Professional Development in Equity-Centered Outcomes Assessment

The Utility of a Meta-Assessment Rubric

Autumn N. Wild
James Madison University
Doctoral Assistant,
Center for Assessment & Research Studies (CARS)
autumn.wild.1@gmail.com | LinkedIn

Riley K. Herr
James Madison University
Doctoral Assistant, CARS
rileykherr@gmail.com | LinkedIn

Sara J. Finney, Ph.D.
James Madison University
Associate Director, CARS
Professor of Graduate Psychology
finneysj@jmu.edu | LinkedIn

Abstract: Core skillsets for student affairs educators have been articulated in several professional standards. However, the expected assessment skillset is not always addressed in graduate programs; in turn, many student affairs educators feel unprepared to engage in outcomes assessment. Our study showcases the utility of a new equity-centered meta-assessment rubric to provide needed assessment training to student affairs educators. Results support the use of this rubric to advance equitable assessment and programming on college campuses.

Keywords: outcomes assessment, equity in assessment, professional development in assessment

For decades, emphasis on accountability and improvement within higher education has steadily grown (Ewell, 2009). With the rising cost of a college education, assessment practices to demonstrate the value of a degree have gained considerable attention. Student affairs educators are particularly scrutinized in this regard, as their contributions to student learning and development are frequently questioned (Biddix et al., 2020; Rincón & Castillo-Montoya, 2018). As a result, the expectations for student affairs educators have grown, with increasing demands for professionals to engage in high-quality assessment practices to prove the value of student affairs programming to enhance student learning and development (Wawrzynski et al., 2015).

Professional Expectations Related to Outcomes Assessment

Student affairs educators are expected to demonstrate a wide range of competencies. These expectations are elucidated via professional standards (Finney & Horst, 2019a, 2019b). There are three main sets of standards for the profession. Two of these sets are personal competency standards: the Assessment Skills and Knowledge (ASK) Standards (American College Personnel Association, 2006), and the ACPA-NASPA Competencies (American College Personnel Association & National Association of Student Personnel Administrators, 2015). The third set is program-focused from the Council for the Advancement of Standards in Higher Education (CAS, 2023).

The ASK Standards (American College Personnel Association, 2006) was the first set of standards developed to articulate the knowledge, skills, and dispositions necessary to measure student learning and development outcomes. Specifically, the ASK Standards detail the knowledge and skills that all student affairs educators should have related to assessment, regardless of functional area. The ACPA-NASPA Competencies (American College Personnel Association & National Association of Student Personnel Administrators, 2015) also reflect core competencies expected of student affairs educators regardless of functional area; however, they are much broader than the ASK Standards. One of these ACPA-NASPA competencies is Assessment, Evaluation, and Research (AER), which focuses "...on the ability to design, conduct, critique, and use various AER methodologies and the results obtained from them, to utilize AER processes and their results to inform practice, and to shape the political and ethical climate surrounding AER processes and uses in higher education" (p. 12). The AER competency enhances the ASK Standards by providing developmental levels of personal competency (foundational, intermediate, advanced).

The CAS Standards (2023) contain a set of General Standards that highlight the essential components of quality student affairs programs and services. Thus, these standards are often used for program reviews. One component of the General Standards is Part 4: Assessment. CAS's communication of the aspects of high-quality outcomes assessment (including the newly added elements of program theory and implementation fidelity) relays the expectation that professionals will be able to effectively design, implement, assess, and improve programming in higher education.

Each of the standards explicitly outlines the expectation that student affairs educators enter the workforce with a strong background and skills in program development, program assessment, and use of results for programmatic improvement. Put simply in the words of Wawrzynski et al. (2015), "...assessment should be a familiar term to new student affairs professionals" (p. 121).

In addition to these expectations regarding assessment, student affairs educators are also expected to be knowledgeable with respect to equity and inclusion (West & Henning, 2023). Both the CAS Standards (2023) and the ACPA-NASPA Competencies (2015) incorporate general (i.e., not assessment-specific) competency areas pertaining to equity (i.e., Part 5 of the CAS General Standards: Access, Diversity, Equity, Inclusion, and Justice; ACPA-NASPA Social Justice and Inclusion Competency). Moreover, the ASK Standards (2006) highlight important factors of equity-minded assessment, including reviewing measures for accessibility and data disaggregation.

Additionally, CAS (2023) includes standards related to graduate program curriculum called "Master's Level Higher Education and Student Affairs Professional Preparation Programs."

> Subpart 5b.5 - Assessment, evaluation, and research (CAS, 2023, p. 720)
>
> Program curriculum must include the study of assessment, evaluation, and research that centers on evidence-based practice to further accountability and continuous improvement. Content must include assessment planning and design; outcome development; qualitative, quantitative, mixed methods, and critical data collection and analysis methods; measurement of learning processes and outcomes;

> assessment of environments and organizations; measurement of program and environment effectiveness; effective reporting; critiques of published studies; integration of social justice; and assessment and change management strategies.
>
> Program content must include opportunities for graduates to learn how to critique a study or evaluation and be able to design, conduct, and report on a sound research study, assessment study, or program evaluation, all grounded in the appropriate literature to improve professional practice and student learning.
>
> Content must include information regarding research ethics and legal implications of research, including the necessity of adhering to a human subjects review.

These curricular standards intertwine assessment and equity. Thus, upon reading these "must" statements expected of graduate curriculum, one would believe student affairs educators would have skills in evidence-based programming, outcomes assessment, and use of results for improvement, all with an equity frame. However, in a review of 111 syllabi from assessment/evaluation courses within student affairs graduate programs, a "limited" alignment with professional standards was observed (ACPA et al., 2024). Notably, only 1.8% of syllabi had a primary theme of equity-centered assessment, indicating a clear discrepancy between expectations set by professional standards and current practice in graduate-level programs.

Lack of Assessment Skills Among Student Affairs Educators

Concerns about assessment-related competencies are further fueled by the finding that many student affairs educators enter the workforce lacking assessment skills (Cooper et al., 2016; Hoffman, 2015). For example, in a recent mixed-methods study examining the perceptions of students in a student affairs graduate program, 30.43% of students reported they did not learn how to "evaluate research," a skill needed to inform programming decisions (Wright-Mair et al., 2022). Additionally, 47.83% of respondents indicated they did not learn how to "understand statistics," and 34.78% indicated they did not learn how to "communicate results," both necessary skills to use assessment data for improvement. In a recent survey (Dean & Langham, 2022), 61% of leadership personnel strongly agreed that they expected new hires to have skills in assessment; however, nearly one-third of respondents indicated that new hires were not adequately prepared with these skills.

Lack of Training in Student Affairs Graduate Programs

One reason student affairs educators may lack assessment skills is insufficient training in graduate programs. Research indicates that Higher Education and Student Affairs (HESA) or College Student Personnel Administration (CSPA) programs often fall short in preparing students for the practical demands of assessment-related work (e.g., Marsden & Eckert, 2018). A survey of entry-level student affairs professionals found that only 24% of respondents reported having a required assessment course in their graduate training program (Dean & Langham, 2022). Additionally, respondents noted that their graduate programs provided minimal coverage of the assessment-related knowledge and skills outlined in the professional standards. In a second study (Wright-Mair et al., 2022), students noted that course sequencing in their graduate training contributed to their perceived preparedness to engage in scholarly activity. Specifically, students were often allowed to enroll in courses without having mastered essential foundational research skills,

rendering them incapable of conducting an empirical project. Also, despite the documented benefits of hands-on assessment experiences for learning and valuing assessment (Renn & Jessup-Anger, 2008; Rincón & Castillo-Montoya, 2018), opportunities for such experiences during graduate studies remain limited (Wright-Mair et al., 2022). As Wawrzynski et al. (2015) noted, one of the barriers preventing student affairs educators from conducting assessment is that they are simply not trained to "do" assessment (p. 127). This lack of training is particularly concerning given that first-time, full-time staff from these graduate training programs comprise 15% to 20% of the student affairs workforce (Renn & Jessup-Anger, 2008).

Although resources on assessment are available (e.g., professional standards, Student Affairs Assessment Leaders massive open online course), the integration of these resources into graduate programs appears to be somewhat limited (Dean & Langham, 2022; Rincón & Castillo-Montoya, 2018). Several factors may contribute to this inefficient use of available resources. First, graduate training programs often prioritize theoretical knowledge over practical skills, leaving students with insufficient experience in conducting real-world assessment applications (Renn & Jessup-Anger, 2008). In addition, there is often a lack of emphasis on navigating institutional politics (e.g., graduate students learn about student affairs "best practice," but not about what to do when institutions resist or do not adhere to "best practice"), which can result in a disconnect between what is taught and what is experienced in professional practice (Renn & Jessup-Anger, 2008). Thus, training is insufficient for professionals to meet expectations, leaving student affairs educators underprepared to contribute meaningfully to evidence-based programming, outcomes assessment, and use of assessment results for program improvement.

Need for Professional Development in Equity-Centered Programming and Assessment

Given that CSPA/HESA graduate programs inconsistently train student affairs educators regarding assessment practices, "on the job" professional development is needed to ensure proper training. As noted by Biddix et al. (2020), some professional staff's only experience with assessment comes from "other duties as assigned" in their job description, without any foundation in the field (p. 168). Thus, supplemental training experiences are necessary to bring new-hires up to speed regarding high quality assessment processes. However, supplemental training in assessment and equity have traditionally been siloed (Henning & Lundquist, 2018). Yet, without training on how to integrate equity and assessment, it is unrealistic to expect student affairs educators to routinely engage in equity-centered assessment practice. Equity-centered assessment aims to "…foster equity, address issues of oppression and privilege, improve student learning, and reshape systems and structures influencing the environments in which students learn" (Heiser et al., 2023). Heiser and colleagues (2023) identified six key characteristics of equity-centered assessment: (a) clarity of purpose and goals, (b) epistemological considerations, (c) recognition and reification of power structures, (d) thoughtful methodological choices, (e) centering student voices, and (f) reflexivity and positionality. Designing assessment efforts with these six characteristics requires guidance and practice.

These issues of lack of experience and competency in programming and assessment were ever-present on our campus. For years, our institution struggled with lack of training of student affairs educators in both evidence-based programming and outcomes assessment. Moreover, there was almost a complete lack of awareness of equity-centered programming and assessment. Thus, high-quality professional development experiences were urgently needed to enhance student affairs educators' proficiency and confidence in equity-centered programming and assessment. We believed this need could be best addressed using a meta-assessment frame, which fulfills the division's goals regarding a culture of evidence and continuous improvement. Specifically, the simple presence of a meta-assessment rubric (described below) communicates the institutions' commitment to evidence-based programming, outcomes assessment, and use of results for improvement. Embedding the rubric into any discussions related to programming and impact on student learning reminds professionals that this resource can guide their work. As noted by Biddix et al. (2020), as assessment is more routinely ingrained into student affairs practice, systems should be put in place to handle the increased demand. Thus, we knew that pursuing a meta-assessment process at our institution would require the creation of associated professional development.

Meta-Assessment Rubric to Guide Professional Development

Meta-assessment evaluates the quality of the assessment process itself (e.g., Fulcher & Good, 2013). Emerging from evaluation literature (Ory, 1992), meta-assessment addresses the growing emphasis on outcomes assessment in higher education by explicitly showcasing the aspects of high-quality assessment processes. Implementing a meta-assessment process typically involves creating a rubric to specify high-quality assessment practices. Ratings from meta-assessment rubrics provide evidence of engagement in the assessment process for accreditation purposes and help administrators identify programs with demonstrated impact versus those with uncertain effectiveness (McDonald, 2010). Moreover, meta-assessment rubrics are valuable for continuous improvement, as they can prompt enhancements to the assessment process or to educational programming (Fulcher et al., 2016).

An additional use of meta-assessment rubrics is for professional development. Meta-assessment rubrics clearly communicate assessment expectations and processes, which is especially important when professionals view assessment as unknown, unexpected, or different from what they perceive as student affairs practice (Castillo-Montoya, 2020). Clear, explicit expectations regarding assessment are also essential given the high turnover of student affairs educators. For example, Bichsel et al. (2023) found that 39% of student affairs educators might seek other employment within the next year. When experienced professionals leave, training for newcomers becomes essential. A meta-assessment rubric offers an efficient way to introduce student affairs educators to the assessment process and best practices. Moreover, training can be designed to align directly with specific aspects of the rubric. With this in mind, we set out to design such a training.

Purpose of the Current Study

The purpose of this study was to empirically evaluate the effectiveness of a professional development session to increase skills related to equity-centered programming and assessment. The evaluation was conducted using a Generalizability study (G-study) to assess the impact of the training. We analyzed assessment report ratings provided by participants using a new meta-assessment rubric specifically designed for student affairs educators.

Materials and Methods

Open Education Resources (OER)

There were three available resources that we used to create and evaluate our professional development in equitable assessment practice: a meta-assessment rubric for program improvement, mock reports, and equity resources. Fortunately, all resources were free and easily accessible to us (and anyone) via an OER Commons website (Finney et al., 2024).

The Student Affairs Assessment Improvement Rubric (SAAIR)

The SAAIR, created by a team of student affairs educators and assessment consultants, supports the goal of continuous improvement (Finney et al., 2024). Containing eight criteria directly aligned with the outcomes assessment cycle (Student Learning & Development Outcomes, Program Theory, Outcome Measures, Implementation Fidelity, Collecting Outcomes Data, Data Analysis, Reporting Results, and Use of Results for Improvement), the rubric was built to operationalize what the creators defined as high-quality assessment practice. Notably, the rubric was designed to align with student affairs professional standards (e.g., ACPA-NASPA Competencies, CAS Standards). As a result, it can serve as a pedagogical tool for professionals, addressing the lack of formal training in assessment. Further, the rubric has an explicit focus on equity. That is, while each criterion of the rubric (e.g., Student Learning & Development Outcomes) has varying sub-criteria, there is an ever-present equity sub-criterion. We found the verbiage included in these equity sub-criteria to be clear, allowing us to incorporate and train student affairs educators on equity-centered actions.

Mock Reports

The OER website that houses the SAAIR provides additional support, including three "mock" assessment reports of varying quality. The three reports align directly with the three levels of development showcased in the rubric: developing, proficient, and exemplary. These reports provide examples of assessment write-ups based on hypothetical, but realistic, student affairs programming. We were able to use the SAAIR to rate the mock reports and create associated keys. These keys were later used in our professional development training.

Equity in Assessment Resource

Lastly, the OER website contains a document of examples of equity-centered actions and associated text. The resource aligns with Heiser and colleagues (2023) six key characteristics of equity-centered assessment. Examples of equity considerations were provided for each criterion of the SAAIR at varying levels of quality (developing, proficient,

exemplary). The example text reflected hypothetical but realistic programming within a division of student affairs. We used this resource as one of the main components of our division-wide professional development, described below.

We believed this set of assessment-related supports could efficiently assist our student affairs division in their quality improvement process. Likewise, these resources would expose and advance considerations regarding how to apply an equity lens when engaging in program improvement efforts. Thus, use of these resources should increase the odds that all students benefit from equitable high-impact programming on our college campuses.

Division-Wide Professional Development on Equity-Centered Assessment

To facilitate engagement in equity-centered assessment practice, it was imperative that professionals (a) could distinguish between the various components of the assessment process and how equity could be considered, (b) could discuss examples of equity considerations that ranged from developing to exemplary, (c) could apply the rubric to assessment reports to identify areas of confusion, and (d) could provide colleagues feedback to improve equity-centered assessment. Thus, we designed a rigorous, interactive three-day professional development for student affairs educators on our campus that was directly aligned with these four outcomes.

Participants and Procedures

All student affairs educators were invited to the May 2024 professional development. Ten student affairs educators representing a variety of offices on campus participated in the professional development. Five of the thirteen Student Affairs offices at our institution were represented at the training (38.46%). Specifically, participants represented the following offices: Dean of Students (1 participant), University Recreation (2 participants), Office of Student Accountability and Restorative Practices (3 participants), University Career Center (2 participants), Student Affairs Technical Services (2 participants). These individuals were released from daily activities in the division to focus on building their skills in equity-centered assessment with the goal of applying these new skills in their offices the next academic year.

The three 9-hour days were structured as follows. First, we started with the basics. Given some student affairs educators were new to assessment practice, several hours during Day 1 were used to frame the “what,” “why,” and “how” of outcomes assessment. Specifically, facilitators spent about two hours articulating the professional standards related to outcomes assessment, how assessment moves the division from a culture of good intentions to a culture of evidence (Culp & Dungy, 2012), and why equity must be integrated into assessment practice.

Then, facilitators walked through each criterion of the SAAIR (e.g., writing student learning and development outcomes, collecting implementation fidelity data), a process that took the remainder of Day 1 (about 5 hours) and the first half of Day 2 (about 3 hours). Facilitators first introduced a criterion of the rubric, walking through the various sub-criteria and discussing how each sub-criterion changed across the levels of development

(developing, proficient, exemplary) (about 20 minutes). Facilitators then provided example text that educators rated for quality using the rubric (about 30 minutes). The example text provided was adapted from the equity resource available from the OER website. After educators completed their individual ratings, facilitators demonstrated how they used the rubric to provide ratings of the text for quality (about 30 minutes). Facilitators also modeled how to provide written feedback that both supported these ratings and encouraged continued practice in equity-centered assessment. Throughout this process, student affairs educators and facilitators engaged in discussion regarding discrepancies in ratings, if applicable. Importantly, ratings given by facilitators were not introduced as the "correct" answer, but rather just an example of the facilitator's perspective.

Next, the three mock reports on the OER website were rated by the ten participants. Student affairs educators were provided time to individually rate and provide constructive qualitative feedback during the second half of Day 2 and the entirety of Day 3 of the professional development. Professionals were prepared for this independent rating task, given the training on the rubric that they received on Days 1 and 2. Once independent ratings were provided to facilitators, participants were allowed to discuss their ratings with a partner and come to a conclusive, final rating via an adjudication process.

Generalizability Study (G-Study)

To assess the effectiveness of the three-day professional development, a generalizability study (G-study) was conducted. This analysis provides the consistency of ratings across the participants and reveals whether the variability observed in ratings was due to the quality of the report (which is desired) versus the harshness of the rater (not desired) or sub-criteria (not desired). The following analyses were completed using the individual pre-adjudicated ratings. It was necessary to ensure the independence of the ratings when conducting this type of analysis.

A G-study provides reliability-like coefficients (e.g., G-Coefficient, Phi Coefficient). These coefficients are analogous to coefficients more commonly used in classical test theory, such as Cronbach's coefficient alpha or KR-21 (Shavelson & Webb, 1991). In addition, a G-study partitions the variance of scores into variation due to the object of measurement (in this case, the mock report) as well as due to raters, rubric sub-criteria, their interactions, and error.

For the purpose of this analysis, all three mock assessment reports were rated by all raters on all sub-criteria. In other words, all reports were crossed with all raters and all sub-criteria, a very robust design. Note, ratings from 3 of the 10 raters were not included in the analyses due to not having provided ratings for every sub-criterion for every report. A small number of raters (typically 2 to 5) is often sufficient for G-studies, as increasing the number of raters does not significantly improve results when the rubric is well-defined, and raters are well-trained (Brennan, 2000). In fact, previous research has shown that differences in reliability estimates were negligible across 2, 4, 5, and 8 raters (Monteiro et al., 2019). Although we could have employed a few raters to evaluate the rubric and training, we allowed all participants to serve as raters because the activity of rating was

part of the professional development (i.e., rating and providing feedback builds skills in evaluating assessment processes).

All effects (mock report, rater, sub-criteria, interactions) were treated as random. Treating the rater effect as random means that raters have been theoretically sampled from a universe of possible raters (i.e., universe of possible student affairs educators). Treating the sub-criteria effects as random means that the sub-criteria were sampled from a universe of possible sub-criteria that represent different parts of the assessment process. Treating the mock reports as random means that the mock reports were sampled from a universe of possible mock reports that represent the various levels of development associated with the rubric (i.e., developing, proficient, exemplary).

Key Matching

In addition to the consistency of ratings, we also evaluated how well the scores from the raters matched the scores we generated as the "keys" for each mock report. We conducted a key-matching analysis to uncover the percentage of raters who exactly matched the key, exactly matched or were no more than 0.5 points off from the key, and exactly matched or were no more than 1 point off from the key. Our goal in executing the key matching analysis was to uncover which sub-criteria proved most difficult to rate. We expected a number of raters to not exactly match with the key, as is common in any application of a rubric. In fact, the requirement to adjudicate ratings with a partner was necessary given expected deviations of some raters from the key. Our goal when evaluating the match between the average independent ratings and the key was to evaluate if the training resulted in about 70% of the ratings being within 0.5 point of the key and over 80% being within 1 point of the key. In turn, there would be a small percentage of ratings (20% to 30%) that would likely be adjusted to better align with the key during the adjudication process. Moreover, by examining the match between the average individual ratings and the key, we could identify areas of confusion that necessitated further professional development.

Results

G-Study

The reliability-like coefficient from the G-study is the Phi Coefficient. Phi ranges from 0 to 1. Our G-study resulted in a Phi Coefficient equal to 0.97, which is very high.

The program (mock report) variance (true score variance) tells us how much program report scores systematically differed from one another. We wanted this value to be high (see Table 1). The program report variance accounted for 37.2% of the total variance in scores, which is moderately high. Further, we wanted the variance due to raters to be low. The results indicated that this was the case for our analysis, with rater variance being only a mere 0.4%. That is, certain raters were not consistently more harsh or lenient in their ratings compared to other raters, which was a welcomed finding of our training.

There was some variance due to rubric sub-criterion (19.6%). This variability was not surprising, given that some sub-criteria (i.e., equity-centered sub-criteria) were typically lower than others, regardless of mock report or rater.

Table 1. *Variance Components from G-Study with Interpretations*

Component (GENOVA Symbol)	Variance	Percent of Variance	Interpretation
P	0.533	37.2%	Variance due to different quality of program reports. Desirable for this percentage of variance to be high. Interpretation: 37.2% of the variance in scores was due to true score variance. Reports consistently rank ordered due to quality (exemplary, proficient, and developing), which is desirable.
R	0.006	0.4%	Variance due to rater harshness. Desirable for this percentage of variance to be low. Interpretation: only 0.4% of the variance was due to rater. It does not appear that some raters were consistently more harsh than other raters.
C	0.282	19.6%	Variance due to differences in scores depending on sub-criteria. Desirable for this percentage of variance to be low. Interpretation: 19.6% of the variance was due to sub-criterion. Some sub-criteria were rated lower or higher, on average, than other sub-criteria, regardless of rater or report.
PR	0.004	0.3%	Variance due to the interaction between program report and rater. Desirable for this percentage of variance to be low. Interpretation: only 0.3% of the variance was due to the interaction between report and rater.
PC	0.125	8.7%	Variance due to the interaction between program report and sub-criteria (i.e., do report scores change depending on the criteria). Desirable for this percentage of variance to be low. Interpretation: 8.7% of the variance in scores was due to different reports having different scores for different criteria.
RC	-0.033	0%	Variance due to the interaction between raters and sub-criteria. Desirable for this percentage of variance to be low (i.e., raters are not particularly harsh depending on the criteria). Interpretation: 0% of the variation in scores was due to different raters rating sub-criteria differently.
PRC	0.518	36.1%	Error

Note. Analyses were conducted using GENOVA Version 3.1, which uses Ordinary Least Squares estimation rather than Maximum Likelihood estimation. Therefore, variances that are negative and small can be interpreted as 0.

Key Matching

Overall, average ratings aligned with the key created for each mock report, indicating that raters were generally well-calibrated. Table 2 shows the percentage of ratings that (a) match the key perfectly, (b) match the key or are at most 0.5 points off from the key, and (c) match the key or are at most 1 point off from the key. Despite some variability across the 3 mock reports, about half of the raters (56%, 58%, and 54%) exactly matched the exemplary, proficient, and developing keys (respectively). A higher percentage of raters (67%, 71%, and 68%) were within 0.5 of the keys. Finally, most raters (82%, 87%, and 91%) were within 1 point of the keys.

Table 2. *Key Matching Results*

Rater	Prime for Life (Exemplary)	Peer 2 Peer (Proficient)	Service Learning (Developing)	Average Rater Match with Key
		Exact Match		
1	61%	61%	55%	59%
2	58%	36%	45%	46%
3	45%	55%	48%	49%
4	58%	61%	55%	58%
5	67%	64%	58%	63%
6	58%	70%	58%	62%
7	48%	64%	58%	57%
Average Match	56%	58%	54%	
		Exact or Within 0.5 Match		
1	70%	73%	76%	73%
2	64%	58%	61%	61%
3	64%	64%	61%	63%
4	67%	73%	67%	69%
5	76%	85%	70%	77%
6	73%	79%	67%	73%
7	55%	70%	73%	66%
Average Match	67%	71%	68%	
		Exact, Within 0.5, or Within 1 Match		
1	76%	85%	94%	85%
2	88%	79%	79%	82%
3	85%	82%	94%	87%
4	79%	82%	88%	83%
5	91%	97%	91%	93%
6	88%	97%	97%	94%
7	70%	85%	94%	83%
Average Match	82%	87%	91%	

Note. Percentages were concatenated across rubric sub-criteria. For example, in the Exact Match section, 61% of Rater 1's sub-criteria ratings for the exemplary report matched the key exactly. On average, across raters, 56% of ratings exactly matched the key for the exemplary report. Across mock reports, Rater 1 matched the key exactly 59% of the time.

Figure 1 shows the average ratings across raters for the exemplary mock report by sub-criterion, along with the key. In general, there were many instances where the key and average participant ratings aligned. However, large (>1 point) discrepancies occurred for the following sub-criteria: malleable, measure development, use of results, comparison to previous findings, and assessment improvement. In addition, two equity-centered sub-criteria (equity concerns regarding programming; equity concerns related to data analysis) deviated from the key by more than 1 point, on average. Despite discrepancies, the pattern of ratings generally aligned with the exemplary key, with 20 sub-criteria having average ratings within 0.5 of the key.

Figure 1. *Exemplary Report: Average Ratings with Key*

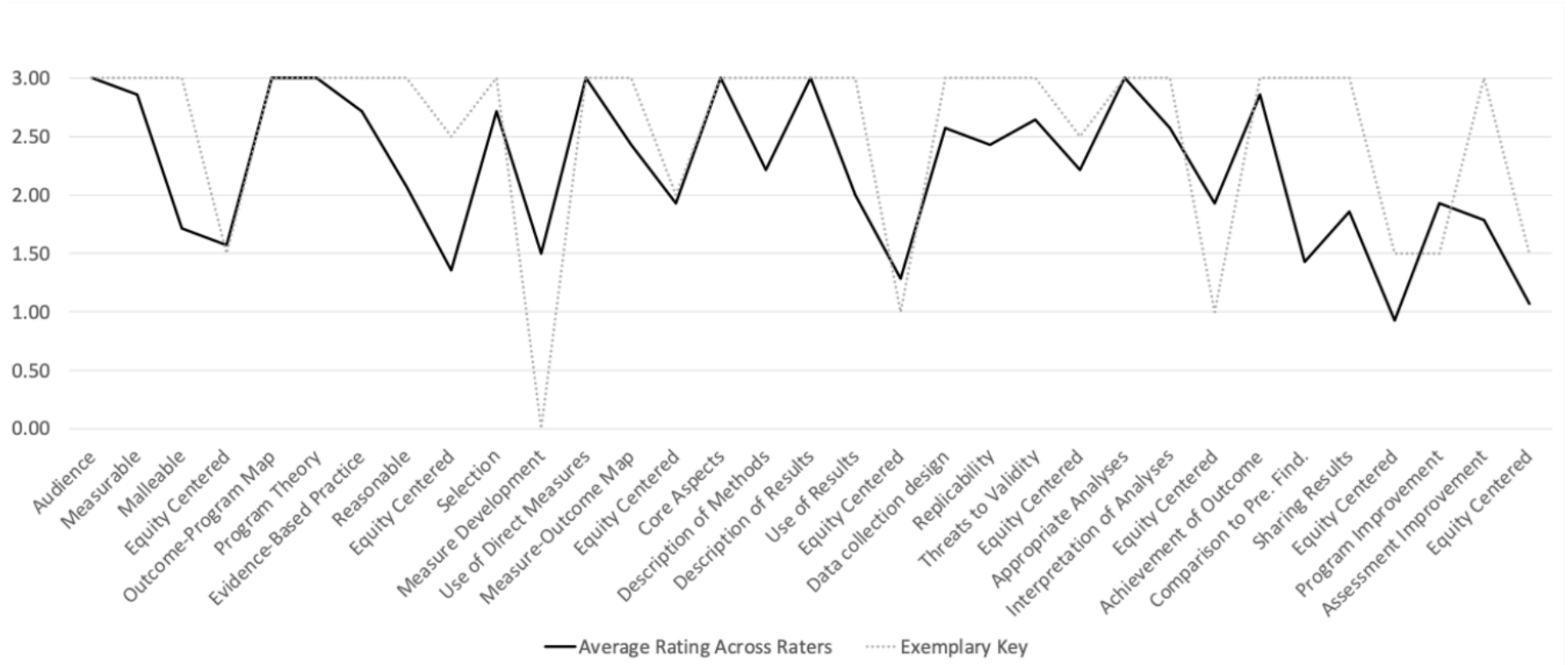
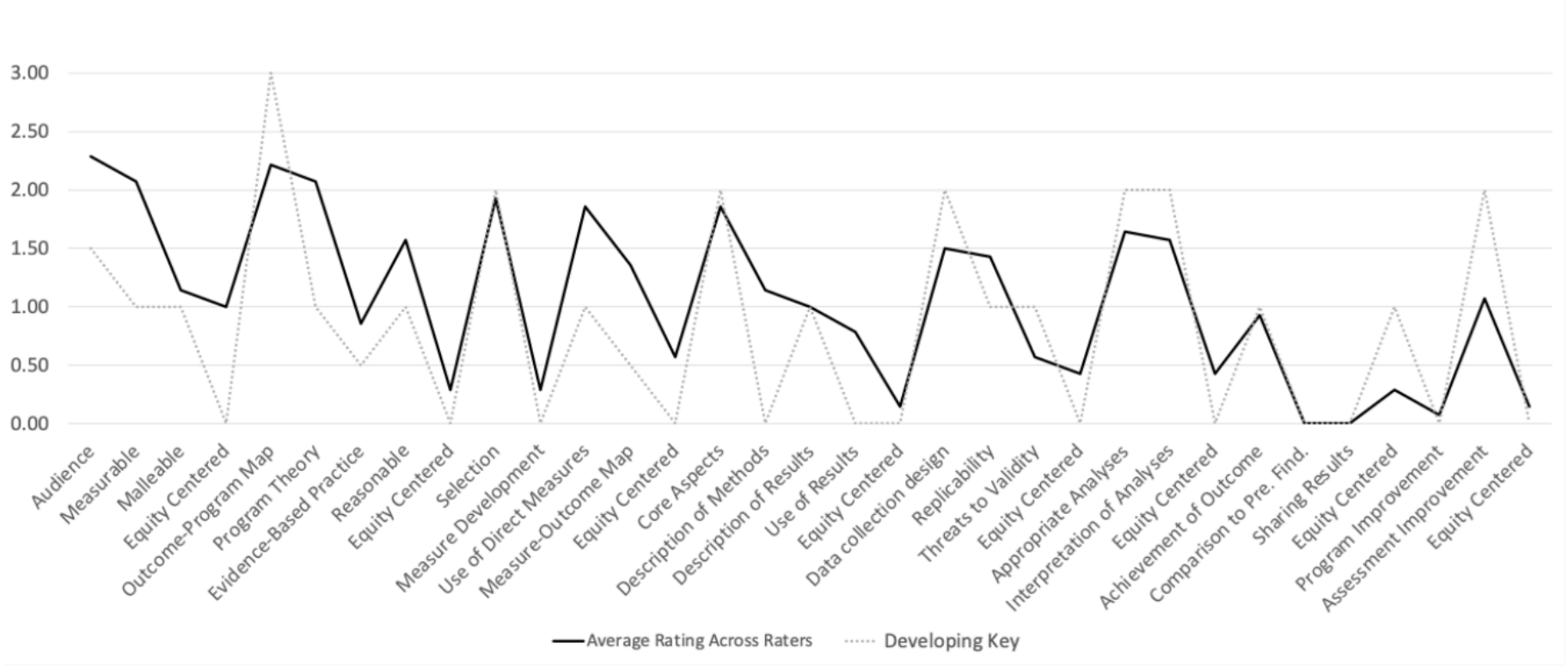


Figure 2 shows the average proficient mock report ratings by sub-criterion, along with the key. Overall, average participant ratings aligned with the proficient key. However, large (>1 point) discrepancies occurred for the measure development and program improvement sub-criteria. Ratings across all other sub-criteria were within 1 point of the key, and 24 sub-criteria had average ratings within 0.5 of the key.

Figure 2. *Proficient Report: Average Ratings with Key*

Figure 3 shows the average developing mock report ratings by sub-criterion, along with the key. The match to the key for the equity-centered sub-criteria was satisfactory, although some instances of non-negligible (i.e., >1 point) discrepancies are present for the following sub-criteria: measurable, the first equity-centered sub-criterion (equity concerns related to student learning and development outcomes), program theory, use of direct measures, and description of methods. Ratings across all other sub-criteria were within 1 point of the key, and 16 sub-criteria had average ratings within 0.5 of the key.

Figure 3. *Developing Report: Average Ratings with Key*



Across all mock reports, there were no sub-criteria that consistently deviated from the key by more than one point. However, the average ratings for the measure development sub-criterion were off by more than 1 point for two of the mock report keys (exemplary and proficient), indicating a need for stronger training on measurement development in future professional development workshops. In addition, all other sub-criterion with non-negligible deviations from their respective keys will be investigated to determine how to minimize these discrepancies in the future, resulting in better calibrations between participants' ratings and keys.

Discussion

We used a newly developed meta-assessment rubric, mock reports, and equity resources to advance the professional development of student affairs educators in equitable programming and assessment. Results of our G-study and key-matching analysis indicated that student affairs educators can distinguish between quality levels of equitable programming and assessment practices. The high consistency of ratings across student affairs educators supports the clarity and utility of the rubric, along with its support materials (e.g., mock reports, equity-centered resources).

Addressing Lack of Training in Assessment via Professional Development

Our training implementing these free OER materials effectively addressed a need on our campus for professional development in equity-centered programming and assessment. In addition to being effective, the accessibility of the online resources enabled us to create the training in just a few months. We found that three days of training was ample to articulate the need and scope of equity-centered assessment and engage in hands-on practice of using the rubric to identify varying quality of programming and assessment processes.

Although the OER resources are free to download, we recognize that there are costs associated with implementing a professional development training (e.g., reserving space to hold the training, providing lunch for participants, hiring experts to facilitate the training if necessary). Despite these costs, we enthusiastically encourage others to apply these resources on their campus. Moreover, our team believes the resources employed in our professional development training are versatile and can effectively support applications beyond traditional in-person full-day professional development contexts. For example, professionals could develop online materials aligned with the OER resources, thereby enhancing accessibility for individuals unable to afford, attend, or implement professional development training in person. Moreover, the training could be spread over a semester or a year (e.g., a couple hours every first Friday of the month) if full days cannot be allotted. In short, although we created and evaluated a three-day, in-person, bootcamp style of professional development, we believe the resources support adapting this training to be virtual and distributed over time.

Further, it may be that some institutions do not use the resources to offer a formal professional development opportunity for the division. Instead, consider models such as the Multilevel Assessment Process (MAP; Strine-Patterson, 2022), where departmental or

divisional leaders could utilize the resources to facilitate strategic planning endeavors. For example, once a student affairs division has identified equity-centered assessment as a priority (which guides the division's purpose and improvement efforts), leadership personnel could create priority assessment teams who serve to explicate and ultimately achieve assessment-related goals aligned with the priority (Strine-Patterson, 2022). These smaller assessment teams could use the OER resources to support assessment efforts at any step of their process.

Following the logic of MAP, leadership should be included in the training focused on equity-centered assessment, so they understand its scope and the resources necessary to support such work. Although division-level leaders (e.g. AVPs, VP) did not participate in our professional development, we strongly encourage others implementing this type of training to include these leaders. Upper administrators play a crucial role in shaping how assessment is perceived and valued by their staff. Thus, upper administrators should reflect on how their discussions and practices in assessment impact their staff (Castillo-Montoya, 2020). Outcomes assessment is often framed as a way to address accountability mandates (e.g., Bresciani, 2011), which may be a "turn-off." Perceiving assessment as a tool to improve programming offered to students and to promote social justice on campus may prompt student affairs educators to regard assessment as a high-impact institutional practice, in turn decreasing resistance or lack of value that some student affairs educators perceive (Henning & Lundquist, 2022; Seagraves & Dean, 2010). This meta-assessment rubric and training could efficiently communicate that messaging to a division, especially if leadership experiences the training.

Addressing Inconsistent Training in Graduate Programs

In addition to the rubric and support resources being used to (re)train current student affairs educators, we argue that HESA or CSPA graduate programs could incorporate these materials into formal training to address the inconsistency in coursework across graduate programs (Dean & Langham, 2022). We believe that courses focused on designing programming could greatly benefit from the rubric and its focus on articulating program theory (Finney & Buchanan, 2021; Pope et al., 2019, 2023) and implementation fidelity (Finney et al., 2021; Fisher et al., 2014; Smith et al., 2019) with equity implications overtly showcased. We envision courses focused on outcomes assessment, data-based decision making, or use of results for improvement could greatly benefit from demonstrating the various levels of high-quality assessment practice (exemplary, proficient, developing), which directly informs claims one can make about the effectiveness of programming.

We envision a couple of ways that instructors could use the resources. First, the three-day training could be directly incorporated into a class. Of course, the training would need to be distributed across class meetings, the ratings of reports could be completed as assignments (in class or outside of class), and this would not consume the semester-long course but rather a portion of the course even if implemented fully. Second, instructors may decide to simply use the rubric to frame the assessment process and use the mock reports to showcase high-quality practice rather than using the resources to formally train students to evaluate assessment practice. This second option has been used on our campus in two courses.

Moreover, the explicit link between the rubric and student affairs professional standards (ACPA-ASK Standards, ACPA-NASPA Competencies, CAS Standards) exposes students to the expectations from their profession. The rubric provides overt "moves" a student affairs educator could make to meet the standards related to assessment, programming, and equity. We imagine this will be empowering to graduate students, as they realize they are developing the competencies that will allow them to meet the standards of the field. This feeling would counter current perceptions of being unprepared in these domains (Wright-Mair et al., 2022). The rubric can help HESA instructors align their courses with national standards, something HESA faculty noted was important when constructing their formal courses (Hunter, 2024).

Conclusion

As the field works to professionalize assessment in student affairs, we need to continue to create and rigorously evaluate training materials for assessment-related professional development opportunities. Moreover, we believe these materials can and should purposefully integrate expected competencies (e.g., assessment and equity). Further, we believe the greatest positive impact on the development of student affairs educators is to share effective professional development materials freely so individuals can return to them as needed. It is our hope that future implementations of these materials expand their effectiveness, and that any new resources are shared freely. Increasing the value and engagement in equitable programming and assessment takes practice. Let's commit to offering high-quality experiences to our colleagues, new and seasoned, in order to move from a "culture of good intentions" toward a "culture of evidence" (Culp & Dungy, 2012).

References

ACPA Commission for Assessment and Evaluation, NASPA Assessment Evaluation and Research Knowledge Community, & Student Affairs Assessment Leaders. (March 2024). *What's next in student affairs assessment? Insights from current graduate education practices.* https://myacpa.org/wp-content/uploads/2024/03/Next_for_SA_Assessment_2024_tagged.pdf

American College Personnel Association. (2006). *ASK standards: Assessment skills and knowledge content standards for student affairs practitioners and scholars.* Author.

American College Personnel Association, & National Association of Student Personnel Administrators (2015). *ACPA/NASPA professional competency areas for student affairs educators*. Authors.

Bichsel, J., Fuesting, M., Tubbs, D., & Schneider, J. (2023, September). *The CUPA-HR 2023 higher education employee retention survey*. College and University Professional Association for Human Resources. https://www.cupahr.org/surveys/research-briefs/higher-ed-employee-retention-survey-findings-september-2023/

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*(2), 157–171. https://dx.doi.org/10.1353/csj.2020.0011

Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement, 24*(4), 339–353. https://doi.org/10.1177/01466210022031796

Bresciani, M. J. (2011). *Making assessment meaningful: What new student affairs professionals and those new to assessment need to know* (Assessment Brief). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Castillo-Montoya, M. (2020). Assessment: The "Wild Card" in student affairs. *Journal of Student Affairs Research and Practice*, *58*(1), 79–93. https://doi.org/10.1080/19496591.2019.1707090

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*(2), 107–117. https://doi.org/10.1080/19496591.2016.1121146

Council for the Advancement of Standards in Higher Education. (2023). *CAS professional standards for higher education* (11th ed.). Author.

Culp, M., & Dungy, G. (Eds.). (2012). *Building a culture of evidence in student affairs: A guide for leaders and practitioners.* NASPA Student Affairs Administrators in Higher Education.

Dean, L. A., & Langham, A. T. (2022). Graduate preparation in student affairs assessment. *New Directions for Student Services, 2022*(178–179), 43–51. https://doi.org/10.1002/ss.20427

Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension* (NILOA Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

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/rpa-archives/

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*(3), 310–325. https://www.tandfonline.com/doi/full/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). Charles C. Thomas Publisher. https://psycnet.apa.org/record/2019-16536-000

Finney, S. J., Stewart, J. P., Wild, A. N., Herr, R. K., & Schaefer, K. E. (2024). A meta-assessment rubric to guide professional development and practice in equitable outcomes assessment. *Research & Practice in Assessment, 19*(1), 45–58. https://www.rpajournal.com/dev/wp-content/uploads/2024/09/Meta-Assessment-Rubric-Equitable-Outcomes-Assessment-RPA.pdf

Finney, S. J., Wells, J., & Henning, G. (2021). *The need for program theory and implementation fidelity in assessment practice and standards* (Occasional Paper No. 51). National Institute for Learning Outcomes Assessment (NILOA). https://www.learningoutcomesassessment.org/wp-content/uploads/2021/03/Occ_Paper_51-1.pdf

Finney, S., Wild, A., & Stewart, J. (2024). *Fostering equity in outcomes assessment: A rubric, mock reports, and training support materials*. Open Educational Resource Commons. Retrieved on November 11, 2024, from https://oercommons.org/courseware/lesson/112329

Fisher, R., Smith, K., Finney, S., & Pinder, K. (2014). The importance of implementation fidelity data for evaluating program effectiveness. *About Campus, 19*(5), 28–32. https://doi.org/10.1002/abc.21171

Fulcher, K., Coleman, C., & Sundre, D. (2016). Twelve tips: Building high-quality assessment through peer review. *Assessment Update, 28*(4), 1–2, 14–16. https://doi.org/10.1002/au.30062

Fulcher, K., & Good, M. (2013, November). *The surprisingly useful practice of meta-assessment.* National Institute for Learning Outcomes Assessment (NILOA). https://www.learningoutcomesassessment.org/wp-content/uploads/2019/08/Viewpoint-FulcherGood.pdf

Heiser, C., Schnelle, T., & Tullier, S. (2023). Defining and operationalizing equity-centered assessment. *Journal of Student Affairs Inquiry, Improvement, and Impact, 6*(1), 4–17. https://doi.org/10.18060/27921

Henning, G., & Lundquist, A. (2018). *Moving towards socially just assessment* (Equity Response). National Institute for Learning Outcomes Assessment (NILOA). https://www.learningoutcomesassessment.org/wp-content/uploads/2019/08/EquityResponse-HenningLundquist.pdf

Henning, G., & Lundquist, A. (2022). Using assessment to advance equity. *New Directions for Student Services, 2022*(178–179), 185–194. https://doi.org/10.1002/ss.20439

Hoffman, J. (2015). Perceptions of assessment competency among new student affairs professionals. *Research & Practice in Assessment, 10*(2), 46–62. https://www.rpajournal.com/perceptions-of-assessment-competency-among-new-student-affairs-professionals/

Hunter, G. F. (2024). HESA faculty members as academic planners: Perspectives on course planning. *Journal of Student Affairs Research and Practice*, *61*(4), 561–576. https://doi.org/10.1080/19496591.2023.2278770

Marsden, C. L. W., & Eckert, E. (2018). Research and assessment: A study of the competencies taught and competencies sought. *The Journal of Student Affairs Inquiry, 4*(1). https://doi.org/10.18060/27913

McDonald, B. (2010). Improving learning through meta assessment. *Active Learning in Higher Education, 11*(2), 119–129. https://journals.sagepub.com/doi/10.1177/1469787410365651

Monteiro, S., Sullivan, G. M., & Chan, T. M. (2019). Generalizability theory made simple(r): An introductory primer to G-studies. *Journal of Graduate Medical Education, 11*(4), 365–370. https://doi.org/10.4300/JGME-D-19-00464.1

Ory, J. (1992). Meta-assessment: Evaluating assessment activities. *Research in Higher Education, 33*(4), 467–481. https://www.jstor.org/stable/40196044

Pope, A. M., Finney, S. J., & Bare, A. K. (2019). The essential role of program theory: Fostering theory-driven practice and high-quality outcomes assessment in student affairs. *Research & Practice in Assessment*, *14*(1), 5–17. https://www.rpajournal.com/rpa-archives/

Pope, A. M., Finney, S. J., & Crewe, M. (2023). Evaluating the effectiveness of an academic success program: Showcasing the importance of theory to practice. *Journal of Student Affairs Inquiry*, *6*(1), 35–50. http://doi.org/10.18060/27924

Renn, K. A., & Jessup-Anger, E. R. (2008). Preparing new professionals: Lessons for graduate preparation programs from the National Study of New Professionals in Student Affairs. *Journal of College Student Development, 49*(4), 319–335. https://doi.org/10.1353/csd.0.002

Rincón, B., & Castillo-Montoya, M. (2018). Learning assessment in student affairs through service-learning. *Research & Practice in Assessment*, *13*(1), 39–50. https://www.rpajournal.com/learning-assessment-in-student-affairs-through-service-learning/

Seagraves, B., & Dean, L. A. (2010). Conditions supporting a culture of assessment in student affairs divisions at small colleges and universities. *Journal of Student Affairs Research and Practice, 47*(3), 307–324. https://doi.org/10.2202/1949-6605.6073

Shavelson, R., & Webb, N. (1991). *Generalizability theory: A primer.* Sage.

Smith, K., Finney, S., & Fulcher, K. (2019). Connecting assessment practices with curricula and pedagogy via implementation fidelity data. *Assessment and Evaluation in Higher Education*, *44*(2), 263–282. https://doi.org/10.1080/02602938.2018.1496321

Strine-Patterson, H. J. (2022). Assessment is a leadership process: The multilevel assessment process. *New Directions for Student Services*, *2022*(178–179), 61–76. https://doi.org/10.1002/ss.20429

Wawrzynski, M. R., Brock, A., & Sweeney, A. (2015). Assessment in student affairs practice. In M. J. Amey & L. M. Reesor (Eds.), *Beginning your journey: A guide for new professionals in student affairs* (pp. 121–142). National Association of the Student Personnel Administrators.

West, A., & Henning, G. (2023). Global implications of student affairs competencies and standards. *New Directions for Student Services, 2023*(183), 31–39. https://doi.org/10.1002/ss.20476

Wright-Mair, R., Saadeddine, R., Peters, C., & Elmes, A. (2022). Higher education and student affairs master's students' perceptions of their preparation for scholarly practice and implications for program improvement: A mixed methods case study. *College Student Affairs Journal*, *40*(2), 13–32.

How to cite this article: Wild, A. N., Herr, R. K., & Finney, S. J. (2024). Professional development in equity-centered outcomes assessment: The utility of a meta-assessment rubric. *Journal of Student Affairs Inquiry, Improvement, and Impact*, *7*(1), 101–120. https://doi.org/10.18060/28702