



AUTHORS

Sara J. Finney, Ph.D.
James Madison University

Gabriel R. Gilmore, MA
James Madison University

Sarah Alahmadi, MS
James Madison University

Abstract

When engaging in outcomes assessment, higher education professionals (i.e., faculty, student affairs educators) are expected to gather reliable data and make valid inferences. Decisions about how to measure student learning and development outcomes impact inferences about the achievement of outcomes and determination of improvement efforts. Professionals may search for existing outcome measures due to lack of experience in the challenging instrument development process and/or the time required to construct a high-quality measure. To support professionals in their search, we created a tool that describes relevant repositories of measures. Given most professionals lack training in psychometrics, we purposefully categorized these repositories by the level of guidance they provide when selecting a measure. That is, in addition to identifying an existing measure and summarizing the measure's psychometric properties, some repositories provide an evaluation of the measure's quality. This resource facilitates the collection of high-quality data that informs valid inferences about student outcomes.

“What’s A Good Measure Of That Outcome?” Resources To Find Existing And Psychometrically Sound Measures

Student learning and development outcomes assessment is challenging and time consuming. The typical outcomes assessment process involves six general steps. The process begins by specifying measurable student learning and development outcomes—what students should know, value/appreciate, or be able to do (Step 1). These outcomes direct the activities completed during the remaining steps of the process. Faculty and student affairs educators map programming to the outcomes (Step 2). Evidence-informed programming (e.g., activities, pedagogies, strategies) that facilitates students achieving the desired outcomes should be intentionally selected (e.g., Finney & Buchanan, 2021; Finney et al., 2021; Horst, et al., 2021; Pope et al., in press; Pope et al., 2019; Smith & Finney, 2020). Once programming is mapped to outcomes, professionals must decide how to measure the outcomes (Step 3). A measure of an outcome (e.g., test, rubric, inventory, observational protocol) can be selected from existing measures or created. Inferences about student learning and development, and, in turn, program effectiveness are drawn from data gathered using these measures. Therefore, careful attention must be paid to how well measures align with intended outcomes, along with the measures' sensitivity to program impact (Bandalos, 2018; Suskie, 2009). The next steps (Steps 4 and 5) involve collecting implementation fidelity and outcomes data (e.g., Gerstner & Finney, 2013; Smith, et al., 2017, 2019). These data are then integrated, analyzed, interpreted, and reported (Step 6). Educators use the results to guide programming changes (Step 7), as the purpose of

CORRESPONDENCE

Email
finneysj@jmu.edu

the assessment process is to make data-based program modifications to improve student learning and development (Fulcher, et al., 2014).

Each step of the assessment process can be unpacked into more precise activities that involve particular skills (e.g., analyzing data, distinguishing between related but different outcomes, evaluating evidence of effectiveness). Our focus in the current paper is on determining how to measure outcomes (Step 3). Correct inferences about student ability, attitudes, skills, and behavior necessitate high-quality measures of those outcomes (Bandalos, 2018). Determining whether a high-quality measure exists or should be created is an essential activity at this step. Creation of a measure that allows for valid inferences requires a deep understanding of the outcome domain (e.g., critical thinking, intercultural competence, quantitative reasoning, career decisiveness, writing ability, ethical reasoning); skills to develop instructions, items, rubrics, or tasks that reflect the construct; an understanding of appropriate reliability and validity evidence, how to collect it, and how to interpret it; and pilot testing to improve the measure's psychometric properties. Although selection of an existing measure does not require skills to create a new measure or the study of its functioning, it does require an understanding of the outcome domain, a recognition of the need for relevant psychometric information, and skills to interpret those psychometric properties. In short, creation or selection of psychometrically-sound outcome measures both entail numerous competencies.

Faculty engaging in outcomes assessment are trained in a variety of disciplines (Leaderman & Polychronopoulos, 2019), and many are not formally trained in outcomes assessment via masters or doctoral programs (Hutchings, 2010; Nicholas & Slotnick, 2018). They instead gain knowledge, skills, and appreciation for assessment via workshops, conference presentations, webinars, and self-directed study (Curtis, et al., 2020). Unfortunately, there is rarely intentional, coherent sequencing of training opportunities, and many are targeted to novices. In turn, these trainings may not result in the depth of understanding and skill necessary for measurement-related concepts (e.g., reliability, validity, standard setting, factor analysis).

Unlike faculty landing in assessment positions from various domains across academic affairs (e.g., English, business), student affairs professionals are expected to understand and practice outcomes assessment (Finney & Horst, 2019a, 2019b). Yet, formal preparation programs may offer little training in measurement (e.g., Biddix et al., 2020; Cooper et al., 2016). According to Jablonski and colleagues (2006), "Even students from some of our best [student affairs] programs are inadequately trained in research, evaluation, and assessment." (p. 187).

Nonetheless, there are expectations regarding responsible practice in educational measurement. The preeminent source is *The Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014), which applies to anyone creating measures, gathering data, and using scores. Moreover, standards or competencies related to measurement and assessment have been created by education organizations (see Table 1).

In student affairs, several organizations have created documents that detail expectations related to the selection or development of outcome measures: the *Assessment Skills and Knowledge (ASK) Standards* (2006) created by the American College Personnel Association (ACPA), the *Professional Competencies* (2015, 2016) created jointly by ACPA and the Student Affairs Administrators in Higher Education Association (NASPA), and the *CAS Standards* (2019) created by the Council for the Advancement of Standards in Higher Education. The difficulty in measuring outcomes of student affairs and co-curricular programs has been acknowledged (ACPA, 2006, p. 4): "In student affairs, the articulation and assessment of student learning has been especially challenging given the complex psychosocial and cognitive constructs that are the hallmarks of our work with students. Messy constructs such as leadership, citizenship, appreciation for diversity, critical and ethical judgement, and a host of interpersonal and intrapersonal intelligences present unique measurement issues." These "measurement issues" require measurement skills.

In the *Assessment Skills Framework*, Horst and Prendergast (2020) outlined the knowledge, skills, and attitudes important for assessment in higher education. They

Although selection of an existing measure does not require skills to create a new measure or the study of its functioning, it does require an understanding of the outcome domain, a recognition of the need for relevant psychometric information, and skills to interpret those psychometric properties.

categorized each domain by skill level: novice, intermediate, and advanced. Regarding the measurement of student learning and development outcomes, there were six domains: evaluate instruments for alignment, evaluate instruments for context and resource considerations, evaluate instruments for reliability and validity, design selected response measures, design non-cognitive measures, and design performance assessments. In Table 1, we listed novice-level skills (i.e., providing basic explanations of concepts). Professionals at the intermediate and advanced levels (not listed) can provide detailed explanations and apply knowledge to real assessment efforts. More general than the *Assessment Skills Framework, The Code of Professional Responsibilities in Educational Measurement* (1995) serves as a guide for anyone engaged in educational assessment, including faculty and staff assessing student learning and development.

In academic affairs, *The Standards for Teacher Competence in Educational Assessment of Students* were developed to guide teacher educators in teacher education programs, to offer a mechanism for self-assessment by teachers, and to serve as a framework for workshop content (Brookhart, 2011). If teachers and teacher educators demonstrated the listed competencies, they may be sought out for consultation by those engaged in higher education outcomes assessment (Kerr et al., 2020). Unfortunately, even the profession of teaching, which involves a tremendous amount of testing and interpretation of scores, does not consistently provide instruction in measurement during formal training (Lukin et al., 2004; Plake et al., 1993; Wise, 1993). If a formal course in measurement is available, the course may not provide instruction on all topics relevant to assessment-related work due to the numerous topics covered in such a course, the diverse needs of students, and the level of preparation of students (Bandalos & Kopp, 2012).

We agree that the competencies listed in Table 1 are necessary to engage in high-quality assessment practice, and, like others (Curtis, et al., 2020), we are concerned that educators practicing outcomes assessment have not engaged in formal training or self-directed study to meet these expectations. Because construction of a new measure is time intensive, requiring training in item writing and measurement prior to creating the measure, it is most efficient to identify existing measures. If no existing measures can be located or none are of sufficient quality, then the time-consuming process of creating a new measure should be pursued. Unfortunately, resources to guide locating and selecting high-quality existing measures are not well-advertised or organized. Thus, to facilitate the assessment of student learning and development outcomes using high-quality measures, we provide a didactic resource to foster the use of repositories of measures.

...to facilitate the assessment of student learning and development outcomes using high-quality measures, we provide a didactic resource to foster the use of repositories of measures.

Our resource differs from previous summaries of available surveys and measures used in post-secondary settings. For example, a 2001 American Council on Education and Association for Institutional Research report summarized the characteristics of 27 national assessments of institutional quality (Borden & Owens, 2001). These assessments include surveys students complete prior to enrollment (e.g., expectations about college), while enrolled in college (e.g., perceptions of college experiences, satisfaction), and after graduation (e.g., reflections on the impact of college). The report also included a few commercial measures of student learning outcomes (e.g., writing, critical thinking). Unlike the measurement repositories we describe below, this report does not discuss the quality of these measures. Although decades old, this report is useful in that it reflects the type of data collected to address accountability and improvement 20 years ago (prevalence of surveys collecting perceptions of college and institutions). Currently, high-quality accountability and improvement efforts emphasize student learning and development outcomes tied to intentional programming, which necessitates high-quality measures of these outcomes.

Description of the New Resource: Organization of Measurement Repositories

To facilitate faculty members', student affairs professionals', and assessment specialists' search for measures, we created a [resource](#) that identifies and organizes measurement repositories relevant to higher education outcomes. Repositories of existing measures differ in their utility; thus, we sorted them into three tiers according to the

Table 1
Professional standards and competences related to the development or selection of outcome measures

ACPA ASK Standards	ACPA NASPA Competencies	CAS Standards	Assessment Skills Framework (novice-level only)	Professional Responsibilities in Educational Measurement	Teacher Competencies in Assessment
Identify strengths and weaknesses of existing measures	Select measures that fit with assessment purposes	Ensure measures and methods are rigorous and reflect the characteristics of validity, reliability, and trustworthiness	Describe basic types of instruments and intended uses (e.g., indirect, direct, selected response, constructed response, cognitive, non-cognitive)	Conduct thorough evaluation of available measures that may be valid for intended uses	Skilled in selecting assessment methods appropriate for instructional decisions
Create measure with effective wording, format, and appropriate administration method	Utilize student learning and development research to inform content and design of assessment tools	Employ multiple measures and methods of data collection	Describe pros and cons of selecting an existing measure versus	Inform users of appropriateness of assessment for intended use, protection of examinee rights, costs, known consequences and limitations	Skilled in developing assessment methods appropriate for instructional decisions
Select most appropriate measure for desired outcome	Facilitate appropriate data collection for assessment purposes	Implement assessment process that is culturally responsive, inclusive, and equitable	Describe advantages and disadvantages of using different types of measures	Select measure based on evidence of technical quality not insubstantial claims	Skilled in administering, scoring, and interpreting results of both externally produced and self-produced assessments
Develop rubrics	Assess legitimacy and validity of various methods of data collection	Use methods and measures that allow for the collection of data that reflect intended outcomes	Match instrument to SLO	Comply with security precaution	Skilled in using assessment results when making decisions about students, instruction, developing curriculum, and improvement
Determine manner in which those with disabilities will use measure	Use culturally relevant and culturally appropriate terminology		Describe pros and cons of using commercial versus non-commercial measures	Plan accommodations for test-takers with disabilities when developing assessments	Skilled in communicating assessment results to students, parents, and other educators
Review a measure for inclusive and accessible language			Acknowledge importance of considering reliability and validity when selecting measure	Ensure assessments are developed to meet technical and legal standards	Skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information
Use measure with rigor appropriate for intended use			Describe common types of reliability and validity evidence	Caution users against most likely misinterpretations/ misuses of data	
			Identify components of multiple-choice item (e.g., stem, distractor)	Correct substantive inaccuracies in assessments as soon as feasible	
			Identify best practices for constructing selected response measures (e.g., use test blueprint, pilot items, revise)	Develop assessments free from bias due to characteristics irrelevant to construct being measured, such as gender, age, ethnicity, disability, SES	
			Identify characteristics of non-cognitive measures (e.g., variety of response options)	Develop score reports that promote understanding assessment results	
			Identify best practice for constructing noncognitive measures	Recommend against assessment likely to be administered, scored, and used in invalid manner for reasons of race, ethnicity, age, gender, disability, language background, SES, or religion	
			Identify basic rubric components (e.g., rating scale, scoring criteria)	Make information available about steps to develop and score assessment, including current information regarding reliability, validity, scoring and reporting	
			Distinguish holistic and analytic rubrics (advantages of each)		

information they provide. Some repositories simply identify measures aligned with a particular outcome and indicate where the measures can be found (what we refer to as Tier Three repositories). Other repositories include a summary of the psychometric information associated with the measure (what we refer to as Tier Two repositories). The most helpful repositories, in our opinion, are those that provide a review or rating of the measure's quality given the psychometric information (what we refer to as Tier One repositories).

The most helpful repositories, in our opinion, are those that provide a review or rating of the measure's quality given the psychometric information (what we refer to as Tier One repositories).

For each repository, we provide its name and web link, description of the resource, information provided about the measures' characteristics and quality, and five example measures. These five measures serve simply as exemplars and a mechanism to quickly access and examine the database.

Each repository is further labeled by the CAS Learning Outcomes Domains. CAS "promotes standards to enhance opportunities for student learning and development from higher education programs and services" (CAS, 2015, para. 1). CAS has developed six student outcome domains: knowledge acquisition, construction, integration, and application; cognitive complexity; intrapersonal development; interpersonal competence; humanitarianism and civic engagement; and practical competence. All six domains are listed for each repository, and the specific domains that the repository includes are bolded and *. For example, the database "emerge" has knowledge acquisition, construction, integration, and application; interpersonal competence; and practical competence bolded and *. Hence, in this repository, you will find measures that align with those specific student learning and development domains. For those who do not use the CAS outcome domains, but rather outcomes specified by the Liberal Education and America's Promise (LEAP) initiative, the Degree Qualifications Profile (DQP), Learning Reconsidered, or other organizations, CAS created a useful [crosswalk](#) of outcomes by organization to show their overlap.

To create our resource, we independently searched the internet for measurement repositories and concatenated the repositories we each found. We independently studied each repository to contribute to its description and example measures before identifying the appropriate tier and relevant CAS outcomes. We then excluded measurement repositories if they did not include measures relevant for the higher education context and population. Two students (one graduate and one undergraduate) examined the new resource and provided us with feedback (e.g., broken links, incomplete directions to access resource). We then piloted the resource during a week-long professional development session offered to United States and international faculty and student affairs professionals.

To create our resource, we independently searched the internet for measurement repositories and concatenated the repositories we each found.

How to Use the New Resource: Didactic Examples

To facilitate familiarity and use of this resource, we walk through two repositories in each tier and explain the type of information and psychometric evaluation they provide.

Tier One

Repositories in Tier One provide psychometric information (e.g., reliability of scores, validity evidence) as well as their own rating of the quality of the measure. This rating can be in the form of a number, statement, or recommendation for use. Ratings may not be provided for every measure but are available for the majority of measures in the repository. We consider repositories in this tier of the highest utility to select evidence-informed existing measures.

Mental Measurements Yearbook (MMY) Series

Tier One houses the Mental Measurements Yearbook (MMY) series, which is published by the Buros Center for Testing (Carlson et al., 2017). The MMY addresses the need for informed test evaluation by offering expert reviews of existing measures. Typically, detailed descriptions of the measures referenced in the MMY are provided, along with two reviews conducted by volunteer professional measurement experts. Volunteer reviewers are selected for each measure based on their domain-specific knowledge and training in measurement and psychometric evaluation. They also must carry a terminal degree (e.g.,

PhD, PsyD, EdD). To qualify for a review in MMY, a measure must be commercial, available in the English language, new or widely used, and provide psychometric qualities (e.g., reliability estimates, validity evidence). Reviews published in the MMY can be accessed through electronic databases, such as EBSCO or Ovid, to which many academic libraries subscribe. Additionally, the Buros Center for Testing offers a *Test Reviews Online* service, through which reviews for a particular test can be purchased.

To demonstrate the utility of MMY for selecting an existing measure, we searched for measures of critical thinking, a common student learning outcome in higher education. *The Cornell Critical Thinking Test* (CCTT; Ennis et al., 1985) was one of the tests identified by our search. The MMY test entry for the CCTT first provides descriptive information about the measure, including authors of the test, publication date, publisher information, purpose, population, scores, administration mode, testing time, price, name of MMY reviewers, yearbook volume in which the test appears, and relevant references (Carlson et al., 2017). Next, the two professional reviews of the measure are provided. The reviews typically summarize the developmental history of the measure, the norming samples, and evidence of technical quality provided by the test developers in the test manual. Following that, the reviewers provide their own commentary and recommendation for use of the measure. Test entries are often concluded with references to articles, manuals, or books that informed the experts' reviews.

For example, the first reviewer of the CCTT noted that “the Cornell Critical Thinking Test provides an objective method for evaluating critical thinking abilities that have been identified as necessary for individuals to respond appropriately to problems encountered in our complex world” (Porter, 2017). The second reviewer stated that the data presented supports the use of the test, but also noted the need for further empirical evidence to support the inferences made based on the test's scores. Specifically, the CCTT may not be appropriate for individual decision-making (Schafer, 2017). Overall, the reviewers support the use of the measure for the purposes of outcomes assessment or program evaluation, but advise against its use for making critical, person-level decisions. Such comprehensive and insightful appraisal of the measure and its appropriate use affords valuable information for informed measure selection.

Evidence-Based Measures of Empowerment for Research on Gender Equality (EMERGE)

EMERGE (2017a) is another Tier One repository that offers expert evaluation of carefully curated measures that assess knowledge, attitudes, and behaviors relating to gender equality and empowerment. Measures housed in the repository were selected with the help of gender equality and empowerment experts and reviews of available literature. For a measure to be included, it must have the following characteristics: quantitative in nature; published in either a national or international survey, or a peer-reviewed journal with impact factor ≥ 1 ; and include empirical evidence for reliability and validity.

To provide ratings of the psychometric quality and utility of the measures, trained EMERGE staff score each measure (EMERGE, 2017b). The psychometric properties rated include the following aspects: formative research (qualitative research, theoretical framework, expert input, and pilot testing), reliability (internal consistency, test-retest, and inter-rater reliability), and validity evidence (content, face, criterion, and construct forms of validity). The scores for the three psychometrics aspects are aggregated into a total score: “Low” ($\leq 33.3\%$) “Medium” (33.4% - 66.6%), “High” ($\geq 66.7\%$), or “No Data” if the measure could not be scored. Another score utilizes information provided by Google Scholar on the number of citations of the measure's primary source: “Low” (< 20 citations), “Medium” (20 - 49 citations), “High” (≥ 50 citations), or “No Data” if the Google Scholar citation record is not available.

An example of a measure found in this repository is the *Illinois Rape Myth Acceptance Scale* (Payne et al., 1999). This measure may be useful for university bystander intervention programs designed to influence outcomes related to intervening in a potential assault. EMERGE provides a brief description of the measure, its purpose, intended population,

Such comprehensive and insightful appraisal of the measure and its appropriate use affords valuable information for informed measure selection.

intended age range, list of items, response scale (e.g., Likert, multiple choice), and the measure's primary citation. EMERGE staff's ratings of the measure's psychometric properties and citation frequency is highlighted and explained (i.e., whether each of the scoring aspects received full or partial points, were not assessed, or were not applicable). This measure received a "high" psychometric score (EMERGE, 2017c), which would support its use as an outcome measure. Another valuable resource found on the EMERGE site is a report explaining how to utilize measurement in the field of gender equality and empowerment, how to identify psychometrically sound measures, and how to adapt a measure to different cultural contexts (Bhan et al., 2017).

Tier Two

Tier Two repositories provide psychometric information (e.g., reliability, validity) for the measures, but do not provide their own rating of the quality of the measures. Psychometrics may not be provided for every measure but are available for most measures in the databases. The majority of the repositories in our resource fall in this category. Below we provide two examples.

RAND Educational Assessment Finder

The RAND Educational Assessment Finder (RAND Corporation, n.d.) requires that included measures reflect interpersonal (e.g., empathy, leadership), intrapersonal (e.g., adaptability, perseverance), or higher-order thinking constructs (e.g., critical thinking, creativity), are appropriate for use in educational settings, and are appropriate for populations of students in the United States. To summarize the psychometric quality of a measure, RAND professionals read the publicly available studies that examined the reliability and/or validity of the measure. The psychometric summaries RAND creates are then shared with the measures' developers to provide any corrections before the summaries are published in the repository. The RAND Education Assessment Finder includes both commercial and non-commercial (i.e., free) measures and identifying free measures is facilitated by the "Fee for Use" search filter (Schweig, et al., 2018).

The Cornell Critical Thinking Test (CCTT; Ennis et al., 1985) can be found in the RAND Education Assessment Finder, but the information this repository provides lacks the expert reviews provided by the Tier One MMY repository. The RAND Education Assessment Finder summarizes the following aspects of the measure: purpose, publication year, administration method, number of items, item format, administration time, available languages, fee, scoring, interpretive information, reliability evidence, validity evidence, links to obtain a copy of the measure, and references (RAND Corporation, 2018). Given the lack of review or rating conducted by measurement experts, the user manual of the RAND repository states that "because interpreting validity evidence is complex and generally requires measurement expertise, users are encouraged to seek input from measurement experts to evaluate the adequacy and relevance of the available evidence for a particular assessment purpose" (Hamilton et al., 2018, p. 11).

ETS Research Report Series

Educational Testing Services (ETS) publishes the ETS Research Report Series journal. This journal, which is freely accessible via the Wiley Online Library, includes resources related to psychometric and statistical methods, educational evaluation, and large-scale assessment. Highly relevant to higher education outcomes assessment are the syntheses of current literature on measures of pertinent student learning outcomes (e.g., quantitative literacy, intercultural competence, written communication, critical thinking). The syntheses contain information on the development of the outcome measures, the available reliability and validity information, target populations, and typically conclude with future directions for assessment in that domain. These reports are classified in Tier Two because they detail the psychometric properties of the measures, but do not include conclusions or interpretations regarding the psychometric quality of the measures.

Tier Two repositories provide psychometric information (e.g., reliability, validity) for the measures, but do not provide their own rating of the quality of the measures.

An example report from the ETS Research Report Series provides the current state of assessment of civic competency and engagement in higher education (Torney-Purta et al., 2015). The report includes current definitions and conceptualizations of the construct in addition to (over 25) available measures. Measures are contrasted in terms of themes, test developer, test format, and length. The report discusses implications related to the reliability of scores and the validity inferences for such a multifaceted outcome. The report ends with a proposed framework for future assessments of civic competency and engagement to facilitate better measurement.

... identifying high-quality existing measures promotes common measurement of outcomes and comparison of results across different programming, teaching approaches, and institutions.

Tier Three

Unlike Tier One and Two, repositories in Tier Three do not provide psychometric information (e.g., reliability, validity) for the measures or their own rating of the quality of the measures. Often, the psychometric information can be found in the linked source articles.

PsycTests

PsycTests is a repository produced by the American Psychological Association (2021). It holds more than 60,000 measures, many of which are free to use. The measures are collected from various sources: directly from authors, peer-reviewed journals, books, dissertations, and websites.

Returning to our example of finding an existing measure of critical thinking, we searched PsycTests. *The Halpern Critical Thinking Assessment* (HCTA; Halpern, 2010) emerged as an option. The repository provided a “Master Test Profile” for the HCTA that included a description of the test, its purpose, the developer’s contact information, and whether it is commercial, among other basic pieces of information. Typically, no information regarding reliability or validity is provided. If psychometric information is available in the original source of the test, the PsycTest entry will include the information, but no professional review of such information is provided. Thus, the amount of information provided by this Tier Three repository is limited compared to that provided by repositories from Tier One and Tier Two.

Assessment and Curriculum Support Center

Another Tier Three repository is the Assessment and Curriculum Support Center at the University of Hawai’i at Mānoa (Assessment and Curriculum Support Center, 2020). This center specializes in assessment of learning outcomes for improvement, and it includes a collection of rubrics used to assess outcomes such as civic knowledge, collaboration, critical thinking, ethical deliberation, integrative learning, information literacy, intercultural knowledge, and others. The repository contains links to the original sources of the rubrics. As such, it is a collection of performance assessments but does not review their psychometric quality. The user is encouraged to collaborate with an assessment expert to evaluate these measures.

Discussion

Our goal was to create a resource of measurement repositories that supports educators’ search for high-quality measures. These repositories can increase efficiency in the outcomes assessment process and the trustworthiness of resulting scores. However, we want to stress that high-quality scores and valid inferences require more than quality measures. Students may have negative attitudes (Zilberberg, et al., 2012; Zilberberg, et al., 2013; Zilberberg, et al., 2014) or emotions (Finney, Perkins & Satkus, 2020; Finney, Satkus, & Perkins, 2020) toward higher education outcomes assessment initiatives. Thus, students may not be motivated to provide valid responses (Barry et al., 2010; Wise & DeMars, 2005) or attend testing sessions (Brown & Finney, 2011; Kopp & Finney, 2013; Swerdzewski et al., 2009). In turn, professionals should engage in strategies to increase examinee motivation (Barry & Finney, 2009; Finney, et al., 2016; Myers & Finney, 2021) or analyses that address the lack of motivation (Swerdzewski et al., 2011; Wise & DeMars, 2010).

Beyond identifying and evaluating existing measures, repositories of measures have additional benefits for faculty, student affairs educators, and assessment specialists. Measurement repositories showcase the various definitions and operationalization of what some professionals assume to be simple outcomes. They force educators to clearly articulate the outcome of interest given the number of different but related outcome measures that exist. They counter vague language describing outcomes, which facilitates alignment between outcomes and effective programming. Moreover, identifying high-quality existing measures promotes common measurement of outcomes and comparison of results across different programming, teaching approaches, and institutions.

Organizations or individuals responsible for a group of programs could consider using [measurement repositories] to identify and endorse a specific set of outcome measures that are both reliable and valid for the populations served across a variety of domains. Endorsing a specific set of outcome measures could allow for consistency in tracking core outcomes or indicators of effectiveness across an array of programs (Acosta et al., 2014, p. 3).

Measurement repositories also showcase the rigorous process of scale development. By reviewing psychometric evidence, they uncover the need for additional psychometric study before trustworthy inferences can be made about student learning and development on our campuses.

References

- Acosta, J., Reynolds, K., Gillen, E., Feeney, K., Farmer, C., & Weinick, R. (2014). The RAND Online Measure Repository for evaluating psychological health and traumatic brain injury programs. *Rand Health Quarterly*, 4(1). <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051979/>
- Assessment and Curriculum Support Center (2020, November). *Rubric Bank*. <http://manoa.hawaii.edu/assessment/resources/rubric-bank/>
- American College Personnel Association. (2006). *ASK standards: Assessment skills and knowledge content standards for student affairs practitioners and scholars*. Washington, DC: Author.
- American College Personnel Association & National Association of Student Personnel Administrators. (2015). *ACPA/NASPA professional competency areas for student affairs educators*. Washington, DC: Authors. https://www.naspa.org/images/uploads/main/ACPA_NASPA_Professional_Competencies_FINAL.pdf
- American College Personnel Association & National Association of Student Personnel Administrators. (2016). *ACPA/NASPA professional competencies rubrics*. Washington, DC: Authors. https://www.naspa.org/images/uploads/main/ACPA_NASPA_Professional_Competency_Rubrics_Full.pdf
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- American Psychological Association (2021). *PsycTests*. <https://www.apa.org/pubs/databases/psyc-tests>
- Bandalos, D.L. (2018). *Measurement theory and applications for social sciences*. Guilford Press.
- Bandalos, D.L., & Kopp, J.P. (2012). Teaching introductory measurement: Suggestions for what to include and how to motivate students. *Educational Measurement: Issues and Practice*, 31(2), 8-13. <https://doi.org/10.1111/j.1745-3992.2012.00229.x>
- Barry, C.L. & Finney, S.J. (2009). Does it matter how data are collected? A comparison of testing conditions and the implications for validity. *Research & Practice in Assessment*, 3, 1-15. <http://www.rpajournal.com/dev/wp-content/uploads/2012/05/A33.pdf>
- Barry, C.L., Horst, S.J., Finney, S.J., Brown, A.R., & Kopp, J.P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing*, 10, 342-363. <https://doi.org/10.1080/15305058.2010.508569>
- Bhan, N., Jose, R., McDougal, L., & Raj, A. (2017). *EMERGE measurement guidelines report 1: What is measurement and how do we quantitatively measure gender equality and empowerment?* Center on Gender Equity and Health (GEH), University of California, San Diego School of Medicine. San Diego, CA. <https://bit.ly/3dzA4Z5>
- Brown, A.R. & Finney, S.J. (2011). Low-stakes testing and psychological reactance: Using the Hong Psychological Reactance Scale to better understand compliant and non-compliant examinees. *International Journal of Testing*, 11, 248-270. <https://doi.org/10.1080/15305058.2011.570884>
- Biddix, J.P., Collom, G.D., & Roberts, D.M. (2020). Scholarship, professional development, and community of practice in student affairs assessment. *College Student Affairs Journal*, 38, 157-171 (EJ1275325). ERIC. <https://files.eric.ed.gov/fulltext/EJ1275325.pdf>
- Borden, V.M. & Owens, J.L. (2001) *Measuring quality: Choosing among surveys and other assessments of college quality*, Washington, DC: American Council of Education and Association of Institutional Research (ED457767). ERIC. <https://files.eric.ed.gov/fulltext/ED457767.pdf>
- Brookhart, S.M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3-12. <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Carlson J.F., Geisinger, K.F., & Jonson, J.L. (Eds.). (2017). *The twentieth mental measurements yearbook*. Buros Center for Testing.
- Cooper, J., Mitchell, D., Jr., Eckerle, K., & Martin, K. (2016). Addressing perceived skill deficiencies in student affairs graduate preparation programs. *Journal of Student Affairs Research and Practice*, 53, 107-117. <https://doi.org/10.1080/19496591.2016.1121146>

- Council for the Advancement of Standards in Higher Education. (2019). *CAS professional standards for higher education* (10th ed.). Washington, DC: Author.
- Curtis, N.A., Anderson, R.D & Van Dyke, R. (2020). A field without a discipline? Mapping the uncertain and often chaotic route to becoming an assessment practitioner. *Research & Practice in Assessment*, 15(1), 1-8. <https://www.rpajournal.com/dev/wp-content/uploads/2020/08/A-Field-Without-A-Discipline.pdf>
- Ennis, R.H., Millman, J., & Tomko, T.N. (1985). *Cornell critical thinking tests level X & level Z: Manual*. Pacific Grove, CA: Midwest Publications.
- Evidence-based Measures of Empowerment for Research on Gender Equality (EMERGE) (2017a). *Overview*. <https://emerge.ucsd.edu/measurement-overview/>
- Evidence-based Measures of Empowerment for Research on Gender Equality (EMERGE) (2017b). *Scoring methodology*. <https://emerge.ucsd.edu/scoring-methodology/>
- Evidence-based Measures of Empowerment for Research on Gender Equality (EMERGE) (2017c). *Illinois Rape Myth Acceptance (IRMA) Scale*. https://emerge.ucsd.edu/r_ag79krbvzx9zcon/
- Finney, S.J. & Buchanan, H.A. (2021). A more efficient path to learning improvement: Using repositories of effectiveness studies to guide evidence-informed programming. *Research & Practice in Assessment*, 16(1), 36-48. <https://www.rpajournal.com/dev/wp-content/uploads/2021/04/Repositories-of-Effectiveness-Studies-to-Guide-Programming.pdf>
- Finney, S.J., & Horst, S.J. (2019a). Standards, standards, standards: Mapping professional standards for outcomes assessment to assessment practice. *Journal of Student Affairs Research and Practice*, 56, 310-325. <https://doi.org/10.1080/19496591.2018.1559171>
- Finney, S.J., & Horst, S.J. (2019b). The status of assessment, evaluation, and research in student affairs. In V. L. Wise & Z. Davenport (Eds.), *Student affairs assessment, evaluation, and research: A guidebook for graduate students and new professionals* (pp. 3-19). Springfield, IL: Charles Thomas Publisher.
- Finney, S.J., Perkins, B.A., & Satkus, P. (2020). Examining the simultaneous change in emotions during a test: Relations with expended effort and test performance. *International Journal of Testing*, 20, 274-298. <https://doi.org/10.1080/15305058.2020.1786834>
- Finney, S.J., Satkus, P. & Perkins, B.A. (2020). The effect of perceived test importance and examinee emotions on expended effort during a low-stakes test: A longitudinal panel model. *Educational Assessment*, 25, 159-177. <https://doi.org/10.1080/10627197.2020.1756254>
- Finney, S. J., Sundre, D.L., Swain, M.S., & Williams, L.M. (2016). The validity of value-added estimates from low-stakes testing contexts: The impact of change in test-taking motivation and test consequences. *Educational Assessment*, 21, 60-87. <https://doi.org/10.1080/10627197.2015.1127753>
- Finney, S.J., Wells, J.B., & Henning, G.W. (2021). *The need for program theory and implementation fidelity in assessment practice and standards*. (Occasional Paper No. 52). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. https://www.learningoutcomesassessment.org/wp-content/uploads/2021/03/Occ_Paper_51-1.pdf
- Fulcher, K.H., Good, M.R., Coleman, C.M., & Smith, K.L. (2014). *A simple paper for learning improvement: Weigh pig, feed pig, weigh pig*. (Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. <https://files.eric.ed.gov/fulltext/ED555526.pdf>
- Gerstner, J.J. & Finney, S.J. (2013). Measuring the implementation fidelity of student affairs programs: A critical component of the outcomes assessment cycle. *Research & Practice in Assessment*, 8, 15-28. <https://www.rpajournal.com/dev/wp-content/uploads/2013/11/SF2.pdf>
- Halpern, D. (n.d.). *HCTA Halpern Critical Thinking Assessment* [Database record]. APA PsycTests. Retrieved February 25, 2021, from <https://doi.org/10.1037/t10940-000>
- Hamilton, L.S., Stecher, B.M., Schweig, J. & Baker, G. (2018). *RAND Education Assessment Finder: User instructions*. RAND Corporation. <https://www.rand.org/education-and-labor/projects/assessments.html>

- Halpern, D.F. (2010). *Halpern Critical Thinking Assessment*. SCHUHFRIED (Vienna Test System): Möedling, Austria.
- Horst, S.J., Finney, S.J., Prendergast, C.O., Pope, A.M. & Crewe, M. (2021). The credibility of inferences from program effectiveness studies published in student affairs journals: Potential impact on programming and assessment. *Research & Practice in Assessment*, 16(2), 17 – 32. <https://www.rpajournal.com/dev/wp-content/uploads/2021/09/The-Credibility-of-Inferences-from-Program-Effectiveness-Studies.pdf>
- Horst, S.J. & Prendergast, C.O. (2020). The Assessment Skills Framework: A taxonomy of assessment knowledge, skills and attitudes. *Research and Practice in Assessment*, 15(1), 1-25. <https://www.rpajournal.com/dev/wp-content/uploads/2020/05/The-Assessment-Skills-Framework-RPA.pdf>
- Hutchings, P. (2010). *Opening doors to faculty involvement in assessment*. (Occasional Paper No. 4). University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Jablonski, M., Mena, S.B., Manning, K., Carpenter, S. & Siko, K.L. (2006) Scholarship in student affairs revisited: The summit on scholarship. *NASPA Journal*, 43, 182-200. <https://doi.org/10.2202/1949-6605.1729>
- Kerr, K.G., Edwards, K.E., Tweedy, J., Lichterman, H.L. & Knerr, A.R. (2020). *The curricular approach to student affairs: A revolutionary shift for learning beyond the classroom*. Sterling, VA: Stylus.
- Kopp, J.P. & Finney, S.J. (2013). Linking academic entitlement and student incivility using latent means modeling. *Journal of Experimental Education*, 81, 322-336. <https://doi.org/10.1080/00220973.2012.727887>
- Leaderman, E.C., & Polychronopoulos, G.B. (2019). Humanizing the assessment process: How the RARE model informs best practices. *Research & Practice in Assessment*, 14, 30-40. https://www.rpajournal.com/dev/wp-content/uploads/2020/03/Humanizing-the-Assessment-Process_r.pdf
- Lukin, L.E., Bandalos, D.L., Eckhout, T.J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23, 26-32. <https://doi.org/10.1111/j.1745-3992.2004.tb00156.x>
- Myers, A.J. & Finney, S.J. (2021). Does it matter if examinee motivation is measured before or after a low-stakes test? A moderated mediation analysis. *Educational Assessment*, 26, 1-19. <https://doi.org/10.1080/10627197.2019.1645591>
- Nicholas, M.C., & Slotnick, R.C. (2018). *A portrait of the assessment professional in the United States: Results from a national survey*. (Occasional Paper No. 34). University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. <https://bit.ly/3dGpOX4>
- Payne, D.L., Lonsway, K.A., & Fitzgerald, L.F. (1999). Rape myth acceptance: Exploration of its structure and its measurement using the Illinois Rape Myth Acceptance Scale. *Journal of Research in Personality*, 33(1), 27-68. <https://doi.org/10.1006/jrpe.1998.2238>.
- Plake, B.S., Impara, J.C., & Fager, J.J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12, 10-12. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Pope, A.M., Finney, S.J., & Bare, A. (2019). The essential role of program theory: Fostering theory-driven practice and high-quality outcomes assessment in student affairs. *Research & Practice in Assessment*, 14, 5-17. <https://www.rpajournal.com/dev/wp-content/uploads/2019/07/L1.pdf>
- Pope, A.M., Finney, S.J. & Crewe, M. (in press). Evaluating the effectiveness of an academic success program: Showcasing the importance of theory to practice. *Journal of Student Affairs Inquiry*.
- Porter, J.Y. (2017). [Test review of Cornell Critical Thinking Tests, Fifth Edition]. In Carlson J. F., Geisinger, K. F. & Jonson, J. L. (Eds.). *The twentieth mental measurements yearbook*. Buros Center for Testing.
- RAND Corporation (n.d.). *RAND Education Assessment Finder*. <https://www.rand.org/education-and-labor/projects/assessments.html>
- RAND Corporation (2018). *Cornell Critical Thinking Test (Level X) (CCTT)*. <https://www.rand.org/education-and-labor/projects/assessments/tool/1971/cornell-critical-thinking-test-level-x-cctt.html>
- Schafer, W.D. (2017). [Test review of Cornell Critical Thinking Tests, Fifth Edition]. In Carlson J. F., Geisinger, K. F. & Jonson, J. L. (Eds.). *The twentieth mental measurements yearbook*. Buros Center for Testing.

- Schweig, J., Baker, G., Hamilton, L. S., & Stecher, B.M. (2018). *Building a repository of assessments of interpersonal, intrapersonal, and higher-order cognitive competencies*. RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR2500/RR2508/RAND_RR2508.pdf
- Smith, K.L., & Finney, S.J. (2020). Elevating program theory and implementation fidelity in higher education: Modeling the process via an ethical reasoning curriculum. *Research and Practice in Assessment*, 15(2), 1-13. <https://www.rpajournal.com/dev/wp-content/uploads/2020/09/Elevating-Program-Theory-and-Implementation-Fidelity-in-Higher-Education.pdf>
- Smith, K.L., Finney, S.J., & Fulcher, K.H. (2017). Actionable steps for engaging assessment practitioners and faculty in implementation fidelity research. *Research & Practice in Assessment*, 12, 71-86. <http://www.rpajournal.com/dev/wp-content/uploads/2018/02/NIB1.pdf>
- Smith, K.L., Finney, S.J., & Fulcher, K.H. (2019). Connecting assessment practices with curricula and pedagogy via implementation fidelity data. *Assessment and Evaluation in Higher Education*, 44, 263-282. <https://doi.org/10.1080/02602938.2018.1496321>
- Suskie, L. (2009). *Assessing student learning: A common sense guide*. San Francisco, CA: Jossey-Bass.
- Swerdzewski, P.J., Harmes, J.C., & Finney, S.J. (2009). Skipping the test: Using empirical evidence to inform policy related to students who avoid taking low-stakes assessment in college. *Journal of General Education*, 58, 167-195. <http://doi.org/10.1353/jge.0.0043>
- Swerdzewski, P.J., Harmes, J.C., & Finney, S.J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education*, 24, 162 – 188. <https://doi.org/10.1080/08957347.2011.555217>
- Torney Purta, J., Cabrera, J.C., Crofts Roohr, K., Liu, O.L., & Rios, J.A. (2015). *Assessing civic competency and engagement in higher education: Research background, frameworks, and directions for next generation assessment* (Research Report No. RR 15 34). Princeton, NJ: Educational Testing Service. <https://doi.org/10.1002/ets2.12081>
- Wise, S.L. (Ed.). (1993). *Teacher training in measurement and assessment skills*. Lincoln, NE: Buros Institute of Mental Measurements.
- Wise, S.L. & DeMars, C.E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, 15, 27–41. <https://doi.org/10.1080/10627191003673216>
- Wise, S.L., & DeMars, C.E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10, 1–17. https://doi.org/10.1207/s15326977ea1001_1
- Zilberberg, A., Anderson, R.A., Finney, S.J., & Marsh, K.R. (2013). American college students' attitudes toward institutional accountability testing: Developing measures. *Educational Assessment*, 18, 208-234. <https://doi.org/10.1080/10627197.2013.817153>
- Zilberberg, A., Anderson, R.A., Swerdzewski, P.J., Finney, S.J., & Marsh, K.R. (2012). Growing up with No Child Left Behind: An initial assessment of the understanding of college students' knowledge of accountability testing. *Research & Practice in Assessment*, 7, 12-25. <https://files.eric.ed.gov/fulltext/EJ1062686.pdf>
- Zilberberg, A., Finney, S.J., Marsh, K.R. & Anderson, R.A. (2014). The role of students' attitudes and test-taking motivation on the validity of college institutional accountability tests: A path analytic model. *International Journal of Testing*, 14, 360-384. <https://doi.org/10.1080/15305058.2014.928301>