As of July 24, 2018: http://resource.tqsource.org/gep/

Child Trends, *Measuring Elementary School Students' Social and Emotional Skills: Providing Educators with Tools to Measure and Monitor Social and Emotional Skills That Lead to Academic Success*, Bethesda, Md., Publication #2014-37, 2014. As of July 24, 2018: https://www.childtrends.org/publications/measuring-elementary-school-students-social-and-emotional-skills-providing-educators-with-tools-to-measure-and-monitor-social-and-emotional-skills-that-lead-to-academic-success/

Core Districts, homepage, undated. As of July 23, 2018: http://coredistricts.org/

Coryn, Chris L., Jessaca K. Spybrook, Stephanie D. H. Evergreen, and Meg Blinkiewicz, "Development and Evaluation of the Social-Emotional Learning Scale," *Journal of Psychoeducational Assessment*, Vol. 27, No.4, 2009, pp. 283–295.

Duckworth, Angela L., and David S. Yeager, "Measurement Matters: Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes," *Educational Researcher*, Vol. 44, No. 4, 2015, pp. 237–251.

Education Endowment Foundation, "Measuring Essential Skills and Non-Academic Outcomes," webpage, 2017. As of July 24, 2018: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/measuring-essential-skills/

Elliott, S. N., M. D. Davies, J. R. Frey, F. Gresham, and G. Cooper, "Development and Initial Validation of a Social Emotional Learning Assessment for Universal Screening," *Journal of Applied Developmental Psychology*, Vol. 55, 2018, pp. 39–51.

Ferguson, Maria, "Washington View: ESSA Opens School Door to Social-Emotional Learning," *Phi Delta Kappan*, Vol. 97, No. 8, 2016, pp. 74–75.

Fredricks, Jennifer, Wendy McColskey, Jane Meli, Joy Mordica, Bianca Montrosse, and Kathleen Mooney, *Measuring Student Engagement in Upper Elementary Through High School: A Description of 21 Instruments*, Greensboro, N.C.: Regional Educational Laboratory Southeast, Issues & Answers series, REL 2011-No. 098, 2011.

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

Gayl, Chrisanne L., *How State Planning for the Every Student Succeeds Act (ESSA) Can Promote Student Academic, Social, and Emotional Learning: An Examination of Five Key Strategies*, Chicago, Ill.: Collaborative For Academic, Social, And Emotional Learning (CASEL), April 2017. As of July 23, 2018: http://www.casel.org/wp-content/uploads/2017/04/ESSA-and-SEL-Five-Strategies-April-2017-041717.pdf

Gough, David, Sandy Oliver, and James Thomas, eds., *An Introduction to Systematic Reviews*, Los Angeles, Calif.: Sage, 2017.

Grant, Sean, Laura S. Hamilton, Stephani L. Wrabel, Celia J. Gomez, Anamarie Whitaker, Jennifer T. Leschitz, Fatih Unlu, Emilio R. Chavez-Herrerias, Garrett Baker, Mark Barrett, Mark Harris, and Alyssa Ramos, *Social and Emotional Learning Interventions Under the Every Student Succeeds Act: Evidence Review*, Santa Monica, Calif.: RAND Corporation, RR-2133-WF, 2017. As of July 23, 2018: https://www.rand.org/pubs/research_reports/RR2133.html

Haggerty, Kevin, Jenna Elgin, and Andrew Woolley, *Social-Emotional Learning Assessment Measures for Middle School Youth*, Seattle, Wash.: Social Development Research Group, University of Washington, 2011.

Heckman, James, Rodrigo Pinto, and Peter Savelyev, "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, Vol. 103, No. 6, 2013, pp. 2052–2086.

Hough, Heather, Demetra Kalogrides, and Susanna Loeb, *Using Surveys of Students' Social-Emotional Learning and School Climate for Accountability and Continuous Improvement*, Stanford, Calif.: Policy Analysis for California Education, 2017. As of July 23, 2018: http://www.edpolicyinca.org/

Jones, Stephanie, *The Taxonomy Project*, Cambridge, Mass.: Harvard University, undated. As of August 1, 2018: https://easel.gse.harvard.edu/files/gse-easel-lab/files/taxonomy_handout_0.pdf

Kyllonen, P. C., "Measurement of 21st Century Skills Within the Common Core State Standards," paper presented at the Invitational Research Symposium on Technology Enhanced Assessments, Washington, D.C., May 2012, May, pp. 7–8.

Levy, Frank, and Richard J. Murnane, *The New Division of Labor: How Computers Are Creating the Next Job Market*, Princeton, N.J.: Princeton University Press, 2005.

Massachusetts Department of Elementary and Secondary Education, *Social and Emotional Learning for All*, Malden, Mass., undated. As of August 2, 2018: http://www.doe.mass.edu/candi/sel/sel-all.pdf

Melnick, Hanna, Channa M. Cook-Harvey, and Linda Darling-Hammond, *Encouraging Social and Emotional Learning in the Context of New Accountability*, Palo Alto Calif.: Learning Policy Institute, 2017.

Nagaoka, Jenny, Camille A. Farrington, Stacy B. Ehrlich, and Ryan D. Heath, *Foundations for Young Adult Success: A Developmental Framework*, Chicago, Ill.: University of Chicago Consortium on Chicago School Research, 2015.

National Center on Education and the Economy, *Tough Choices or Tough Times: The Report of the New Commission on the Skills of the American Workforce*, Washington, D.C., executive summary, 2007. As of August 2, 2018: http://www.ncee.org/wp-content/uploads/2010/04/Executive-Summary.pdf

National Collaboration for Youth, *A Shared Vision for Youth: Common Outcomes and Indicators*, Washington, D.C., 2012. As of August 2, 2018: http://www.collectiveimpactforum.org/sites/default/files/NCYCommonOutcomes_0.pdf

National Research Council, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, Washington, D.C.: National Academies Press, 2012.

Noltemeyer, Amity, Allison Dimick, Marissa Smith-Millman, Kevin Shaw, and Katelyn Palmer, *Mental Health, Social-Emotional, and Behavioral Screening and Evaluation Compendium*, Oxford, Ohio, Center for School-Based Mental Health Programs, Miami University, 2015.

Partnership for 21st Century Skills, *21st Century Skills, Education, and Competitiveness: A Resource and Policy Guide*, Tucson, Ariz., 2008.

Performance Assessment Project, *The Performance Assessment Resource Bank*, homepage, undated. As of July 23, 2018: https://www.performanceassessmentresourcebank.org/

Reeves, Richard V., Joanna Venator, and Kimberly Howard, *The Character Factor: Measures and Impact of Drive and Prudence*, Washington, D.C.; Center on Children and Families, 2014.

Saavedra, Anna Rosefsky, and V. Darleen Opfer, "Learning 21st-Century Skills Requires 21st-Century Teaching," *Phi Delta Kappan*, Vol. 94, No. 2, 2012, pp. 8–13.

Sedlacek, William, *Measuring Noncognitive Variables: Improving Admissions, Success and Retention for Underrepresented Students*, Sterling, Va.: Stylus Publishing, 2017.

SEL Solutions at American Institutes for Research, *Are You Ready to Assess Social and Emotional Development?* Washington, D.C., 2010.

Soland, Jim, Laura S. Hamilton, and Brian M. Stecher, *Measuring 21st Century Competencies: Guidance for Educators*, Santa Monica, Calif.: RAND Corporation and Global Cities Education Network: Asia Society, November 2013.

Stecher, Brian M., and Laura S. Hamilton, *Measuring Hard-to-Measure Student Competencies: A Research and Development Plan*, Santa Monica, Calif: RAND Corporation, RR-863-WFHF, 2014. As of August 2, 2018: https://www.rand.org/pubs/research_reports/RR863.html

Stillman, S. B., P. Stillman, L. Martinez, J. Freedman, A. L. Jensen, and C. Leet, "Strengthening Social Emotional Learning with Student, Teacher, and Schoolwide Assessments," *Journal of Applied Developmental Psychology*, Vol. 55, 2018, pp. 71–92.

U.S. Department of Education, *Massachusetts Consolidated State Plan Under the Every Student Succeeds Act (ESSA)*, Washington, D.C., 2017. As of July 23, 2018: https://ed.gov/admins/lead/account/stateplan17/macsa2017.pdf

West, Martin R., *Should Non-Cognitive Skills Be Included in School Accountability Systems? Preliminary Evidence from California's CORE Districts*, Washington, D.C.: Center on Children and Families, Evidence Speaks Reports 1, No. 13, 2016.

William and Flora Hewlett Foundation, "Deeper Learning," webpage, undated. As of August 2, 2018: http://www.hewlett.org/programs/education/deeper-learning

Wilson-Ahlstrom, Alicia, Nicole Yohalem, David DuBois, Peter Ji, Barbara Hillaker, and David P. Weikart, *From Soft Skills to Hard Data: Measuring Youth Program Outcomes*, Washington, D.C.: Forum for Youth Investment, 2014.

Yuan, Kun, Brian M. Stecher, and Laura S. Hamilton, *The Feasibility of Developing a Repository of Assessments of Hard-to-Measure Competencies*, Santa Monica, Calif.: RAND Corporation, RR-1204-WFHF, 2015. As of July 23, 2018: https://www.rand.org/pubs/research_reports/RR1204.html

About This Report

This report describes the development of an online repository of social and emotional learning assessments that contains more than 200 curated and publicly available assessments, along with associated descriptive and evaluative information. The repository was designed to benefit three distinct user groups. For practitioners, the repository enables users to explore what assessments are available and obtain key information about what they are designed to measure, how they operate, what demands they place on students and teachers, and what kinds of uses their scores support. For researchers, the repository is a place to identify assessments—both operational and under development—related to a given construct, quickly scan the evidence about their reliability and validity, and contact the developers for more information. For policymakers, the repository offers a source of information to see whether assessments exist to support various policy options.

The authors are grateful to the generous funders of this work: The William and Flora Hewlett Foundation, The Wallace Foundation, The Raikes Foundation, the Einhorn Family Charitable Trust, and the Overdeck Family Foundation. Our Advisory Committee provided useful feedback on the design and functioning of the online repository. The committee consisted of Lauren J. Bierbaum, Katie Buckley, Michelle Clark, Mari Jones, Sally Grubb Kingston, Clark McKown, Jim Soland, and Jeremy Taylor. We also want to acknowledge the helpful feedback we received from Cathy Stasz, Chris Doss, and Jeremy Taylor, and the support from Arwen Bicknell in preparing the report for publication.

Repository development work has been conducted by RAND Education and Labor, a division of the RAND Corporation that conducts research on early childhood through postsecondary education programs, workforce development, and programs and policies affecting workers, entrepreneurship, and financial literacy and decisionmaking. Funding to support the repository has been provided by the William and Flora Hewlett Foundation, The Wallace Foundation, The Raikes Foundation, the Einhorn Family Charitable Trust, and the Overdeck Family Foundation. These foundations are part of the Funders' Collaborative for Innovative Measurement (FCIM), a group of 15 private foundations and one federal agency partner, which was created during White House meetings held in 2015 to explore areas for coordination, collaboration, and information-sharing that will advance the state of and appropriate usage of interpersonal and intrapersonal measures by educators, policymakers and researchers.

More information about RAND can be found at www.rand.org. Questions about this report should be directed to Jonathan_Schweig@rand.org, and questions about RAND Education and Labor should be directed to educationandlabor@rand.org.

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors. RAND® is a registered trademark.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

For more information on this publication, visit www.rand.org/t/RR2508.

© Copyright 2018 RAND Corporation

www.rand.org

RR-2508-FCIM