

The Feasibility of Developing a Repository of Assessments of Hard-to-Measure Competencies

Kun Yuan, Brian M. Stecher, Laura S. Hamilton



For more information on this publication, visit www.rand.org/t/rr1204

Published by the RAND Corporation, Santa Monica, Calif.

© Copyright 2015 RAND Corporation

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 its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.html.

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.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

Preface

In 2010, the William and Flora Hewlett Foundation undertook a new strategic initiative focused on students' mastery of core academic content and their development of "deeper learning" skills (e.g., critical thinking, problem solving, collaboration, communication, and learning how to learn). The foundation engaged the RAND Corporation to conduct research related to the conceptualization and measurement of skills for deeper learning. The current project builds on earlier work done at RAND and elsewhere to explore the feasibility of and challenges associated with building a repository of assessments of hard-to-measure competencies, such as those associated with deeper learning. This project is designed to inform decisions about a larger effort to build such a repository to help practitioners, researchers, and policy-makers assess hard-to-measure competencies for various purposes.

Other reports written by RAND authors related to the Hewlett Foundation's interests in deeper learning include

- Brian M. Stecher and Laura S. Hamilton's *Measuring Hard-to-Measure Student Competencies: A Research and Development Plan* (RAND, 2014, www.rand.org/t/RR863)
- Kun Yuan and Vi-Nhuan Le's *Measuring Deeper Learning Through Cognitively Demanding Test Items: Results from the Analysis of Six National and International Exams* (RAND, 2014, www.rand.org/t/RR483)
- Susannah Faxon-Mills, Laura S. Hamilton, Mollie Rudnick, and Brian M. Stecher's *New Assessment, Better Instruction? Designing Assessment Systems to Promote Instructional Improvement* (RAND, 2013, www.rand.org/t/RR354)
- Jim Soland, Laura S. Hamilton, and Brian M. Stecher's *Measuring 21st Century Competencies: Guidance for Educators* (Asia Society, 2013)
- Kun Yuan and Vi-Nhuan Le's *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests* (RAND, 2012)
- Anna Rosefsky Saavedra and V. Darleen Opfer's "Learning 21st-Century Skills Requires 21st-Century Teaching" (*Phi Delta Kappan*, 2012).

The research reported here was conducted in RAND Education, a unit of the RAND Corporation, under a grant from the William and Flora Hewlett Foundation.

Contents

Preface	iii
Figures and Tables	vii
Summary	ix
Acknowledgments	xv
CHAPTER ONE	
Introduction	1
Background	1
Key Questions	2
Overview of Methods	3
Limitations of This Study	4
Structure of This Report	4
CHAPTER TWO	
Collection of Assessments for the Repository	5
Step 1: Identify Constructs of Interest	5
Step 2: Develop Measure-Selection Criteria	8
Step 3: Search for Measures	10
Step 4: Select Measures for the Repository	14
Step 5: Obtain Information About Measures Selected	14
Step 6: Organize Information About Measures	16
Summary	17
CHAPTER THREE	
Design of the Repository User Interface	19
Example Assessment Repository Websites	19
Repository Functions	19
Desired User-Interface Features	24
Summary	28
CHAPTER FOUR	
Findings and Recommendations	29

APPENDIXES

A. The Database of Measures Developed in This Feasibility Study	35
B. Measure Profile for California Critical Thinking Disposition Inventory	41
C. Descriptions of Example Websites Reviewed	45
Abbreviations	51
References	53

Figures and Tables

Figures

S.1.	An Overview of the Procedure to Build and Maintain the Repository and Major Considerations at Each Step	xiii
3.1.	Screen Shot of the Search Interface of ROMR	25
3.2.	An Example of Comparisons of Multiple Measures on the RAND Online Measure Repository	26
3.3.	The Template Used to Present Information About Measures Listed on ROMR.....	27
3.4.	Use of Tabs to Display Information About a Measure on the DREAM Website.....	28
4.1.	An Overview of the Procedure to Build and Maintain the Repository and Major Considerations at Each Step.....	33

Tables

S.1.	Steps in Collecting Assessment and Challenges in Doing So	xi
S.2.	Functions and Key Features of an Assessment Repository Website	xii
2.1.	Steps in Collecting Assessment and Challenges in Doing So	6
2.2.	Lists of Specific Constructs of Hard-to-Measure Competencies Included in the Feasibility Study	7
2.3.	Dimensions on Which Measures Might Vary.....	11
2.4.	Categories of Information for Selected Measures and Examples.....	15
3.1.	The 12 Example Repository Websites	20
3.2.	Functions and Key Features of an Assessment Repository Website	21
3.3.	Summary of Major Functions of Eight Assessment Repositories	22
A.1.	Main Features of Measures Selected for the Repository in the Feasibility Study	36

Summary

Is It Feasible to Build a Repository of Assessments of Hard-to-Measure Competencies?

Policymakers have called on U.S. schools to develop students' higher-order cognitive skills and their interpersonal and intrapersonal competencies to keep pace with rapid developments in science and technology and increasing global economic competition. Many of these skills—including critical thinking, problem solving, collaboration, communication, and self-regulated learning—are considered “hard to measure” because they are not amenable to familiar, efficient testing approaches, such as multiple-choice testing, that are traditionally used for measuring academic achievement or because they lack clear definitions that can be easily translated into test specifications. Educators' lack of access to simple, efficient, high-quality measures hinders current efforts to promote these higher-order skills. While there are a number of relevant measures available or under development, few are widely known among educators. Furthermore, many new measures are the subject of current research and development, and information about the availability of these measures and about their suitability for use in educational contexts is not widely available outside the research community.

One step toward making measures available to educators would be to establish a database of existing assessments of hard-to-measure competencies that includes both descriptive information about the measures and available assessment materials.¹ One can envision a curated database of publicly available measures and associated descriptive and evaluative information. If well constructed, such a repository would allow potential users to search for measures that fit their needs, view the measures, and review information about their quality and use in various settings. The repository would also help potential developers identify gaps where measure development is needed, and it would provide developers with an outlet to share their tools and encourage widespread use.

To advance these efforts, the William and Flora Hewlett Foundation asked RAND to study the feasibility of developing a repository of assessments of hard-to-measure competencies and to identify the major challenges associated with its development. The feasibility study focused on two aspects of building and maintaining a repository of measures. First, we examined the procedures needed to collect, review, document, and catalog assessments for a computerized database. As part of this effort, we built a small database of assessments of hard-to-measure competencies, focusing on measures applicable to K–12 students in typical school settings in the United States. This was a practical choice to limit the scope of the feasibility study, and a working repository might also include assessments related to higher education,

¹ We use the terms *measure* and *assessment* interchangeably.

workforce preparation, and so on. Second, we examined web-based archives to identify the best functional modules and user-interface features to incorporate into a repository of measures. The feasibility study emphasized the collection of assessment information and materials more than the design of a website. In the second part of the study, we reviewed existing websites to identify important features, but we did not build a web-based repository, nor did we fully estimate the cost of doing so. Overall, this feasibility study was exploratory in nature and was limited to answering the following questions:

1. How feasible is it to build a repository of assessments of hard-to-measure competencies?
2. What are the potential challenges in building such a repository?
3. What design features will make the repository most useful to potential users?
4. What does experience from doing this feasibility study suggest for building such a repository in a future project?

We tested multiple searching strategies to assemble information about assessments of hard-to-measure competencies, using as a starting point an existing list of 130 measures collected as part of a prior RAND project on 21st-century competencies (Soland, Hamilton, & Stecher, 2013). We also tested general searching strategies by looking for additional assessments of hard-to-measure competencies in a variety of places, including databases of educational and psychological assessments, research articles and reports, projects funded by the Institute of Education Sciences and the National Science Foundation, the websites of commercial assessment providers, and other assessment repository websites. For each measure, we tried to obtain as much information as possible from public sources, and we interviewed assessment developers and providers for more information, when needed.

To explore the practical issues of developing an online repository, we reviewed 12 relevant repository websites and compared their features. We also interviewed ten potential users, including teachers, school administrators, researchers, and leaders of nonprofit education organizations, to collect feedback about their needs for such a repository.

What Did We Learn?

Overall, the results are encouraging, but we identified some challenges to be overcome in assembling the information needed for a repository.

It Is Possible to Find Assessments of Hard-to-Measure Competencies, but There Will Be Challenges in Gathering Relevant Information

Our preliminary efforts to assemble a repository convinced us that a large number of relevant assessments could be identified and cataloged, but gathering all the information about them was not always easy. Through the searches we conducted, we found measures that assess cognitive, interpersonal, and intrapersonal competencies and that could be included in a repository. Some are available in other databases that include, but are not limited to, the kinds of measures our study focuses on. For instance, the American Psychological Association's PsycTESTS® database provides access to more than 20,000 assessments that have been used in psychological research, some of which are assessments of hard-to-measure competencies that could be applied in K–12 education settings. Similarly, the Buros Center for Testing at the University of

Nebraska, Lincoln, has an assessment-review database that includes information about more than 3,500 assessments, including many measures in the domains of interest to this study. We also found relevant measures by searching through university-based research centers and commercial publishers. Our experience suggests that one could identify a substantial number of seemingly appropriate measures from which to further select measures for the repository. We also identified some challenges at each step along the way.

We identified six steps that should be followed to collect assessments and relevant information and prepare them for the repository, starting with clearly identifying the constructs of interest, including developing search criteria and conducting searches, and concluding with collecting and organizing the information about each identified measure. At each step along the way, we also encountered challenges, as shown in Table S.1.

To Be Effective, an Online Repository Should Include Five Key Functions—Search, Display, Evaluate, Use, and Connect

As noted, we reviewed 12 websites—eight repositories of assessment tools, two repositories of online learning courses, and two collections of audio and video files—to identify key functions that would be essential for an effective repository of assessments. Based on this review and our own understanding of functions that an online repository needs, we identified five major functions that an effective assessment repository would need to perform; we also derived a few key features for each (see Table S.2).

Interviews with potential repository users confirmed the potential value of such a repository and highlighted the importance of these five functions to make the repository usable.

Table S.1
Steps in Collecting Assessment and Challenges in Doing So

Step	Challenge
Identify the major constructs whose assessments are to be included in the repository	Difficulty in identifying a finite set of clearly defined constructs to guide the search for and selection of measures
Develop a set of criteria to guide the search and to determine which measures should be selected for the repository	Lack of clarity about the intended users and their needs Need to establish measure-selection criteria when certain features and information about measures are not available
Systematically search for potential measures of these constructs and collect basic descriptive information, which will be used to narrow the set of measures included	Wide variation in the form, focus, and function of existing measures, which will require the developers of the repository to make difficult choices Need to find noncommercial assessments and newly developed assessments, which are likely to be the measures of greatest interest when it comes to hard-to-measure competencies
Select measures for the repository based on the criteria established in step 2 and the basic descriptive information collected in step 3	Need to make decisions when information that is needed for such decisions is unavailable (e.g., it might be difficult to tell which age group a measure is appropriate for)
Collect a common set of information about the measures identified for inclusion in the repository	Need to find all the information one would want to include in the repository
Organize the collected information in ways consistent with the design of the repository	Need to coordinate the information being assembled with the functionality of the web-based infrastructure of the repository

Table S.2
Functions and Key Features of an Assessment Repository Website

Website Function	Key Feature
Search (i.e., locate measures of interest)	Provide multiple search options, including options to narrow the search iteratively Display the searching filters and results in two parts, vertically, on the same page Support the sorting of searched results Enable users to choose and compare multiple measures Cross-list measures under multiple, similar construct names
Display (i.e., show information about measures)	Use the same template to present detailed information about measures Use multiple tabs to display different types of information about an assessment so that users can learn all aspects of an assessment without switching to another web page
Evaluate (i.e., access information about quality and utility)	Provide guidance on how to interpret evidence of technical quality and how to select measures for particular uses Provide technical-quality evidence and reviews of assessments
Use (i.e., allow users to use assessment information and document and administer selected measures online)	Allow users to export information about measures they are interested in Provide links to the developer's or provider's website Allow users to administer measures they select online and collect assessment results in a convenient way
Connect with a larger user community	Send regular newsletters to inform users about updates in the repository Allow users to provide comments about measures and participate in discussion forums to share their experience with using these measures

Most interviewees indicated that they and their colleagues would be interested in using this repository to find measures of higher-order cognitive skills, interpersonal competencies, and intrapersonal competencies. Although some users are exploring how to assess these competencies using performance assessments and observation rubrics, they described a great need for support to find quality measures of these competencies, integrate the measures into the existing curricular and learning models, administer the measures online to collect data, and provide actionable suggestions based on the results. Although a repository cannot by itself accomplish all of these aims, it can serve as an important first step in helping users find measures and evaluate their suitability for information decisionmaking. Users consider a powerful search function, information about the measure and its technical properties, guidance on how to use the repository, and support for online administration of the assessments to be important features of the repository. Of course, our user sample was small and not intended to be representative of the larger population of likely users, but it does provide some suggestive evidence of potential demand for a repository with the functions we have described in this report.

A Number of Actions Will Need to Occur to Develop an Actual Repository

In one sense, our study might be considered a pilot test for building a repository. We went through most of the steps required to construct an actual repository. Figure S.1 summarizes the steps we think will be necessary to develop a repository.

Figure S.1
An Overview of the Procedure to Build and Maintain the Repository and Major Considerations at Each Step

1. Define the goals and purposes of the repository	<ul style="list-style-type: none"> Identify users and user needs Identify major functions of the repository website
2. Identify the right expertise	<ul style="list-style-type: none"> Have an expert committee provide guidance on key decisions for the repository
3. Identify constructs of interest	<ul style="list-style-type: none"> Develop an operational definition and identify associated terms for each construct selected
4. Develop measure-selection criteria	<ul style="list-style-type: none"> Specify actions when information about selection criteria is unavailable
5. Search for measures of constructs of interest	<ul style="list-style-type: none"> Use multiple searching strategies and sources
6. Select measures for the repository	<ul style="list-style-type: none"> Collect information about measures involving measure-selection criteria and record the decisionmaking process
7. Obtain information and materials for the measures selected	<ul style="list-style-type: none"> Include measures without adequate evidence about their technical quality but provide guidance to users about how to select measures based on available information
8. Organize information and materials collected	<ul style="list-style-type: none"> Use a template to organize information and materials collected for measures
9. Merge the assessment database and web software	<ul style="list-style-type: none"> Have the substantive team and the technology team work together to jointly design the assessment database and the user interface of the repository website
10. Disseminate information about the repository to potential users	<ul style="list-style-type: none"> Disseminate through multiple channels to reach a wide range of potential users
11. Update the repository	<ul style="list-style-type: none"> Update the information and materials about measures included and add new measures Conduct regular technological reviews of the website

NOTE: Some steps in the measures for searching and selecting might have a recursive relationship and might require the repetition of steps 3–9.

Acknowledgments

We would like to thank our interview participants for sharing their thoughts about assessment of hard-to-measure competencies and potential use of the repository.

We are grateful to the sponsor of this project, the William and Flora Hewlett Foundation, and the support and feedback we received from the foundation's Education Program staff, especially Marc Chun.

We thank James Soland at Northwest Evaluation Association and Jonathan Schweig and Cathy Stasz at RAND for their thoughtful reviews and comments. We would also like to thank Paul Steinberg for his edits and Diane Bronowicz for her administrative assistance.

Introduction

Background

Rapid developments in science and technology and increasing global economic competition have raised expectations about the skills and attributes that individuals need for successful workforce participation (National Center on Education and the Economy, 2007). Many educators, business leaders, scientists, and policymakers have called for efforts to ensure that students demonstrate both academic achievement and a broader range of the skills that are believed to be important for career success and the prospects of the U.S. economy (Apple, 2008; Cisco, 2008; National Research Council, 2012; Microsoft Partners in Learning, Pearson Foundation, & Gallup, 2013). Although different names have been given to this set of skills—*21st-century competencies*, *social-emotional skills*, *deeper learning competencies*, to mention a few (see Partnership for 21st Century Skills, 2008; National Research Council, 2012; the William and Flora Hewlett Foundation, n.d.)—all the labels refer to higher-order cognitive skills, interpersonal competencies, and intrapersonal competencies that are not typically measured by the standardized achievement tests that dominate current accountability and assessment systems in K–12 education. These higher-order cognitive skills of interest include critical thinking, problem solving, collaboration, communication, and self-regulated learning skills, among others. Many could be considered hard to measure,¹ because they are not amenable to familiar, efficient testing approaches, such as multiple-choice testing, that are traditionally used for measuring academic achievement or because they lack clear definitions that can be easily translated into test specifications.

U.S. educators at all levels are under growing pressure to promote these hard-to-measure competencies, and there have been various domestic efforts to support such work, including the Deeper Learning Initiative of the William and Flora Hewlett Foundation (Huberman, Bitter, Anthony, & O’Day, 2014; Yuan & Le, 2012, 2014).² Efforts to develop these competencies among students would benefit from high-quality measures of these competencies

¹ Although some of the measures that we reviewed and included in the feasibility study are self-reported scales and might not be considered to be hard to measure, we used the term *hard to measure* mainly to capture the challenge of assessing most of the higher-order cognitive skills, interpersonal competencies, and intrapersonal competencies that are not typically measured by commonly used standardized achievement tests. We use the terms *measure* and *assessment* interchangeably.

² We are aware of the international efforts to promote these hard-to-measure competencies and of projects to develop innovative measures of such competencies, such as the assessment of collaboration skills in the Program for International Student Assessment (PISA) and assessments of hard-to-measure competencies developed for college students and working professionals. We reviewed some of these assessments in this study. However, based on the scope for this feasibility study, they were not included in our database developed. The developers of the repository may define a larger scope of measures to be included in the repository based on the repository users and their needs.

(Soland, Hamilton, & Stecher, 2013), which can help practitioners assess individual students' progress while also informing decisions about improvement in instruction and professional development. However, available educational and psychological assessments do not adequately address this need (National Research Council, 2012; Stecher & Hamilton, 2014). Although various individuals and groups have made efforts to develop new tools to assess one or more of these hard-to-measure competencies (Herman & Linn, 2013), neither the quantity nor the quality of existing assessments of hard-to-measure competencies is sufficient to meet the needs of practitioners, researchers, and policymakers (National Research Council, 2012; Stecher & Hamilton, 2014). In addition, much of the current approach to measuring these competencies is conducted in a fragmented style, with specific organizations measuring the constructs they are interested in, on a small scale, in their specific contexts and with little coordination across these diverse efforts. Substantially more research and development are needed to produce a sufficient set of quality assessments of hard-to-measure competencies and expand the emphasis given to the development and measurement of these hard-to-measure competencies that are important for learning and working in the future; doing so is a long-term mission, because of various technical, political, and economic challenges (National Research Council, 2012; Stecher & Hamilton, 2014).

One step toward completing this long-term mission is to establish a database of existing assessments of hard-to-measure competencies so that potential users can search for measures that fit their needs and potential developers can identify gaps that exist. To advance these efforts, the Hewlett Foundation asked RAND to conduct a study to assess several aspects of the feasibility of developing a repository of assessments of hard-to-measure competencies and to identify the major challenges associated with it. The major focus of this study was to collect and catalog a sample of assessments and information about their quality and use this exercise to explore potential challenges in developing this repository. To a lesser extent, we also identified features that the repository should adopt to ensure that it meets potential users' needs.³ When completed, such a repository can help practitioners, researchers, and policymakers find measures they need, based on associated descriptive and evaluative information, and provide them with assessment materials, if available, to assess hard-to-measure competencies for various purposes. Moreover, the process of building this repository can provide the field with an understanding of the current status of existing measures, which will be useful in setting priorities for future research and development.

Key Questions

For this study, we considered a repository to be a curated database of publicly available measures and associated descriptive and evaluative information that is easy to search. In particular, we focused on the substantive and technological aspects of building a repository of measures. For the substantive aspect, we explored the procedures to collect assessments for the repository and built a small database of assessments of hard-to-measure competencies that can be used to assess K–12 students in typical school settings in the United States. For the technological

³ Obtaining a comprehensive understanding of the nationwide demand for such a repository and the cost to develop and maintain this repository was beyond the scope of this study. However, we did interview a small group of potential users when addressing the minor focus of this study.

aspects, we examined what functional modules and desirable user interface features a repository of measures should have, based on 12 example repository websites. In this feasibility study, we gave more emphasis to the substantive aspect than to the technological aspect. We did not build a web-based repository, nor did we fully estimate the cost of doing so. Overall, this feasibility study was exploratory in nature and was limited to answering the following questions:

1. How feasible is it to build a repository of assessments of hard-to-measure competencies?
2. What are the potential challenges in building such a repository?
3. What design features will make the repository most useful to potential users?
4. What does experience from doing this feasibility study suggest for building such a repository in a future project?

Overview of Methods

Building such a repository is likely to involve two major activities. One is to assemble information about available assessments of hard-to-measure competencies, which entails searching for and identifying measures to include, gathering descriptive and evaluative information about them, and organizing the information into a common format for presentation. The other major activity involves designing and developing a website to allow practitioners, researchers, and policymakers to access the information that was collected and organized. Our study explored the feasibility and challenges associated with both activities, with greater emphasis on the first.

We used multiple searching strategies to prepare input materials for the repository. We started with a list of 130 measures collected in a database as part of a prior project on 21st-century competencies (Soland, Hamilton, & Stecher, 2013). We also searched in databases of educational and psychological assessments, such as the American Psychological Association (APA) PsycTESTS® database, for additional measures that could be used to assess these competencies. In addition, we reviewed projects recently funded by the Institute of Education Sciences and the National Science Foundation that might be developing assessments of hard-to-measure competencies. Moreover, we searched for potential measures of interest from commercial assessment providers, other repository websites, and databases of research articles and reports. (See more details in Chapter Two.) We reviewed information obtained about each potential measure of interest and contacted assessment developers and providers for more information when needed.

To explore the practical issues of developing an online repository, we reviewed 12 repository websites and compared their features. We also interviewed ten potential users to collect feedback about users' needs for such a repository. These users included teachers, school administrators, researchers, and leaders of nonprofit education organizations. Interviewees included potential users whose organizations have done work to promote hard-to-measure competencies, as well as potential users who were affiliated with organizations that have not explicitly addressed these competencies. We interviewed users about their needs for such a repository and how they might use it if it existed.

Limitations of This Study

A comprehensive examination of user demands for such a repository is beyond the scope of this project. Based on discussions with the Hewlett Foundation and with others who have participated in meetings convened by the foundation, we have inferred that practitioners, researchers, and policymakers could benefit from such a repository to find tools to address their assessment needs. Although results of the small-scale group of user interviews suggested that these potential users and their colleagues would very likely use the repository if it were built, the lack of data from a larger and more-representative sample limits the extent to which our findings can be generalized to support inferences about likely demand for such a repository.

A thorough cost analysis for building and maintaining this repository is also not covered by the available resources for this study. We provided some general suggestions about how to plan for financially maintaining the repository after it is built, based on the example repository websites we reviewed.

Structure of This Report

The remainder of this report describes the tasks we carried out, the challenges we identified, and the recommendations we developed about the next steps in building the repository. Chapter Two describes activities and findings about collecting and preparing materials for the repository, addressing the first and second key questions about the feasibility of, and challenges associated with, developing a repository of assessments of hard-to-measure competencies. Chapter Three presents activities and findings related to the design of the repository website, addressing the third key question, about design features that will make the repository most useful to potential users. Results of potential user interviews are presented throughout Chapters Two and Three, when relevant. Chapter Four provides recommendations about developing and managing the repository. The chapter addresses the fourth key question, regarding what the feasibility study suggests for future efforts to develop and maintain such a repository.

In addition, Appendix A shows the database of assessments we selected in this project. Appendix B shows an example of a measure profile we prepared for each selected measure. Appendix C provides more-detailed information about the 12 example websites we reviewed.

Collection of Assessments for the Repository

We conducted exploratory searches for about 150 assessments of hard-to-measure competencies and collected information about these measures to examine the feasibility of and challenges associated with developing a repository of such measures. Our experience shows that there are quite a good number of measures that could be included in this repository, and it is feasible to find diversified measures that assess different types of hard-to-measure competencies.¹

Based on our experience in this study, we also identified six steps that need to be undertaken to collect assessments and assemble relevant information about them for the repository. First, developers must identify the major constructs to be included in the repository. Second, they need to develop a set of criteria to select measures for the repository. Third, they should systematically search for potential measures of these constructs and collect basic information about each measure that will be used to determine whether a measure should be included in the repository. Fourth, a set of measures needs to be selected from the pool of candidate measures identified through the search, based on the selection criteria established in step 2 and the basic information about each measure collected in step 3. Fifth, developers must decide on the types of information that should accompany selected measures on the repository website and then collect that information for the selected measures, to the extent that it is available. Finally, the information needs to be organized to prepare for the construction of the repository website. In this chapter, we describe our approaches to each of these six steps and identify the challenges encountered at each step (see Table 2.1 for an overview).

Step 1: Identify Constructs of Interest

The first step to collect and prepare assessment materials for the repository is to identify the constructs that the repository will cover. In the feasibility study, we started from a list of constructs from an earlier project on 21st-century competencies (Soland, Hamilton, & Stecher, 2013). Entries in this database covered 19 constructs under the three categories of hard-to-measure competencies, drawing on a framework developed by the National Research Council (National Research Council, 2012; see Table 2.2). Also, given the great interest in assessing interpersonal and intrapersonal competencies, we searched for additional measures that assess interpersonal and intrapersonal constructs, such as emotional intelligence, oral-

¹ Although there are other aspects that should be considered to assess the feasibility of building such a repository, such as cost and user need, this particular project focused on the feasibility of collecting and cataloging assessments and associated information for the repository. We also conducted a small-scale user study to learn about user needs for such a repository and perceptions about the potential features of this repository website.

Table 2.1
Steps in Collecting Assessment and Challenges in Doing So

Step	Challenge
Identify the major constructs whose assessments are to be included in the repository	Difficulty in identifying a finite set of clearly defined constructs to guide the search for and selection of measures
Develop a set of criteria to guide the search and to determine which measures should be selected for the repository	Lack of clarity about the intended users and their needs Need to establish measure-selection criteria when certain features and information about measures are not available
Systematically search for potential measures of these constructs and collect basic descriptive information, which will be used to narrow the set of measures included	Wide variation in the form, focus, and function of existing measures, which will require the developers of the repository to make difficult choices Need to find noncommercial assessments and newly developed assessments, which are likely to be the measures of greatest interest when it comes to hard-to-measure competencies
Select measures for the repository based on the criteria established in step 2 and the basic descriptive information collected in step 3	Need to make decisions when information that is needed for such decisions is unavailable (e.g., it might be difficult to tell which age group a measure is appropriate for)
Collect a common set of information about the measures identified for inclusion in the repository	Need to find all the information one would want to include in the repository
Organize the collected information in ways consistent with the design of the repository	Need to coordinate the information being assembled with the functionality of the web-based infrastructure of the repository

communication skills, goal setting, and self-regulated learning (Stecher & Hamilton, 2014). Table 2.2 shows the specific constructs that we included in the feasibility study. Results from the small-scale group of user interviews also suggest that these users are interested in measures of all three categories of hard-to-measure competencies. Although a few organizations are exploring measuring these competencies through performance-assessment and observational rubrics, most schools and other organizations lack a systematic approach to measuring these competencies.

One of the most significant challenges we encountered was the need to identify a finite set of clearly defined constructs to guide the search for and the selection of measures. There is an extensive body of literature on hard-to-measure competencies for K–12 students (such as Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Yeager, 2015; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Moffitt et al., 2011), and many of these competencies have been found to be associated with future success in academic or other realms. To narrow the scope of our pilot study, we focused our efforts on identifying assessments that met the needs of K–12 education practitioners and researchers; so we chose to only include constructs that have been measured in a K–12 educational setting among the general population of students, excluding assessments that are intended primarily to inform placement into special-education or gifted programs. This choice led us to exclude some constructs that are related to the list presented in Table 2.1 but that are not generally used in typical education settings, such as aspects of personality that are primarily in the realm of psychology or psychiatry. We narrowed the search decision to enable us to carry out the feasibility study within a limited scope and budget, but a working repository could be devel-

Table 2.2
Lists of Specific Constructs of Hard-to-Measure
Competencies Included in the Feasibility Study

Category	Construct
Cognitive	Academic learning
	Creativity
	Critical thinking
Interpersonal	Communication
	Collaboration
	Emotional intelligence ^a
	Global awareness
	Leadership
Intrapersonal	Ethics
	Goal orientation
	Goal setting ^a
	Grit
	Learning how to learn
	Mind-sets
	Motivation
	Resilience
	Self-control
	Self-efficacy
Self-regulated learning ^a	

^a We searched for additional measures for these constructs.

oped to cover a much broader set of domains.² Other repository developers might make different decisions about how to define the scope of a repository, but the main point is that developers must create explicit definitions before undertaking a search for measures.

Even when the constructs are selected, it is not always straightforward to determine how a specific construct should be defined. In the published literature, a construct is commonly defined differently by different authors or in different contexts. For instance, the construct of emotional intelligence has been defined in different ways: Mayer, Caruso, and Salovey (2000) defined it as the ability to reason with and about emotions, while Boyatzis, Goleman, and Rhee (2000) viewed it more in terms of personality traits, including conscientiousness, emotional stability, extraversion, and openness, as well as motivation and leadership. Given the diversity in definitions for the same construct, it is necessary to obtain an overview of such differences

² In fact, by broadening the scope of measures beyond those typically used in K–12 settings, a repository might serve as a means to generate awareness among K–12 practitioners of constructs and measures that are more prevalent in other contexts but that might be useful to address in K–12 schools.

to guide the selection of constructs of interest. In fact, researchers often achieve clarity about a construct when they try to figure out how to measure it. One of the potential benefits of the repository is that it might create pressure for assessment developers to operationalize construct definitions in ways that allow more-direct comparison and lead to clarification. It may also be helpful to provide such overviews to potential repository users to help them select measures that fit their needs. Existing articles that provide such overviews on related constructs would provide a good starting place.

A comprehensive review of the literature also helped identify multiple terms that have been used to name the same or similar construct to ensure the right scope of search for measures. For instance, such phrases as *self-control*, *self-management*, and *self-leadership* have been used to describe the set of constructs that explain individual differences in how people manage their perceptions, emotions, and behaviors to successfully achieve personal goals (Norris, 2008; Briesch & Chafouleas, 2009; Myrseth & Fishbach, 2009; Steward, Courtright, & Manz, 2010).

This diversity in labels assigned to similar constructs is common and can confuse potential repository users, who might not be aware of the ways in which different constructs overlap. For instance, previous research showed that constructs, such as conscientiousness, perseverance, and self-control, were positively related to individual performance (Moffitt et al., 2011; Educational Testing Service [ETS], 2012; Duckworth et al., 2007). Recent research also found that self-control and perseverance were clustered under the construct of conscientiousness (Robert, Chernyshenko, Stark, & Goldberg, 2005; MacCann, Duckworth, & Roberts, 2009). Understanding these overlap areas is important for ensuring that the repository includes a comprehensive set of measures of each construct of interest; consequently, the search function in the repository should be designed to address these overlapping constructs (e.g., by cross-listing measures under multiple similar constructs, where appropriate).

Step 2: Develop Measure-Selection Criteria

It might not be desirable on cost grounds to include in the repository all of the measures of the constructs that are identified in step 1. For instance, if the repository focuses on measures for K–12 students, it is not the best use of limited resources to search for measures developed for postsecondary students or working professionals. Moreover, measures included in the repository and their potential uses need to be aligned with the needs of the intended repository users. That is, if the repository is developed for use in the K–12 settings, including assessments for postsecondary students and working professionals might not address most users' needs and could make the search process more time-consuming or difficult. Thus, it makes sense to develop a set of criteria (e.g., measures developed for K–12 students) to select measures that meet the intended users' needs for the repository. Specifically, if measure-selection criteria are available before researchers start searching for measures, researchers can focus on finding information that will tell them, at the initial stage of the searching process, whether a measure meets the selection criteria (i.e., whether a measure was developed for K–12 students). If not, then they can exclude a measure from the search list and move on to other measures, which will save time.

One step that should be taken before developing measure-selection criteria is to identify the intended users and, to the extent possible, determine the needs that the repository plans to

support.³ In the feasibility study, we had to make some assumptions to develop these measure-selection criteria. We assumed that this repository would have three major groups of users: practitioners, researchers, and policymakers. Practitioners might use these measures about regular students as formative or summative assessments and might use the results to adjust instruction or guide professional development. In addition, practitioners who are in school or district leadership positions might use the assessments to determine how well an organization (e.g., a school or a district) is promoting the development of hard-to-measure competencies. Researchers might use these measures to study the effectiveness of certain intervention programs, to develop new measures, or to conduct research to examine the technical quality of existing measures. Finally, policymakers might use results from these measures to inform legislation or for other policy-related purposes. Results from the small-scale user interviews suggest that these assumptions are consistent with these potential users' intended uses of measures from the repository.

These potential users might need to administer assessments of hard-to-measure competencies to different assessment populations, such as K–12 students in schools or in out-of-school settings, postsecondary students, and working professionals. We assumed that the repository would mainly support the assessment of hard-to-measure competencies in typical K–12 settings and developed the following measure selection criteria for this study:

- measures that assess higher-order cognitive skills (such as critical thinking and creativity), interpersonal, and intrapersonal skills
- measures that can be administered to individual K–12 students
- measures that can be administered as stand-alone assessments
- measures that are not used for special-education diagnosis purposes.

One challenge in developing measure-selection criteria is how to select measures when certain features of and information about measures do not exist or cannot be accessed easily, such as established systems to deliver assessment materials, administer the assessment, score the results, and provide final reports (i.e., the infrastructure system to support the use of measures); a comprehensive set of evidence about the technical quality of measures (i.e., reliability and validity); and complete assessment materials, including assessment booklets, answer forms, scoring manuals, and guidance on how to interpret the results. For the feasibility study, we decided not to restrict the search based on these features. Moreover, because one of the goals of this repository is to support future research and development work on existing measures, the inclusion of measures without a mature infrastructure system to support use or without adequate technical-quality evidence could encourage further research and development work on these measures.

The measure-selection criteria reflect decisions we made to carry out this small-scale study. Admittedly, the decision to include measures that meet our selection criteria but do not have infrastructure systems, evidence about their technical quality, or assessment-related documents might not be best for constructing an actual repository. A future full-scale repository-

³ As this small-scale study focused on the feasibility of collecting and cataloging information about assessments and assessment materials, if available, understanding the nationwide need for such a repository was beyond the scope of this study. However, at the beginning of the effort to build the repository, we recommend defining the intended users and their needs that the repository plans to support. See more details in Chapter Four.

development effort could address this limitation by applying different selection criteria or by providing users with guidance to help inform their selection and use of measures.

Step 3: Search for Measures

The third step is to search for measures that assess the constructs of interest. In addition to carrying out the search for the measures themselves, this step should also include an effort to collect any information that would be used to determine each measure's suitability for inclusion in the repository, such as the population to which the measure is applicable and whether it can be used as a stand-alone assessment. It is not necessary at this stage to gather all of the supporting documents that would eventually be included for measures that are selected for the repository (e.g., evidence of technical quality); that process is described under step 5.

We used multiple strategies to search for measures in the feasibility study. We used the Google search engine to look for information about measures that were included in a database built in a prior project on 21st-century skills (Soland, Hamilton, & Stecher, 2013). We used measure and developer names as key words in these searches. Based on results from these online searches, we conducted further searches in databases of journal articles and through search engines to gather information about studies that have used these measures or tested their technical quality.

We then carried out several steps to search for additional measures. First, we conducted online searches using construct names as key words. Second, we looked through several databases of psychological and educational assessments, such as the APA PsycTESTS database; the database from the Buros Center for Testing at the University of Nebraska, Lincoln; and repository websites that we found during earlier searches (e.g., the Consortium for Research on Emotional Intelligence in Organizations). Third, we searched for products from major commercial for-profit and nonprofit assessment providers, such as ETS, Pearson, and ACT. Fourth, we searched the databases of projects funded by the Institute of Education Sciences and the National Science Foundation for projects that focused on developing measures of interpersonal and intrapersonal skills. Fifth, we searched large research databases, such as Education Resources Information Center (ERIC) and APA PsycARTICLES®, for studies that might have developed measures of relevant constructs using the construct names as key words. Finally, we looked for measures of constructs of interest, such as emotional intelligence, through articles that gave overviews of research on this construct; then, we searched for and reviewed articles that provide technical-quality evidence about the measures we found.

Overall, we found a good number of measures that assess cognitive, interpersonal, and intrapersonal competencies and that could be included in the repository. Some of these measures are available in other databases; these measures include, but are not limited to, the kinds of measures that our study focuses on. For instance, the APA PsycTESTS database provides access to more than 20,000 assessments that have been used in psychological research, some of which are assessments of hard-to-measure competencies that could be applied in K–12 education settings. Similarly, the Buros Center for Testing at the University of Nebraska, Lincoln, has an assessment-review database, which includes information about more than 3,500 assessments.

On the one hand, this large pool of measures, along with a variety of other sources, such as university-based research centers and commercial publishers, addressed the concern that

there were not enough assessments of hard-to-measure competencies to be included in the repository. On the other hand, the large number of measures and the knowledge, time, and resources required to understand how to search for, select, and use the appropriate measures for different purposes suggest that it is important to build this repository to make it easy for potential users, especially practitioners who are already busy with other educational responsibilities, to find appropriate assessments of hard-to-measure competencies that fit for their needs.

The measures we examined in the feasibility study varied on a number of dimensions that might be relevant to the design of the repository (see Table 2.3 for an overview). These dimensions are often used to categorize educational and psychological assessments. They are worth mentioning here because they can serve as search filters to guide repository users to locate measures of interest. The developer of the repository should carefully think about how to categorize measures and what search filters should be used to efficiently locate measures of users' interest, and the developers' efforts to search for measures should include efforts to obtain information on each of these dimensions. We discuss these dimensions below.

Target assessment population refers to the group of individuals that a measure was developed for. There are multiple subcategories under this dimension. For instance, measures can be categorized based on whether they were developed for clinical populations (e.g., individuals with learning difficulties) or nonclinical populations (e.g., middle school students, in general). Most of the measures we reviewed here are for nonclinical populations. However, if the repository is going to support the assessment of hard-to-measure competencies among clinical populations, there are measures that meet this need as well.

It is also possible to categorize the target population in terms of age or education level (e.g., elementary, secondary, and postsecondary students). In the feasibility study, we focused on measures for K–12 use. We also found a number of assessments of hard-to-measure compe-

Table 2.3
Dimensions on Which Measures Might Vary

Dimensions	Details
Target assessment population	Whether a measure was designed for clinical or nonclinical populations Test-taker populations defined by the ages or education levels of assessment takers whom the measures intended to serve
Item format	Whether items included in an assessment are multiple-choice items, extended-response items, Likert-scale items, and performance-assessment tasks
Length of an assessment	The number of items included in a measure
Time required to complete a measure	The time required to complete a measure
Requirement for technology for administration, scoring, and reporting	Whether technology is involved in the assessment administration, scoring, and reporting
Level of results being reported	Whether results are produced at the individual level
Commercialized versus noncommercialized assessments	Whether users have to pay for access to and use of assessment materials for research or educational purposes
Whether an assessment can be used as a stand-alone assessment	Whether an assessment has to be used together with other learning materials or programs

NOTE: These variations can be used as search filters.

tencies for postsecondary students. For instance, ETS developed the WORKFORCE® Assessment for Job Fit to measure six competencies that are considered important for performance in the workplace, such as initiative and perseverance, responsibility, teamwork, and citizenship. However, this measure was developed for college-level students and, consequently, is not included in our pool for this feasibility study.

Item format refers to the types of items that each measure uses to collect data about assessment takers, such as multiple-choice items, extended-response items, Likert-scale items, and performance-assessment tasks. Multiple-choice and Likert-scale items are more frequently used than other types of items, because it is easier to collect and score results on these items than on other types of items. This phenomenon exists for most educational and psychological assessments, as well as for the specific measures we reviewed in this study.

Measures we reviewed also varied substantially in their lengths and the time required completing them. For instance, some scales include only five or fewer items and take less than two minutes to complete. Then again, another measure that includes multiple sections might take a few hours to complete. Typically, it takes less time to complete a short assessment than a long assessment. However, the length of an assessment is not necessarily positively related to the time required to complete it. For instance, an assessment of creativity might have three questions, but it might take half an hour to complete it.

Whether technology is involved in the assessment administration, scoring, and reporting is another dimension that the measures we reviewed varied on. A number of providers offer multiple modes to administer an assessment, including a paper-and-pencil mode; a web- or computer-based mode; and a combination of both modes for administration, scoring, and reporting. For instance, assessment takers of the California Critical Thinking Disposition Inventory (CCTDI) can complete CCTDI in a paper-and-pencil mode; send the results to the provider, Insight Assessment,⁴ for scoring; and obtain reports that Insight Assessment provides. Assessment takers can also complete CCTDI online and obtain the results.

Measures can also be categorized based on whether their results are produced at the individual level. Most of the measures we reviewed are used to produce individual-level scores. For some measures that are administered to individuals, the results are used to produce organization-level results. For instance, the Situational Outlook Questionnaire, developed based on the early works of Dr. Göran Ekvall (1991), is administered to individuals in an institution to assess its members' perceptions of the climate for creativity, innovation, and change in the institution.

Another major dimension that measures we reviewed differ on is whether an assessment is a commercial assessment. Commercial assessments are those that users have to purchase, including assessment-related documents (e.g., assessment booklets, scoring forms, and manuals) and professional support in administration, scoring, and reporting. The assessment developer or a commercial-assessment provider is in charge of providing such materials and services. For instance, users of assessments developed by Insight Assessment have to pay that organization to administer its assessments of critical thinking skills and dispositions, score the results, and prepare the final report.

⁴ Insight Assessment is a California-based company that specializes in delivering assessments of critical thinking skills and dispositions for a wide range of assessment takers, from elementary school students to professionals in different specialized fields, such as business, law, and health. See more details at its website, <http://www.insightassessment.com>.

In comparison, for noncommercial assessments, users can obtain assessment-related documents and instructions regarding how to score and interpret the results free of charge for research or educational purposes. Some of these assessments were funded through government grants. Their assessment booklets, answering forms, and manuals were available from the developers' websites. Potential users can download such materials and use them for their own purposes, based on certain conditions (e.g., research and educational purposes, as long as original sources are cited appropriately). Other assessments were developed and used in certain research projects. Potential users can obtain information about such assessments from journal articles or research reports. These users will have to contact the developer to obtain the complete assessment booklets and manuals, and they might have to design their assessment booklets using items available from the assessment developer and score and interpret the results based on the instructions available.

The measures we reviewed also varied in terms of whether they can be used as stand-alone measures. The database that resulted from the prior project (Soland, Hamilton, & Stecher, 2013) included 17 entries that were related to technology-based learning programs, such as the Auto Tutor and iSTART programs (Graesser, McNamara, & VanLehn, 2005). These programs included measurement components designed to assess the specific knowledge and skills that were intended to be developed through the use of such programs. The close connection between the learning and measurement components in these programs makes it difficult to use these measures as stand-alone measures. Moreover, if a measure is embedded in a specific instructional activity and is designed to measure student performance on that activity, scores might not generalize beyond the narrow context of the learning programs.

We did encounter some challenges in locating measures of potential interest. For instance, noncommercial assessments are more challenging to find than commercial assessments. Although the APA PsycTESTS database is a source for noncommercial assessments, it is not complete (i.e., we found measures, through the Google search engine, that were not included in the APA PsycTESTS database). Also, assessments developed a long time ago that were not commercialized might be challenging to find. It is also harder to find newly developed measures or measures under development that have not been included in existing assessment databases. Therefore, it is necessary to continue using multiple searching strategies so that the search results from different strategies complement each other.

Another challenge is that the process to conduct searches and collect, review, and evaluate search results can be very time-consuming. For example, online searching on a particular competency (e.g., emotional intelligence) produces large numbers of "hits" (more than 26,000) that have to be reviewed and assessed. It requires extra effort to search for descriptive information about noncommercial assessments and newly developed assessments. It requires even more effort to collect, review, and evaluate technical-quality evidence for the measures selected, especially for noncommercial assessments. In addition, although most noncommercial assessment developers do permit free use of their assessments for educational and research purposes, it can be time-consuming to obtain assessment materials and negotiate the rights to share materials for other uses.

When searching for the measures of constructs identified in step 1, it is helpful to collect information about those dimensions included in the selection criteria. With such information, decisions can be made about whether to include a measure in the list for further information collection. This leads to the next step in the process: select measures for the repository.

Step 4: Select Measures for the Repository

In step 4, we applied the measure-selection criteria developed in step 2, based on basic information collected about measures in step 3, such as the measured constructs and target assessment-taker population, and identified those measures that should be included in the repository.

The main challenge we encountered when selecting measures to include in the repository was the lack of information on certain key attributes of measures, such as the age group of assessment takers, to make a decision for some measures, especially for noncommercial assessments. For instance, one assessment we identified had been developed to measure students' motivation for pursuing postsecondary education. The developer did not specify the target assessment population. The assessment items (measuring students' motivation for attending colleges) seem applicable to both college students and high school students (e.g., "University is a way for me to get into an interesting and satisfying career"; Dennis, Cavillo, & Gonzalez, 2008). However, in an article that reported on this measure, the assessment was administered only to college students. The author did not discuss the potential of administering this measure to high school students.

This situation did not happen often. But when it happened, we contacted the developer for clarification. However, for about ten measures, either the developers we contacted were not responsive or the contact information was outdated, requiring us to explore other ways to locate and reach them. For measures that the developers did not specify as intended for use in K–12 settings but that seem applicable to K–12 populations, we included only those measures that had been tested on K–12 samples. However, a decision could have been made in the opposite direction if the measure-selection criteria require including measures for college students or measures that have been tested on adults but that have the potential to be used in K–12 settings.

Step 5: Obtain Information About Measures Selected

After we selected measures for the repository, we needed to collect additional information about them, to display on the repository website. We had collected some information related to the measure-selection criteria during step 3, to search for measures, such as the constructs assessed and the target assessment population. Such information was not sufficient for users to understand and select measures for their purposes. Thus, we collected additional information about measures in step 5 to help users better understand measures and select the ones that fit for their needs.

We reviewed example repository websites and assessment databases to see what information they presented. (See Chapter Three for a comprehensive list of the example repository websites we reviewed.) The reviews suggested that we provide a range of descriptive information, information about technical quality, and examples of assessment materials, as shown in Table 2.4.

We tried to collect all three categories of information for each measure selected for the feasibility study. And we used multiple strategies to search for that information. Most of the time, we looked for such information from the sources where we found these measures. We conducted additional searches using search engines, when necessary. We also searched for and reviewed journal articles and research reports about studies that used the selected measures. In

Table 2.4
Categories of Information for Selected Measures and Examples

Categories	Examples
Descriptive information	<ul style="list-style-type: none"> Target construct(s) that a measure intends to assess The target population that a measure is designed for The number and type of questions included in a measure The time required to administer a measure How to administer the measure, score it, and report the results Cost to purchase and administer a measure, as well as score and report the results How to obtain additional information about the measure
Evidence about the technical quality of a measure	<ul style="list-style-type: none"> Evidence about reliability, including relevant sources of error Evidence about validity for specific purposes Evidence about fairness to various populations of assessment takers
Assessment materials	Assessment booklets, scoring forms, and manuals

addition, we contacted the developers and the provider for actual assessment booklets, scoring forms, manuals, and information that we could not find through other channels.

Our search results showed that the availability of descriptive information varied, but that descriptive information was typically more widely available than the other two categories of information listed in Table 2.4. It required greater effort to find, from journal articles and research reports, descriptive information for newly developed noncommercial measures than it did for commercial measures. Often, these journal articles and research reports did not provide all the descriptive information we needed. Moreover, the websites of commercial-assessment providers also varied in the details of descriptive information. In some situations, we had to contact the providers for information we could not find on their websites.

The availability of evidence about technical quality varied even more than the descriptive information. For measures that have been used on a large scale and for a long time, it was common to see some evidence about their technical quality. For newly developed measures, research articles or reports were almost the only sources that provided some evidence about measures' technical quality. It required additional efforts to search for other studies that had used these measures and reported on their technical quality.

In addition, evidence about internal consistency reliability, especially in the form of Cronbach's alpha coefficient, was more frequently reported than other reliability coefficients. This might have resulted from the fact that common statistical analysis programs, such as SPSS, Stata, and SAS, make it easy to calculate the Cronbach's alpha coefficient. However, lack of information about other types of reliability coefficients, such as test-retest reliability or rater agreement, limits the extent to which potential users can assess the technical quality of a measure for their use.

Moreover, evidence about validity—including evidence based on the content of an assessment, response processes, the internal structure of an assessment, and relationships between scores on an assessment and other variables—was less readily available than evidence about

reliability. Evidence about the internal structure of an assessment is more frequently available (typically in the form of results from factor analysis), compared with other sources of evidence about validity. In general, most measures do not have much evidence to support the validity for their intended purposes, which poses problems when repository users select measures for their purposes and highlights the need for more efforts to study the quality of assessments of hard-to-measure competencies.

The availability of actual assessment booklets, scoring forms, and manuals also varied quite a bit. For commercial measures, although paid customers can access complete assessment booklets, scoring forms (and sometimes scoring algorithms), complete manuals, and reports, developers and providers were commonly unwilling to provide copyrights to a third party, such as the owner of the repository. Developers of noncommercial measures often provided access to all assessment-related documents to potential users, but they granted free use only for research and educational purposes. Interested users will have to contact the developer for assessment materials and further discuss their plans for using these measures for nonresearch purposes.

Overall, collecting descriptive information was relatively easier than collecting evaluative information and actual assessment documents. However, we often found that the available information about a measure does not answer all the questions about its quality and utility. In these cases, we had to contact the developer or the provider to obtain information that was not readily available from public sources. In addition, we had to carefully review the available information about a measure to make sure we provided a comprehensive and objective description of it, based on information the developer or provider offered.

In terms of challenges for future developers, it might be harder to collect evaluative information and the actual assessment materials. Although developers of many noncommercial assessments let people use these assessments for free for educational or research purposes, the procedures to obtain assessment materials can be time-consuming. Moreover, it might take substantial effort to search for, review, and evaluate evidence of technical quality about measures, especially for noncommercial assessments.

Step 6: Organize Information About Measures

The final step is to organize collected information in a way that facilitates its presentation on the repository website.

In this feasibility study, we organized collected information in three formats. First, we developed a database (see Appendix A) to organize information about the key dimensions that potential users can use to search for measures of interest. In this database, each row provides a record for a given measure, while each column records a measure's attribute on a key search dimension, such as constructs measured, target assessment population, assessment format (i.e., paper and pencil only, online, or both), and languages in which a translated version is available. These search dimensions reflect the major features of the assessment that someone might use to identify an assessment as appropriate for his or her purpose.

Second, we produced a measure profile in Word files to present descriptive information about measures (see an example in Appendix B). We developed a template that used a simple question-and-answer format to provide descriptive and available evaluative information about each selected measure. In the measure profile, we also documented the sources where we obtained such descriptive information.

Third, where assessment materials are available, we downloaded such documents. Most of the time, the original file was in a Word or PDF format. We used the same format as the original documents.

There are three major benefits to organizing information about selected measures this way. First, cataloging information about selected measures on key dimensions in an Excel file allows the information technology (IT) team to use such information to develop the search engine for the repository website. Second, providing descriptive and evaluative information in a Word document allows easy review and revision, compared with providing such information in an Excel file. Providing descriptive and evaluative information about each measure using the same format also helps users locate the information they need, once they are familiar with the format. Third, providing assessment documents in a Word or PDF format allows for easy review, download, and adaptation by potential users.

Summary

Overall, our experience with the exploratory searches suggests that it is feasible to find a diversified set of assessments for hard-to-measure competencies to be included in the repository. Our experience also provided some important findings and challenges associated with each step in preparing materials for the repository. First, it is critical yet challenging to define a finite list of constructs of interest and associated terms to guide the search for measures. Second, it is necessary to use multiple searching strategies and sources to maximize the chance to identify all potential measures that assess the constructs of interest. Third, it is important to have a clear set of criteria to guide the selection of measures; however, the information needed to make a decision about a measure might not be readily available, and will require additional efforts to obtain. Fourth, descriptive information about measures is the most likely to be available, whereas actual commercial-assessment documents and evaluative information about the validity of an assessment are the least likely to be available.

Design of the Repository User Interface

In this chapter, we describe our efforts to develop an effective design for the repository's user interface. Rather than starting from scratch, we reviewed 12 existing assessment databases, assessment repositories, and online course catalogs. From these examples, we identified the major functional modules an assessment repository needs and the desirable features of the user interface.

Example Assessment Repository Websites

We reviewed 12 websites as a basis for developing our design ideas (see Table 3.1). Eight of the 12 websites are repositories of assessment tools. Two (Child Language Data Exchange System [CHILDES] and Databrary) are repositories of audio and video files, and two (Skilledup and MIT OpenCourseWare) are repositories of online learning courses. We mainly drew on the eight assessment repositories to identify the functional modules that an assessment repository should have. We studied all 12 websites to look for effective user-interface features. (See Appendix C for more details about each website.)

Repository Functions

We identified five major functions an effective assessment repository needs to perform (see Table 3.2 for an overview): (1) search (i.e., locate measures of interest), (2) display (i.e., show information about measures), (3) evaluate (i.e., access information about quality and utility), (4) use (i.e., allow users to use assessment information and documents and administer selected measures online), and (5) connect with a larger user community. Results from the user interviews also suggest that these five functions are aligned with how the interviewed users plan to use the repository website. Table 3.3 provides an overview of the functions available in each of the eight assessment repositories we reviewed.

A repository must have a good “search” function to allow users to browse for assessments relevant to their interests. All the websites we reviewed had extensive search capabilities. The search function is especially important when a repository lists a large number of items, such as the APA PsycTESTS database and the Buros Center for Testing's online assessment review database. Most websites allow users to conduct open searches (i.e., search by typing in key words) or prespecified searches (i.e., search by selecting key features of the assessments, such as grade level and subject). A “save” function would also be a helpful feature to include, so that

Table 3.1
The 12 Example Repository Websites

Repository Website	Description
APA PsycTESTS database	This database contains more 20,000 tests assessments used in psychology and education, most of which are noncommercial tests or assessments used in research projects. Website: http://www.apa.org/pubs/databases/psyc-tests/
Buros Center for Testing at the University of Nebraska, Lincoln	The Buros Center hosts an online assessment review database that provides third-party reviews of more than 3,500 commercial assessments. Website: https://marketplace.unl.edu/buros/
RAND Online Measure Repository (ROMR)	This database contains 171 measures related to psychological health and the treatment of traumatic brain injuries. Website: http://www.rand.org/multi/military/innovative-practices/measure.html
Performance Assessment Links in Science (PALS)	PALS contains science-performance assessment tasks for grade levels K–12. These tasks are referenced to the National Science Education Standards and state assessment frameworks in Illinois and Texas. Website: http://pals.sri.com
Stanford Center for Assessment, Learning and Equity (SCALE) Performance Task Database	This database contains performance assessment tasks in English language arts, mathematics, science, and social studies at K–12 levels. Website: http://performanceassessment.stanford.edu
Assessment Repository of the Society of Clinical Psychology (ARSCP)	ARSCP lists assessment tools that are publically available, have evidence of reliability and validity, and could be used by clinical and academic psychologists. Website: https://www.div12.org/assessment-repository/
Directory and Repository of Educational Assessment Measures (DREAM)	DREAM collects assessment instruments that are relevant to health science education. Website: https://www.mededportal.org/about/initiatives/dream/
The National Institutes of Health (NIH) Toolbox	The NIH Toolbox® contains measures of cognition, emotion, motor skills, and sensation for ages 3 to 85. It aims to provide a standard set of common measures to be used in different studies. Website: http://www.nihtoolbox.org/Pages/default.aspx
Child Language Data Exchange System (CHILDES)	CHILDES is a repository of first-language acquisition data. It contains transcripts and audio and video files in more than 20 languages, from 130 different corpora. Website: http://childes.psy.cmu.edu
Databrary	Databrary is a database of video, audio, and related metadata used to study human and animal development, for a variety of purposes. Website: https://nyu.databrary.org
Skilledup	Skilledup is a repository of online learning courses, available worldwide. Website: http://www.skilledup.com
MIT OpenCourseWare	This website is a repository of online learning courses from the Massachusetts Institute of Technology (MIT). Website: http://ocw.mit.edu

users could begin searches, save the results, and return to those searches later. Interviewed users also thought that the repository should have a powerful search function to help them find what they need. Because the same construct might be labeled using different names, it is necessary to cross-list measures under multiple associated names to maximize the chance that users interested in a particular construct will be able to find all relevant measures through one search using the construct name.

Displaying information about assessments is another key function of all the example websites. The types of information they provide include descriptive information and assess-

Table 3.2
Functions and Key Features of an Assessment Repository Website

Website Function	Key Feature
Search (i.e., locate measures of interest)	<p>Provide multiple search options, including options to narrow the search iteratively</p> <p>Display the searching filters and results in two parts, vertically, on the same page</p> <p>Support the sorting of searched results</p> <p>Enable users to choose and compare multiple measures</p> <p>Cross-list measures under multiple, similar construct names</p>
Display (i.e., show information about measures)	<p>Use the same template to present detailed information about measures</p> <p>Use multiple tabs to display different types of information about an assessment so that users can learn all aspects of an assessment without switching to another web page</p>
Evaluate (i.e., access information about quality and utility)	<p>Provide guidance on how to interpret evidence of technical quality and how to select measures for particular uses</p> <p>Provide technical-quality evidence and reviews of assessments</p>
Use (i.e., allow users to use assessment information and document and administer selected measures online)	<p>Allow users to export information about measures they are interested in</p> <p>Provide links to the developer's or provider's website</p> <p>Allow users to administer measures they select online and collect assessment results in a convenient way</p>
Connect with a larger user community	<p>Send regular newsletters to inform users about updates in the repository</p> <p>Allow users to provide comments about measures and participate in discussion forums to share their experience with using these measures</p>

ment documents. Descriptive information includes major characteristics of assessments, such as the constructs assessed, age level of the appropriate assessment-taker population, the format of the assessment items, and administration requirements. All eight repositories of assessments provide descriptive information about the measures they contain.

The other type of information is assessment documents, such as assessment booklets, answer sheets, scoring algorithms or rubrics, and manuals. Among the eight assessment repositories,

the Buros Center for Testing, ROMR, and the ARSCP do not provide any assessment documents. The APA PsycTESTS database provides assessment documents except for those that are proprietary. Given that this database focuses on noncommercial assessments, users can access full or partial assessment booklets for more than three-

Comments from Potential Users About Repository Functions

"It should be highly searchable and filterable. You can easily get to the particular kind of things you are looking for."

"Advanced search function. Multiple filters also make it easy to find what I want."

"The number one thing would be a strong search function. A fairly sophisticated one."

Table 3.3
Summary of Major Functions of Eight Assessment Repositories

Example Repository Website	Search to Locate Measures	Display Information About Measures		Evaluate the Quality and Utility of Measures	Use Assessment Information and Documents			
		Descriptive Information About Measures	Assessment Documents		Download Assessment Information	Provide Additional Materials	Administer Measures Online and Collect Data	Connect with a Larger User Community
APA PsycTESTS database	Open and prespecified	Yes	Yes, when available	Available technical evidence is provided	Yes	Links to related articles	No	Yes
Buros Center for Testing at the University of Nebraska, Lincoln	Open and prespecified	Yes	No	Third-party review is provided	Yes	Links to the website of the provider, when available	No	No
RAND Online Measure Repository (ROMR)	Open and prespecified	Yes	No	Available technical evidence is provided	No	User guides on how to use the repository and evaluate the technical quality of measures, links to related articles, and glossary of assessed constructs	No	No
Performance Assessment Links in Science (PALS)	Open and prespecified	Yes	Yes	Information provided not very useful	Yes	Examples of student work	No	No
Directory and Repository of Educational Assessment Measures (DREAM)	Open and prespecified	Yes	Yes	Third-party review is provided	Yes	Links to related articles	No	Yes
Stanford Center for Assessment, Learning and Equity (SCALE) Performance Task Database	Open and prespecified	Yes	Yes	Require evidence to be included	No	Teacher guide on assessment administration	No	No
Assessment Repository of the Society of Clinical Psychology (ARSCP)	Browse only	Yes	No	Require evidence to be included	No	Links to related articles	No	Yes
The National Institutes of Health (NIH) Toolbox	Prespecified	Yes	Yes	Require evidence to be included	Yes	Links to related articles, online training videos, and user manuals	Yes	Yes

NOTES: Open = search by typing in key words; prespecified = search by selecting key features of the assessments, such as grade level.

Comments from Potential Users About Information Needs

“Information about what this measure is designed to measure, what settings it is supposed to be used for, what user cases have been validated for, and if it’s validated, what is the technical information at the item level, if possible.”

“Not just list information about assessments, but also descriptions of how this tool is used for learning and developmental purposes in one or more schools. It’s important to show these assessments in the context of learning, rather than only for accountability.”

quarters of the assessments listed in this database. Users can also obtain instructions, from linked research articles, on how to score and interpret the results. The other four assessment repositories provide access to assessment documents for all measures listed.

Providing information about the quality and utility of assessments and guidance on how to interpret the quality of assessments are the main components of the evaluate function. There are three different approaches to providing quality information about assessments. The first approach draws on evidence from published journal articles or reports. The APA PsycTESTS database and ROMR use this approach. They provide summaries of the evidence of technical quality

or a link to related articles that provide such evidence. The second approach provides third-party reviews of measures, in addition to evidence available from published journal articles. The Buros Center for Testing and DREAM belong to this group. The third approach pre-screens assessments based on evidence of reliability and validity before including measures in the repository. However, information about the technical quality is not presented in the database; users need to search for such information on their own. The NIH Toolbox, the ARSCP, and the SCALE Performance Task Database fall into this category.¹

The reviewed websites provide examples, from research articles, of evidence about the utility of assessments. For instance, the NIH Toolbox also provides a bibliography of articles that have used measures included in the NIH Toolbox, along with summaries or full-text documents for the linked articles and presentations that used measures included in the NIH Toolbox.

In addition, example websites provide guidance to users on how to interpret evidence about the technical quality of assessments. For example, the ROMR provides guidelines to users on how to interpret the reliability and validity of measures. Such guidance is helpful for users without training in measurement to evaluate the quality of the measures they are interested in and select the ones that meet their needs.

The use function includes a variety of ways to support users to utilize assessment information and documents. For instance, the APA PsycTESTS database allows users to save and export their search results. The SCALE Performance Task Database provides teacher guides on how to implement a selected task. The NIH Toolbox provides multiple online training programs and various user manuals to help users understand how to use the measures it includes. The Buros Center for Testing provides links to the websites of the assessment providers, when

¹ It is difficult to evaluate the technical quality of the measures included in PALS because the information provided about their technical quality does not include commonly used reliability coefficients or evidence about validity.

available. CHILDES and Databrary also provide information about computer programs that can be used to analyze the video and audio files included in their databases.

One special type of use function is to support users in administering selected measures online. The NIH Toolbox is the only assessment repository that provides such assessment-administration functionality. The NIH Toolbox website provides a platform, called Assessment Center, to support users in administering their selected measures and collecting data online. An iPad app will also be available in 2015 to support the use of measures included in the NIH Toolbox. Such user-support services to administer measures and collect data may encourage use of measures included in the NIH Toolbox and help its team collect extensive data on its measures from different studies and use them for research purposes. In addition, certain services, such as technical assistance during the use of Assessment Center, will charge fees from users. On the one hand, this provides one source of revenue to sustain the repository in the long run. On the other hand, offering online administration of assessments requires more technological capacity and is likely to increase the cost and complexity of the repository.

The connection functions refer to functions to connect individual users with the larger user community. For example, the DREAM repository provides monthly newsletters to keep users informed of updates. DREAM and the NIH Toolbox website also connect with users through social media tools, such as Twitter. The Databrary repository also organizes professional research conferences. In addition, Skilledup allows its users to share their experience with a course by providing comments. Such functions allow users to join in the larger user community to share their experiences with using measures listed in the repository. This feature is not crucial for a well-functioning repository but could be helpful for facilitating interactions among users, which in turn could help promote improvements in the measures, as well as future development.

Allowing users to register and have a personal account on a repository might be beneficial for achieving multiple functions. For instance, this personal account can allow users to customize the information they receive, manage and use the information they requested, participate in discussion forum, provide comments to share their experience in using measures listed in the repository, and complete other functions provided by the repository system. Five of the assessment repositories we reviewed—the APA PsycTESTS database, the Buros Center for Testing, PALS, DREAM, and the NIH Toolbox—provide users customization functions. Users of the APA PsycTESTS database can save their search records and results for later use. Users of the Buros Center for Testing’s assessment review database will need to pay for access to third-party assessment reviews through their personal accounts. Users of the NIH Toolbox who are interested in using the Assessment Center to administer their selected measures and collect data online will need to achieve such purposes through a personal account.

Desired User-Interface Features

When reviewing the 12 websites, we identified a number of user-interface features that made the sites more or less effective, based on our experience. While any effective online database should have features like these, we mention them here because they seem particularly relevant to an assessment repository and the kinds of interactions that practitioners are likely to have with it.

First, websites that offer two or three search options can better accommodate the varied needs of potential users. For instance, the APA PsycTESTS database provides both a simple and

an advanced search interface. The simple version allows users to search for measures with simple key words. The advanced version offers more choices to conduct the search, such as searching by specific key words according to different features of an assessment (e.g., title, author, and year). Moreover, users can conduct further searches among results from the initial search using key words or by selecting characteristics of measures, such as the age group of assessment takers.

Second, it is helpful to display both the search criteria and the search results in two parts, vertically, on the same page. For example, the APA PsycTESTS database, ROMR, and Skilledup divide the screen into two vertical parts, with one panel providing detailed search categories for users to modify their searches and the other displaying the results (see Figure 3.1). In comparison, the PALS displays the search criteria and results on a single screen in a long continuous display, with the search categories on the top of the page. Thus, users have to scroll up to the top of the page to modify the search criteria.

Third, positive search results have to be displayed in some order, and it seems helpful to allow users to indicate their preferred order. For instance, Skilledup allows users to sort searched courses by the relevance of the course to the search criteria and by the ratings of courses. The Buros Center for Testing also allows users to sort searched results in multiple ways, such as alphabetically by assessment name.

Fourth, allowing users to compare multiple measures of potential interest will help users select the ones that best fit their needs. This means that users can select multiple measures they are interested in and have an overview of how these measures compare with each other on key dimensions—such as the purposes, target assessment population, item format, and time required to complete an assessment—on the same page. The websites of the Buros Center for

Figure 3.1
Screen Shot of the Search Interface of ROMR

Using the RAND Online Measure Repository

Measures can be browsed alphabetically or the list of 171 measures can be narrowed down using a **keyword search** and the seven criteria in the filter below.

Measures can be browsed by specific clinical **domains** such as depression, posttraumatic stress disorder or traumatic brain injury, **age group** of respondent (adult or child), whether there is a **fee for using the measure**, the **number of items**, the individual, clinical, or trained professional **who can administer** the measure, the measure **respondent** (e.g., unit leader, child, caregiver), and whether the measure has ever been **used with a military population**.

For more information please see the [User Guide for the Measure Repository](#) or the [Glossary of Terms](#).

If you have not selected any filter criteria all 171 measures will be listed alphabetically.

Showing 1 - 10 of 171

	COMPARE SELECTED
<p>Acute Stress Checklist for Children (ASC-Kids) The Acute Stress Checklist for Children (ASC-Kids) was designed as a self-report measure of acute stress reactions in children and adolescents who have recently experienced a traumatic event. Read More ></p>	<input type="checkbox"/> Compare
<p>Adolescent Cognitive Style Questionnaire (ACSQ) The Adolescent Cognitive Style Questionnaire (ACSQ) is a 12-item assessment of cognitive vulnerability to depression among adolescents. Read More ></p>	<input type="checkbox"/> Compare
<p>Adult Suicidal Ideation Questionnaire (ASIQ) The Adult Suicidal Ideation Questionnaire (ASIQ) is a 25-item, self-administered questionnaire used to assess the acuity and severity of suicidal ideation. Read More ></p>	<input type="checkbox"/> Compare

To narrow or broaden the search chose options from the lists below.

Keyword Search

- Domain

- PTSD
- Depression
- Anxiety
- TBI
- Suicidal thoughts

SOURCE: Screenshot of the “Search the RAND Online Measure Repository” section of the RAND Online Measure Repository.

Testing's assessment review database and ROMR allow users to compare multiple assessments of interest (see Figures 3.1 and 3.2). Skilledup also allows users to compare up to four courses. This feature is helpful for users to compare several possible options on key dimensions and identify the ones that better fit their needs. However, it might be necessary to set a limit on the number of measures that can be compared at once to ensure a clear review of the comparison results.

Moreover, it will be useful to provide a check box for each measure on the page where all search results are displayed, which would allow users to choose the measures that they want to compare. The Buros Center's website allows users to select measures for comparison only on the page where information about individual measures is displayed. Users have to click through several pages to select multiple measures for comparison. In comparison, ROMR and Skilledup allow users to select items for comparison on the page where all search results are displayed (see Figure 3.1). Users can compare lessons simply by clicking a check box associated with each item. This design feature reduces the amount of time required for the selection process. Overall, we think that the best search function will have these four features as part of its design.

Our review of the websites also revealed differences in how information is provided and some ways to make the display more effective. First, it is helpful to categorize information about measures into categories (such as purposes, assessment time, and cost) and present this

Figure 3.2
An Example of Comparisons of Multiple Measures on the RAND Online Measure Repository

RAND Online Measure Repository

	Assessment of Depression Inventory (ADI)	California Psychological Inventory-Depression (CPI-D)	Beck Hopelessness Scale (BHS)
General Information on the Measure			
Brief Summary of the purpose of the measure	The Assessment of Depression Inventory (ADI) is a 39-item measure of depression. It is a self-report assessment used to screen and assess progress during treatment for depression.	The California Psychological Inventory-Depression (CPI-D) was constructed retroactively from 33 items of a 480-item, longitudinal California Psychological Inventory (CPI) questionnaire, which measures depression and	The Beck Hopelessness Scale (BHS) is a 20-item, self-administered questionnaire used to measure and screen for hopelessness thinking patterns. More specifically, it seeks to identify and quantify patterns of negative expectations concerning one's self and one's future life.
	≈ Show Full Text	≈ Show Full Text	≈ Show Full Text
Measure Domain	Depression	Depression	Depression Suicidal Thoughts
Age Group	Adults	Adults	Adults
Military Population	Yes	Yes	Yes
Measure Administration			
Respondent	Individual	Individual	Unit Leader (Military only) Individual

SOURCE: Screenshot of the RAND Online Measure Repository, after choosing to compare the Assessment of Depression Inventory, the California Psychological Inventory-Depression, and the Beck Hopelessness Scale.

information in a consistent format so that users can easily locate the specific information they need. All example websites use a template to display information about items. We also used a template to collect and manage information about selected measures in the feasibility study. Figure 3.3 shows the template used for measures listed through ROMR.

Second, using multiple tabs to display different types of information about an assessment makes it easy for users to learn all aspects of an item without going to another web page. As described earlier, there are multiple categories of information that can be displayed about each measure. Some websites display descriptive information about measures and links to other documents, such as the assessment booklet and scoring rubric, on one page. Users will need to switch to another web page to review the linked documents and will have to click the “back” button or switch browser tabs to return to the main page and view other documents. The DREAM website uses tabs to display different pieces of information about a measure (e.g., “General Information” and “Outcomes”; see Figure 3.4). Users do not need to move between different web pages to review all information about a measure. Users can bring information from different tabs to the front by clicking on the tab title. This feature reduces the number of steps users need to take to understand all aspects of a measure.

Figure 3.3
The Template Used to Present Information About Measures Listed on ROMR

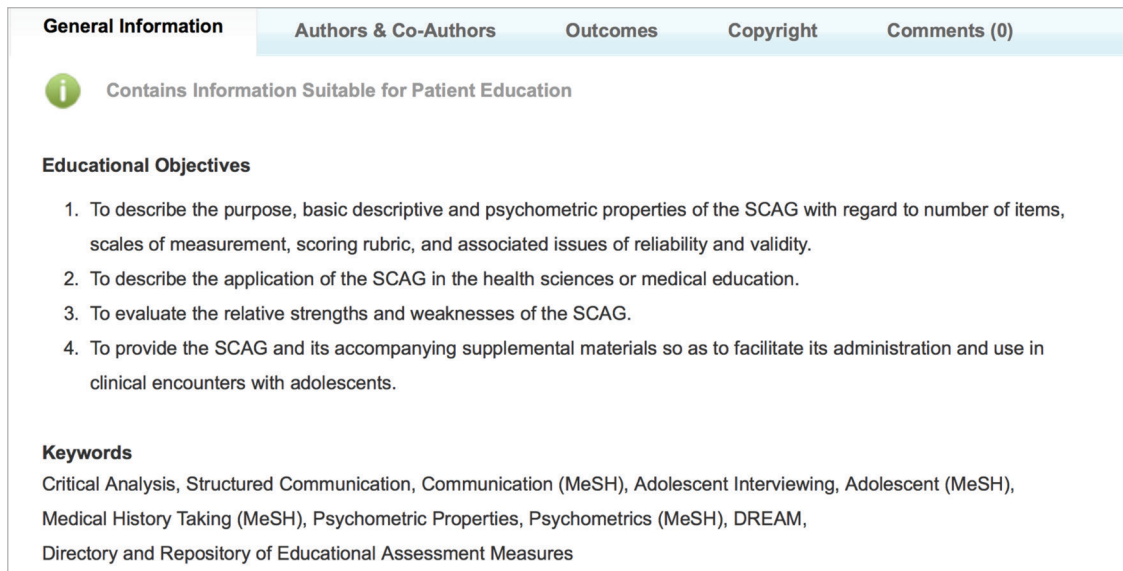
RAND Online Measure Repository

California Psychological Inventory-Depression (CPI-D)	
General Information on the Measure	
Brief Summary of the purpose of the measure	The California Psychological Inventory-Depression (CPI-D) was constructed retroactively from 33 items of a 480-item, longitudinal California Psychological Inventory (CPI) questionnaire, which measures depression and anxiety across a population.
Measure Domain	Depression
Age Group	Adults
Military Population	Yes
Measure Administration	
Respondent	Individual
Who Can Administer the Measure?	Self-administered
Method of Administration	Survey questionnaire (pen/paper)
Scoring the Measure	
Number of Items	33
Measure Subscales	No information available in the references reviewed.
Response options and anchors for each scale that comprises the measure	Consistent with CPI questions, the response options for these 33 items are TRUE-FALSE, indicating the presence or absence of each symptom, thought, or behavior identified in the item. Eight of the 33 items included in the Depression scale are reverse - scored.
Reliability and Validity of the Measure	
Sample minimum	244
Sample maximum	1044

SOURCE: Screenshot of the search result for the California Psychological Inventory-Depression, from the RAND Online Measure Repository.

RAND RR1204-3.3

Figure 3.4
Use of Tabs to Display Information About a Measure on the DREAM Website



SOURCE: MedEdPortal's Dream website, as of September 2015.

NOTE: The tabs included in this figure are "General Information," "Authors & Co-Authors," "Outcomes," "Copyright," and "Comments." SCAG = Structured Communication Adolescent Guide; MeSH = Medical Subject Headings.

RAND RR1204-3.4

For the overall design of the website, we identified two user-interface features worth considering while designing the repository. First, it is helpful to provide tabs to the major sections of the website on every page of the repository so that users can easily redirect to other sections from any page. For instance, the MIT OpenCourseWare website has six tabs on every page, including "Home," "Courses," "About," "Donate," "Featured Sites," and a search box. Users can easily jump to any other sections when they need to.

Second, it is helpful to reduce the amount of information processing for users by using icons, shortcuts, and so on. For instance, PALS uses graphic icons to represent a common set of documents associated with each measure. Skilledup uses graphic icons and brief category names for the search function to reduce the amount of text processing for users. However, excessive use of icons might increase the cost of constructing the website and slow down the processing speed. Developers of the repository should consider both the pros and cons of using icons when designing the repository.

Summary

We reviewed 12 repository websites and identified a number of useful functional modules and user-interface features that should be incorporated into an assessment repository. Together, these elements can help users identify measures they are interested in, obtain information about measures they need, support their use of measures they select, and connect to the larger community to develop hard-to-measure competencies among K–12 students.

Findings and Recommendations

The primary purpose of this study was to assess the feasibility of building a repository of assessments of hard-to-measure competencies that would be useful to practitioners, researchers, and policymakers. We investigated this issue in a variety of ways, including talking with experts, exploring existing online repositories, and developing a small-scale database of measures that we identified and reviewed. In this process, we learned a great deal about the challenges associated with building such a repository and about effective ways to approach the task, and we identified a few important questions we were unable to answer. In this chapter, we summarize our findings about feasibility and offer some practical recommendations about developing and maintaining such a repository.

The most general finding from our study is that it is feasible to build a repository of assessments of hard-to-measure skills. With some limitations and considerations, the necessary conditions for a repository exist: Many measures have been developed, it is usually possible to find descriptive information about them, it is often possible to obtain copies that can be reviewed, and it is sometimes possible to find evidence about their technical quality and other information that one would want to provide to prospective users. Moreover, it is possible to organize the information so that potential users can find out what they want to know, and it is possible to create a website whose architecture facilitates access, review, and selection. Furthermore, none of the existing repositories easily fills the role.

On the other hand, we identified a number of considerations associated with building such a repository. In the following paragraphs, we discuss a range of theoretical and practical considerations and offer our recommendations about how to address them.

1. Goals and purposes. Decisions about which measures to include in the repository and what information to collect and present for each measure are impossible to make without a clear set of goals and purposes. Thus, developers should create a statement that specifies the types of users and user needs the repository aims to satisfy and should include both short- and long-term goals to guide the rest of the project work. Some key questions that need to be answered in the statement are: Which users are the repository designed to serve—teachers, administrators, researchers, or policymakers? Which uses should it support—formative feedback for daily instructional improvement, summative assessment of program impact, periodic profiling of student growth? Is the repository designed to be comprehensive and contain all available measures of all relevant constructs, or is it focused in some manner—for example, measures that are most relevant to student outcomes, measures that meet a threshold in terms of technical quality, or measures of particular constructs, such as “learning how to learn”?

2. Expertise. Key planning decisions—both conceptual and technical—require a range of expertise that is difficult to find in one individual or one institution. To make key decisions

about the repository, including setting goals, selecting constructs of interest, finalizing the definition of constructs and associated names, and developing criteria for selecting measures, it would be important to consult with researchers, educators, psychometricians, and subject-matter experts. One way to ensure adequate expertise is to convene a technical advisory committee that includes researchers familiar with educational and psychological measurement, educators, and policymakers.

It is also necessary to have multiple teams of experts for different project tasks and to make sure that they work closely with each other. Specifically, the substantive team that collects measures for the repository should include researchers familiar with educational and psychological measurement and with research on hard-to-measure competencies. IT professionals with expertise in data architecture and web design should develop the infrastructure system and the user interface of the repository. These teams need to work to ensure that the goals of the repository, the assessments collected for the repository, and the function and designs of the repository are systematically aligned with each other, through a well-functioned and user-friendly website, to provide the information that potential users need.

3. Searching. Developers should set up rules for how to search for measures, and their associated information, to add to the repository, as well as guidelines about what to do when information about the selection criteria is unavailable. Searching for measures electronically can be effective, but it is not sufficient. Standard electronic-search approaches have limitations when it comes to noncommercial assessments and newly developed assessments. Thus, the rules for searching should include the use of multiple searching strategies and sources.

The rules for searching should be followed consistently but should also be modified over time, as appropriate. In particular, new search strategies and sources may become applicable after deciding what measures should be included in the repository. For instance, we decided not to include measurement components in technology-based learning programs. If the plan for the larger study to develop the repository includes these measurement programs, it would be necessary to also search in journals about technology-based learning to find such programs and associated measures.

Given that the process to conduct searches and collect, review, and evaluate results can be time-consuming, it is necessary to budget adequate personnel and time for these tasks. Time budgeted should include both the time required to search for measures, review searched information, and organize information for the repository and the time needed to review literature about similar constructs, develop operational definitions for constructs of interest, and identify terms associated with each construct of interest.

4. Selection. Developers should set up criteria for which measures should be selected for the repository. The goal statement of the repository and the intended users of the repository will guide the development of measure-selection criteria. The dimensions on which potential measures vary from each other should be considered when developing these criteria. In addition, it is also necessary to specify how decisions should be made when the key dimensions of an assessment are unknown (e.g., there is a lack of technical information on validity).

We also recommend that the repository developers err on the side of being more inclusive than less inclusive. In particular, include all the measures that were found that assess the constructs of interest; do not select measures based on the availability of technical-quality evidence or complete assessment documents. This will ensure that there will be a diversified pool of measures to be included in the repository. Users will have the opportunity to learn all possible options that they have in order to measure their constructs of interest. However, this

may also contribute to future research and development on the assessment of hard-to-measure competencies.

If the decision is to include measures regardless of the availability of technical-quality evidence and the assessment documents, the developer of the repository should also provide some guidance to repository users on how to interpret evidence about the technical quality of measures (such as the guidance that ROMR provides) and how to select measures based on a variety of factors, such as the purposes of an assessment, the technical quality of measures, and other logistical concerns (Soland, Hamilton, & Stecher, 2013).

5. Common features of measures. It is important to set up a template that specifies what information should be displayed in the repository for each measure, but it is rarely the case that all the desired information can be found. For instance, it is hard to find and to gain access to information about technical quality and, in some cases, to provide potential users with access to assessment documents, especially for commercial assessments. It is important to recognize that it will not be possible to obtain the technical details that people might want, especially technical evidence about recently developed measures and documents for commercial assessments.

However, developers should include measures that have enough information for users to decide whether they are relevant, even if some key information is missing. Some dimensions that will be important for users to understand whether a measure is appropriate for their intended uses are the purposes of a measure, the constructs assessed, the target assessment population, the time and format required for assessment administration, the cost, and how to score the results and receive final reports.

Moreover, some alternative solutions may be adopted to alleviate the inconvenience resulting from the gaps in the information provided for measures. For instance, links to the assessment provider's website or contact information of the developer will help potential users obtain the measures they are interested in, directly from the provider or developer. The technical team that will maintain the repository can use change detection and notification tools for online content, such as Google Alert, to track new findings about measures that are selected for the repository and update the repository when new technical evidence is available.

6. Maintaining the repository. Once the initial repository is operational, effort will be needed to maintain it. Maintenance involves refreshing the database with new information about measures that are already in the repository, such as new versions of existing measures. Maintenance also involves adding measures that have been developed, especially those that may appear because of new developments in technology, such as the use of artificial-intelligence technology to develop game-based assessments (Lester et al., 2013). To maintain and improve the technological aspect of the repository website, developers should conduct regular technological reviews, such as checking whether links provided are still active and accurate and updating the technical functionality of the online portal and interface. In addition, it may also be helpful to regularly review other repository websites for new available functions that have resulted from innovations in technology. Developers should also scan the landscape of hard-to-measure competency assessments to identify significant changes that warrant updates to the repository.

7. Funding the repository. In terms of the funding to support the maintenance of the repository, multiple options should be considered. One option is to draw on continuing support from one or multiple foundations. Another option is to provide fee-based services to users to make the website self-sustainable in the long run. While we do not think that users will continue to pay a fee for access to the information contained in the repository once they (or

a colleague) have identified measures, it seems possible that online test administration could be fee-based. For instance, users might need support to administer and score noncommercial assessments. It is possible to provide such additional services using technology-based platforms, such as the Assessment Center of the NIH Toolbox, and charge fees from users to support the operation of the repository. A third option is a combination of both options. For example, funding from foundations may play a major role in supporting the operation of the repository at the beginning, when the repository is growing its user group. However, as more users use the fee-based services, the repository may become self-sustainable and rely less on support from foundations.

To attract more users to this repository and make it self-sustainable, efforts will need to be made to market the repository to the right users and ensure that the services provided meet their needs. Disseminating information about the existence and purposes of this repository will also help keep the repository up-to-date. Researchers or other developers whose instruments are not included or are included in an outdated form might not be aware of that fact until they have had a chance to review the repository, so widespread dissemination can serve as a mechanism to promote frequent updating.

Information about this repository should be disseminated to potential users through multiple channels. For instance, information about this repository should be disseminated through major news sources (e.g., *Education Week*) and conferences for practitioners, researchers, and policymakers (e.g., the annual conference of the American Educational Research Association)—as well as through user groups of networks with a similar mission to promote the development of hard-to-measure competencies. Allowing website visitors to subscribe to email newsletters about updates to the repository might also help engage existing users. However, the most powerful force to attract users, engage them, and make the website self-sustainable in the long run will be primarily determined by the usefulness of the repository website and the services it provides to potential users.

Together, these considerations and the steps described in Chapter Two suggest an overall procedure to build and maintain this repository, as shown in Figure 4.1. Some steps might have a recursive, instead of a linear, relationship. Repetition of steps 3–9 might be required when building and updating the repository. This procedure map serves as a guide for the next step toward building a working repository. The endeavor to develop this repository will require close collaboration among funders, developers, and potential users. Once built, the repository should provide a good source for practitioners, researchers, and policymakers to promote the development and assessment of hard-to-measure competencies.

Figure 4.1
An Overview of the Procedure to Build and Maintain the Repository and Major Considerations at Each Step

1. Define the goals and purposes of the repository	<ul style="list-style-type: none"> • Identify users and user needs • Identify major functions of the repository website
2. Identify the right expertise	<ul style="list-style-type: none"> • Have an expert committee provide guidance on key decisions for the repository
3. Identify constructs of interest	<ul style="list-style-type: none"> • Develop an operational definition and identify associated terms for each construct selected
4. Develop measure-selection criteria	<ul style="list-style-type: none"> • Specify actions when information about selection criteria is unavailable
5. Search for measures of constructs of interest	<ul style="list-style-type: none"> • Use multiple searching strategies and sources
6. Select measures for the repository	<ul style="list-style-type: none"> • Collect information about measures involving measure-selection criteria and record the decisionmaking process
7. Obtain information and materials for the measures selected	<ul style="list-style-type: none"> • Include measures without adequate evidence about their technical quality but provide guidance to users about how to select measures based on available information
8. Organize information and materials collected	<ul style="list-style-type: none"> • Use a template to organize information and materials collected for measures
9. Merge the assessment database and web software	<ul style="list-style-type: none"> • Have the substantive team and the technology team work together to jointly design the assessment database and the user interface of the repository website
10. Disseminate information about the repository to potential users	<ul style="list-style-type: none"> • Disseminate through multiple channels to reach a wide range of potential users
11. Update the repository	<ul style="list-style-type: none"> • Update the information and materials about measures included and add new measures • Conduct regular technological reviews of the website

NOTE: Some steps in the measures for searching and selecting might have a recursive relationship and might require the repetition of steps 3–9.

The Database of Measures Developed in This Feasibility Study

Table A.1 shows the database of measures we developed in this feasibility study. More details about each measure, including links to the developer or provider and related citations, are provided in the measure profile we developed for each assessment (see Appendix B).

Table A.1
Main Features of Measures Selected for the Repository in the Feasibility Study

Measure Name	Category of Construct(s) Measured	Construct(s) Measured ^a	Grade Level(s) of Target Assessment Takers	Format	Language	Assessment Administration Format	Is a Fee Required to Use This Measure? ^b
The Academic Motivation Scale	Intrapersonal	Motivation	Elementary school; middle school; high school	Likert-scale item	English	Paper and pencil	No
Bar-On Emotional Quotient Inventory	Interpersonal	Emotional intelligence	Middle school; high school	Likert-scale items	English	Paper and pencil; computer	Yes
Behavior Assessment System for Children	Interpersonal; intrapersonal	Communication; collaboration; emotional intelligence; self-control; resilience	Preschool; elementary school; middle school; high school	Likert-scale items	English; Spanish	Paper and pencil; computer	Yes
California Critical Thinking Disposition Inventory	Cognitive	Critical thinking	Elementary school; middle school; high school	Multiple-choice items	English; Spanish; others	Paper and pencil; computer	Yes
California Measure of Mental Motivation	Cognitive	Critical thinking	Preschool; elementary school; middle school; high school	Multiple-choice items	English; Spanish; others	Paper and pencil; computer	Yes
Center for Clinical and Translational Science and Training (CCTST) M-Series for Children and Youth	Cognitive	Critical thinking	Elementary school; middle school	Multiple-choice items	English; Spanish	Paper and pencil; computer	Yes
Challenging Situations Task	Interpersonal	Communication; collaboration	Preschool	Multiple-choice items	English	Paper and pencil	No
Children's Self-Efficacy for Peer Interaction Scale	Interpersonal	Communication; collaboration	Elementary school	Likert-scale items	English	Paper and pencil	No
Clifton Strengths Explorer	Cognitive; interpersonal; intrapersonal	Critical thinking; communication; emotional intelligence; motivation; grit	Middle school; high school	Likert-scale items	English	Computer	Yes
Cornell Critical Thinking Test	Cognitive	Critical thinking	High school	Multiple-choice items	English	Paper and pencil	Yes

Table A.1—Continued

Measure Name	Category of Construct(s) Measured	Construct(s) Measured ^a	Grade Level(s) of Target Assessment Takers	Format	Language	Assessment Administration Format	Is a Fee Required to Use This Measure? ^b
Creative Self-Efficacy Scale	Intrapersonal	Self-efficacy	Middle school; high school	Likert-scale items	English	Paper and pencil; computer	No
Creativity Assessment Packet	Cognitive	Creativity	Elementary school; middle school; high school	Drawing; Likert-scale items	English	Paper and pencil	Yes
CWRA+	Cognitive; interpersonal; intrapersonal	Critical thinking; written communication	High school	Performance-assessment task; multiple-choice items	English	Online administration	Yes
Friendship Quality Questionnaire	Interpersonal	Communication; collaboration	Elementary school	Likert-scale items	English	Computer	No
Global Empathy Scale	Interpersonal	Global awareness	High school	Likert-scale items	English	Paper and pencil; Computer	No
Graduation Portfolio System	Cognitive; interpersonal; intrapersonal	Academic learning; critical thinking; communication; collaboration; global awareness	Preschool; elementary school; middle school; high school	Performance-assessment frameworks and rubrics	English	Paper and pencil	Yes
Grit Scale	Intrapersonal	Grit	Elementary school; middle school; high school	Likert-scale items	English; Chinese; French; German	Paper and pencil	No
Homework Management Scale	Intrapersonal	Self-regulated learning	Middle school; high school	Likert-scale items	English	Paper and pencil	No
Implicit Theories of Intelligence Scale for Children	Intrapersonal	Mind-set	Elementary school; middle school; high school	Likert-scale items	English	Paper and pencil	No
KIPP Character Report Card	Interpersonal; Intrapersonal	Communication; emotional intelligence; grit; mind-set; self-regulated learning	Elementary school; middle school; high school	Likert-scale items	English	Paper and pencil	Yes

Table A.1—Continued

Measure Name	Category of Construct(s) Measured	Construct(s) Measured ^a	Grade Level(s) of Target Assessment Takers	Format	Language	Assessment Administration Format	Is a Fee Required to Use This Measure? ^b
Mayer-Salovey-Caruso Emotional Intelligence Test	Interpersonal	Emotional intelligence	High school	Multiple-choice items	English; Spanish; Others	Paper and pencil; computer	Yes
Moral Competence Test	Cognitive; interpersonal	Critical thinking; collaboration	Elementary school; middle school; high school	Likert-scale items	English	Paper and pencil; computer	Yes
Patterns of Adaptive Learning	Interpersonal ^a	Goal orientation; self-efficacy	Elementary school; middle school; high school	Rating-scale items	English	Paper and pencil	No
Self-Regulated Strategies for School Writing Tasks	Intrapersonal	Self-regulated learning	High school	Rating-scale items	English; Portuguese	Paper and pencil	No
Social Competence and Behavior Evaluation	Interpersonal	Communication; emotional intelligence	Preschool; elementary school	Likert-scale items	English	Paper and pencil	No
Social Skills Improvement System	Interpersonal; intrapersonal	Communication; emotional intelligence; self-control	Elementary school; middle school; high school	Likert-scale items	English; Spanish	Paper and pencil; computer	Yes
SCALE Performance Task Database	Cognitive	Academic learning; critical thinking	Preschool; elementary school; middle school; high school	Performance-assessment task	English	Paper and pencil	Yes
Test of Everyday Reasoning	Cognitive	Critical thinking	High school	Multiple-choice items	English; Spanish	Paper and pencil; computer	Yes
Thinking Creatively in Action and Movement	Cognitive	Creativity	Elementary school; middle school; High school	Constructed-response items	English	Paper and pencil	Yes
Time Management Scale for Middle School Students	Intrapersonal	Time management	Middle school	Likert-scale items	English	Paper and pencil	No

Table A.1—Continued

Measure Name	Category of Construct(s) Measured	Construct(s) Measured ^a	Grade Level(s) of Target Assessment Takers	Format	Language	Assessment Administration Format	Is a Fee Required to Use This Measure? ^b
Torrance Test of Creative Thinking	Cognitive	Creativity	Elementary school; middle school; high school	Constructed-response items	English	Paper and pencil	Yes
VIA Youth Survey	Cognitive; interpersonal; intrapersonal	Creativity; communication; collaboration; leadership; emotional intelligence; grit; self-regulated learning ^b	Elementary school; middle school; high school	Likert-scale items	English	Computer	No
Virtual Environment for Social Information Processing Assessment Tool for Upper Elementary and Middle School Children	Interpersonal	Social-information processing	Elementary school; middle school	Performance-assessment task in virtual environment	English	Computer	Unclear ^c
Work Readiness Credential	Cognitive; interpersonal; intrapersonal	Academic learning; critical thinking; collaboration; communication; self-regulated learning	High school	Multiple-choice items	English	Computer	Yes

^a Constructs measured are assigned based on descriptions of measures available from the developer or provider.

^b No = measures can be used for research and educational purposes for free.

^c This measure is still under development. Information about whether fee is required for use is unavailable.

Measure Profile for California Critical Thinking Disposition Inventory

In this appendix, we show an example of the profile we prepared for each measure selected in this project. This measure profile is for CCTDI.

What Does It Intend to Measure?

CCTDI is part of the California Critical Thinking Skills Test Family. It is intended to measure the dispositional aspect of critical thinking. According to the publisher, people with high scores on CCTDI have a strong desire to apply critical-thinking skills to solving problems in real life.

CCTDI contains seven scales about an individual's willingness to think critically: the disposition toward truth seeking or bias (i.e., Truthseeking), toward open-mindedness or intolerance (Open-Mindedness), toward anticipating possible consequences or being heedless of them (Analyticity), toward proceeding in a systematic or unsystematic way (Systematicity), toward being confident in the powers of reasoning or mistrustful of thinking (Confidence in Reasoning), toward being inquisitive or resistant to learning (Inquisitiveness), and toward mature and nuanced judgment or toward rigid simplistic thinking (Maturity of Judgment).

Whom Is It Intended For?

CCTDI is appropriate for a wide range of assessment takers, including secondary and postsecondary school students, working professionals, and the general public.

What Type of Items Does the Measure Have?

CCTDI items ask respondents to rate how much they agree or disagree with statements about opinions, beliefs, values, and expectations related to the use of critical thinking in decisionmaking.

How Is It Administered?

Respondents can complete CCTDI in a paper-and-pencil mode or through a computer-based testing interface developed by Insight Assessment.

How Long Does It Take to Complete This Assessment?

The assessment takes 20–30 minutes to complete.

How Is This Measure Scored?

Insight Assessment provides scoring services.

What Do the Scoring Results Include?

Insight Assessment produces the final report, which provides an overall score of critical-thinking disposition and scores on all the subscales. The report also provides a categorical interpretation of the strength of the overall and subscale scores included in the report.

Insight Assessment also provides additional analysis for group assessment takers, such as descriptive statistics of the demographic characteristics of assessment takers (if collected) and their scores on CCTDI. Electronic data files with all assessment takers' demographic information and assessment scores are also available to group assessment takers. The Insight Assessment website (www.insightassessment.com) provides more information about the details of the score report.

How Much Does It Cost to Purchase, Administer, and Score This Measure?

Contact Insight Assessment for more details.

What Evidence Is Available About the Reliability and Validity of This Measure?

Reliability

- Internal consistency reliability: According to Insight Assessment, CCTDI has been used to assess individuals' willingness to apply their critical-thinking skills in a variety of settings, including elementary and secondary schools, colleges and universities, and working professionals. The internal consistency reliability coefficients ranged from 0.8 to 0.98.
- Test-retest reliability: Insight Assessment also reported that the test-retest reliability coefficient averaged 0.86. However, Insight Assessment's website does not provide detailed information about the study in which the reliability of CCTDI was examined. Therefore, information about the sample of assessment takers and the length of time between pre- and postassessment was unavailable.

Validity

- Evidence based on relations to other variables: Details of evidence about the validity of CCTDI are not publicly available from Insight Assessment. Interested users may purchase the assessment manual from Insight Assessment. One study showed that scores on CCTDI had a strong correlation, of 0.66, with scores on the California Critical Thinking Skills Test (Facione, Facione, & Sanchez, 1994). However, the samples used in this study were nurses. It is uncertain whether the results can be generalized to K–12 students.

The Insight Assessment website provides more details about the reliability and validity evidence for CCTDI.

Whom to Contact to Obtain This Measure?

Contact Insight Assessment to obtain the measure.

Source: Insight Assessment.

Reference: N. Facione, P. Facione, & C. Sanchez. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, 33(8), 345–350.

Descriptions of Example Websites Reviewed

This appendix provides detailed information about the content and functions of the websites of 12 example repositories we reviewed in this study.

APA PsycTESTS Database

APA developed and maintains the APA PsycTESTS database. This database contains more than 20,000 assessments that have been used in psychological research. More than three-quarters of the assessments included are assessments developed by researchers but not made commercially available. Users can search for assessments using simple key words or multiple key words on several dimensions, such as assessment title, author name, format, and assessment-taker population. They can also browse by key categories, such as constructs measured and year developed. The database provides descriptive and evaluative information about assessments and links to associated research articles or the developer's website for assessments listed. It also provides actual assessment documents for noncommercial assessments, when such documents are available. Users can use noncommercial assessments for educational and research purposes for free. Users need to contact the developers if they want to use noncommercial assessments for other purposes. Users can find more information about how to obtain commercial assessments from the links to the developers' websites. Users can register personal accounts and save their search results into their accounts. They can also email, print, and export their search results for other uses. Users have to subscribe to APA's databases and pay a membership fee to access and use this database.

More information about this database is available at <http://www.apa.org/pubs/databases/psyctests/index.aspx>

Buros Center for Testing

The Buros Center for Testing at the University of Nebraska, Lincoln, has an assessment review database¹ that provides reviews of more than 3,500 commercially available assessments. Users can search assessments based on key words or browse by assessment name and the categories of the assessed constructs. This database provides descriptive information and third-party reviews

¹ The Buros Center for Testing also publishes *Mental Measurement Yearbook* and *Tests in Print*. Both provide information about commercially available, English assessments. However, these two publications are available only in hard copy. It is likely that they draw on the information in the assessment review database.

of assessments. Users can view the descriptive information for free, but they have to pay to review the third-party reviews of the assessments. The database does not provide the actual assessment documents. It provides links to the provider of each assessment, if that information is available. Users can contact the developer if they are interested in any particular assessment. The website also allows users to select and compare multiple measures. However, the dimensions on which measures are compared are limited to author, publisher, purpose, and publication date. Users will need to register a personal account to pay for third-party reviews. Users can subscribe to a rich site summary (RSS) to obtain updates in the database.

More information about this database is available at <https://marketplace.unl.edu/buros/>

The RAND Online Measure Repository (ROMR)

ROMR is an online searchable database that contains measures related to psychological health and traumatic brain injury (TBI). It was developed to support the monitoring and evaluation of programs designed to address military personnel's psychological health and TBI. Currently, ROMR hosts 171 measures. Users can locate measures by searching with key words or filters on key dimensions, such as constructs assessed, age group of assessment takers, and whether a fee will be charged for use. ROMR uses a template to display information about all measures included in its database. This template covers information about the purpose of the measure, constructs assessed, targeted assessment takers, assessment items, how to administer a measure selected, and technical evidence available from published journal articles or reports. ROMR allows users to select and compare multiple measures on all dimensions in the template. ROMR provides information about how to obtain the actual assessment materials. It is the responsibility of interested users to obtain such materials if needed. Access to ROMR is free to the public.

More information about this database is available at <http://www.rand.org/multi/military/innovative-practices/measure.html>

Performance Assessment and Links in Science (PALS)

PALS is a repository of performance-assessment tasks for science. This repository was developed and hosted by SRI International. The National Science Foundation provided funding for this project. The PALS repository linked its listed tasks for science-performance assessment to the National Science Education Standards and other standards frameworks, such as the state assessment standards used in Illinois and Texas. Users locate tasks by clicking the standards or subcomponents of standards, grade levels, and content areas. For each task listed, PALS provides descriptive and evaluative information about the task, assessment documents, and examples of students' work. The information provides a brief description of what students do during the task, expectations for their performance, and administration procedures. Assessment documents include actual task descriptions, with instructions to students and scoring rubrics. Evaluative information provides brief descriptions about the development process, the number and representativeness of students who had taken a task, and results from reliability and validity analyses. However, most tasks listed in the database do not have actual reliability

and validity analyses results, although every task in PALS has a linked web page, “Technical Quality Information.”

More information about this database is available at <http://pals.sri.com>

Stanford Center for Assessment, Learning and Equity (SCALE) Performance Task Database

The SCALE Performance Task Database includes two groups of performance-assessment tasks. The first group includes tasks developed for the Innovation Lab Network (ILN). The other group includes tasks that SCALE collects from online sources. Both components include performance assessment tasks in English language arts (ELA), mathematics, science, and social studies, at the K–12 level. Tasks are linked to the Common Core State Standards (CCSS) in ELA and mathematics; the Next Generation Science Standards; and the College, Career, and Civic Life (C3) Framework for Social Studies State Standards. All tasks are categorized on key characteristics, such as source (i.e., developer), grade level, subject, and course. Users locate tasks of interest by selecting the characteristics of tasks (e.g., ELA tasks for grade 10) or by providing key words. For each task, descriptive information and assessment documents are provided. Assessment documents include instructions for students, teacher guides, and scoring rubrics. The database developer also claims to have tested these tasks against “rigorous” quality criteria to ensure the technical quality of tasks.

More information about this database is available at <http://performanceassessment.stanford.edu>

Assessment Repository of the Society of Clinical Psychology (ARSCP)

The ARSCP database is developed and maintained by the Society of Clinical Psychology. The database includes assessment tools that can be used by clinical and academic psychologists. These assessments need to have published psychometric evidence of reliability and validity. Moreover, both the assessment tool and the technical evidence are publically available. ARSCP lists all measures on one web page. Users can browse the list to identify assessments of interest. For each assessment, ARSCP indicates the age group(s) of applicable assessment users (i.e., child, adolescent, and adult) and the construct(s) assessed and provides a link to an article that presents the technical evidence for this assessment. Users can provide feedback to ARSCP via email.

More information about this database is available at <https://www.div12.org/assessment-repository/>

Directory and Repository of Educational Assessment Measures (DREAM)

DREAM contains assessment instruments used in health science education. This database is developed and maintained jointly by MedEdPORTAL Publications and Georgia Regents University. Users can search for measures of interest using key words or selecting filters on dimensions, such as assessed constructs and assessment methods. For each measure listed in the data-

base, DREAM provides descriptive and evaluative information, as well as actual assessment documents. Descriptive information includes the constructs measured, the format of assessment items, the scores produced, and a reference to the article that described the technical evidence of this measure. Users can obtain evaluative information about the measure from the linked article. They can also obtain a peer-reviewed analysis of the measure from DREAM. Assessment documents include the assessment itself and available supplementary materials for assessment administration. Users need to register for an account of the Association of American Medical Colleges to access the materials. Registered users can share their experience with other users about measures they have used in their research. Website visitors can subscribe to email newsletters to receive updates. Users can also remain connected with this database through Twitter.

More information about this database is available at <https://www.mededportal.org/about/initiatives/dream/>

The National Institutes of Health (NIH) Toolbox

The NIH Toolbox contains 42 measures of cognition, emotion, motor skills, and sensation for ages 3 to 85. These measures were selected by a group of experts based on the available technical evidence. Users locate the measures of interest by selecting the constructs and subcomponents of the constructs. Descriptive information and actual assessment documents are provided for measures. Descriptive information includes what an assessment measures and the appropriate age group for respondents. Although explicit evaluative information is not provided for measures in the NIH Toolbox, users can find technical-quality evidence about each measure based on the information in the NIH Toolbox's technical manual. Assessment documents include actual assessment booklets, administration instructions, and scoring and technical manuals.

Users can administer these measures using assessment documents from the website. The NIH Toolbox provides various materials to help users utilize measures included in its database, including online presentations, demonstration videos, and training manuals.

Users can also administer these measures using online support from the NIH Toolbox website. They can use an online administration platform, Assessment Center, to administer the measures they select. An iPad app is being developed and is estimated to be available in 2015 for data collection. Users need to register an account on Assessment Center to use it. Technical support is available for a fee to administer assessment tools through Assessment Center and resolve other problems when using measures included in the NIH Toolbox.

NIH Toolbox also has several features to keep users engaged. It keeps a running list of researchers, research projects, publications, and conference presentations that used measures included in the NIH Toolbox. It invites visitors to follow it on Twitter and give feedback about the website.

More information about this database is available at <http://www.nihtoolbox.org/Pages/default.aspx>

Child Language Data Exchange System (CHILDES)

CHILDES is a central repository of speech audio, video, and text transcript files for first-language acquisition. The website presents information in ten functional tabs. Interactions with the website are relatively straightforward. Users can join a membership list by providing their email addresses. The website provides three categories of information. The first category is about the CHILDES repository, instructions for potential users and contributors, and contact information. The second category is the audio, video, and text transcript files that users can download from the repository. These files are grouped into multiple folders, by language and semantic topics, such as bilingual. Users locate files of interest by browsing through multiple folders. The third category is about programs that users can utilize to analyze speech files, links to other relevant websites, and instructions about how to use files from the repository for teaching purposes.

More information about this database is available at <http://childes.psy.cmu.edu>

Databrary

Databrary is an open-data library for storing, managing, preserving, analyzing, and sharing video and other data files for developmental science. Videos collected in this database come from various developmental studies about different age levels, ranging from infants to adults. This repository serves as an open source for researchers to learn how to collect, manage, and analyze video data for developmental research.

Videos and associated files are organized by projects. For each project, the website provides a brief description, covering the purpose, method, results, conclusion, and the reference. Users can search for videos using simple key words or browsing the catalog. This website also provides a user guide, online training, and in-person training workshops at professional conferences. A sister repository, Datavyu, provides support on how to code and visualize video data. Users need to register at this website to access files in the repository.

Visitors of this website can subscribe to newsletters to receive updates about the repository. Users can also follow Databrary on popular social-media programs, such as Facebook and Twitter.

More details about this database are available at <https://nyu.databrary.org>

Skilledup

Skilledup is a website that provides information about online courses. Courses are separated into four categories: creative, technology, general education, and business. Users can access courses by browsing within each category or searching by features of the courses, such as the format. Users can also compare up to four courses. For each selected course, the website provides a simple description, information about the provider, and four related courses.

This website has several helpful user-interface features. First, the search interface is simple and easy to use. Users see a large search box in the middle of the screen when they log in to the website. Second, the website uses graphic icons to categorize information and simple phrases to minimize the workload for text processing. Third, this website uses a large font size for key

information, such as course title. This makes it easy to identify the location of critical information. Fourth, the website provide quick links to its major functional modules at the same location on every page. This feature makes it easier for users to jump from any page to the function they need.

Users need to register and log in to access courses they are interested in. Visitors of this website can subscribe to email updates. They can also obtain updates from the website through popular social media programs, including Facebook, Twitter, Google+, and LinkedIn.

More information about this database is available at <http://www.skilledup.com>

MIT OpenCourseWare

MIT provides anyone connected to the Internet with materials for more than 2,200 courses through this open-course website. Users can search for course by department, topic, and MIT course number, as well as by key words. For each course selected, the course home page uses multiple tabs to provide access to all the information about the course, including a brief summary, syllabus, calendar, readings, assignments, and exams. Users can download all course materials with one click.

Users can access all the materials on this website for free. The website receives support from multiple foundations and private supporters. It also calls for donations. Visitors of the MIT OpenCourseWare website can receive updates from the website by subscribing its newsletters or following it on popular social media programs, such as Google+ and Twitter.

More information about this database is available at <http://ocw.mit.edu>

Abbreviations

APA	American Psychological Association
ARSCP	Assessment Repository of the Society of Clinical Psychology
CCTDI	California Critical Thinking Disposition Inventory
CHILDES	Child Language Data Exchange System
DREAM	Directory and Repository of Educational Assessment Measures
ETS	Educational Testing Service
IT	information technology
MIT	Massachusetts Institute of Technology
NIH	National Institutes of Health
PALS	Performance Assessment Links in Science
ROMR	RAND Online Measure Repository
SCALE	Stanford Center for Assessment, Learning and Equity

References

- Apple. (2008). *Apple classroom of tomorrow—today: Learning in the 21st century*. Cupertino, CA: Author. Retrieved from http://ali.apple.com/acot2/global/files/ACOT2_Background.pdf
- Boyatzis, R. E., Goleman, D., & Rhee, K. S. (2000). Clustering competence in emotional intelligence. In R. Bar-On, & J. D. A. Parker (Eds.), *The handbook of emotional intelligence: Theory, development, and assessment, and application at home, school, and in the workplace* (pp. 343–362). San Francisco, CA: Jossey-Bass.
- Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988–2008). *School Psychology Quarterly*, *24*(2), 106–118. <http://dx.doi.org/10.1037/a0016159>
- Cisco. (2008). *Equipping every learner for the 21st century*. San Jose, CA: Author. Retrieved from <http://www.cisco.com/web/about/citizenship/socio-economic/docs/GlobalEdWP.pdf>
- Dennis, J. M., Calvillo, E., & Gonzalez, A. (2008). The role of psychosocial variables in understanding the achievement and retention of transfer students at an ethnically diverse urban university. *Journal of College Student Development*, *49*(6), 535–550.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101.
- 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.
- Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432.
- Educational Testing Service. (2012). Relationships between Big Five and academic and workforce outcomes. Princeton, NJ: Author. Retrieved from https://www.ets.org/s/workforce_readiness/pdf/21334_big_5.pdf
- Ekvall, G. (1991). The organizational culture of idea-management: A clear climate for the management for ideas. In J. Henry & D. Walker (Eds.), *Managing innovation* (pp. 73–79). London: Sage.
- Facione, N., Facione, P., & Sanchez, C. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, *33*(8), 345–350.
- Faxon-Mills, S., Hamilton, L. S., Rudnick, M., & Stecher, B. M. (2013). *New assessment, better instruction? Designing assessment systems to promote instructional improvement* (RR-354-WFHF). Santa Monica, CA: RAND Corporation.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, *40*, 225–234.
- Herman, J., & Linn, B. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC Assessment Consortia* (CRESST Report 823). Los Angeles: National Center for Research on Evaluation, Standards, & Student Testing, University of California, Los Angeles. Retrieved from <http://www.cse.ucla.edu/downloads/files/CRESSTReport823.pdf>

- Huberman, M., Bitter, C., Anthony, J., & O'Day, J. (2014). *The shape of deeper learning: Strategies, structures, and cultures in deeper learning network high schools*. Washington, DC: American Institute of Research. Retrieved from <http://www.air.org/resource/shape-deeper-learning-strategies-structures-and-cultures-deeper-learning-network-high>
- Lester, J. C., Ha, E. Y., Lee, S. Y., Mott, B. W., Rowe, J. P., & Sabourin, J. L. (2013). Serious games get smart: Intelligent game-based learning environments. *AI Magazine*, 34(4), 31–45. doi: <http://dx.doi.org/10.1609/aimag.v34i4.2488>
- MacCann, C., Duckworth, A. L., & Roberts, R. (2009). Empirical identification of the major facets of conscientiousness. *Learning and Individual Differences*, 19, 451–458.
- Mayer, J. D., Caruso, D., & Salovey, P. (2000). Selecting a measure of emotional intelligence: The case for ability scales. In R. Bar-On, & J. D. Parker (Eds.), *Handbook of emotional intelligence* (pp. 320–342). New York: Jossey-Bass.
- Microsoft Partners in Learning, Pearson Foundation, & Gallup. (2013). *21st century skills and the workplace: A 2013 Microsoft Partners in Learning and Pearson Foundation Study*. Redmond, WA; New York, NY; Washington, DC: Authors.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108, 2693–2698. doi:10.1073/pnas.1010076108
- Myrseth, K. O., & Fishbach, A. (2009). Self-control: A function of knowing when and how to exercise restraint. *Current Directions in Psychological Science*, 18(4), 247–252.
- National Center on Education and the Economy. (2007). *Tough choices or tough times: The report of the New Commission on the Skills of the American Workforce* (executive summary). Washington, DC: Author. Retrieved from http://www.helios.org/uploads/docs/ToughChoices_EXECSUM.pdf
- National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.
- Norris, S. E. (2008). An examination of self-leadership. *Emerging Leadership Journeys*, 1(2), 43–61. Retrieved from https://www.regent.edu/acad/global/publications/elj/vol1iss2/ELJ_V1Is2_Norris.pdf
- Partnership for 21st Century Skills. (2008). *21st century skills, education and competitiveness: A resource and policy guide*. Tucson, AZ: Author. Retrieved from http://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf
- Robert, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology*, 58, 103–139.
- Saavedra, A. R., & Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Phi Delta Kappan*, 94(2), 8–13.
- Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st century competencies: Guidance for educators*. New York, NY: Asia Society. Retrieved from <http://asiasociety.org/files/gcen-measuring21cskills.pdf>
- Stecher, B. M., & Hamilton, L. S. (2014). *Measuring hard-to-measure student competencies* (RR-863-WFHF). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/research_reports/RR863
- Steward, G. L., Courtright, S. H., & Manz, C. C. (2010). Self-leadership: A multilevel review. *Journal of Management*, 37(1), 185–222. doi:10.1177/0149206310383911
- The William and Flora Hewlett Foundation. (n.d.). *Deeper learning*. Retrieved from <http://www.hewlett.org/programs/education/deeper-learning>

Yuan, K., & Le, V. (2012). *Estimating the percentage of students who were tested on cognitively demanding items through the state achievement tests* (WR-967-WFHF). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/working_papers/WR967

Yuan, K., & Le, V. (2014). *Measuring deeper learning through cognitively demanding test items: Results from the analysis of six national and international exams* (RR-483-WFHF). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/pubs/research_reports/RR483