

Original Publication

 OPEN ACCESS

Collecting Validity Evidence: A Hands-on Workshop for Medical Education Assessment Instruments

Caroline R. Paul, MD*, Michael S. Ryan, MD, MEHP, Gary L. Beck Dallaghan, PhD, Thanakorn Jirasevijinda, MD, Patricia D. Quigley, MD, MME, EdS, Janice L. Hanson, PhD, Amal M. Khidir, MD, Jean Petershack, MD, Joseph Jackson, MD, Linda Tewksbury, MD, Mary Esther M. Rocha, MD

*Corresponding author: crpaul@wisc.edu

Citation: Paul CR, Ryan MS, Beck Dallaghan GL, et al. Collecting validity evidence: a hands-on workshop for medical education assessment instruments. *MedEdPORTAL*. 2019;15:10817.

https://doi.org/10.15766/mep_2374-8265.10817

Copyright: © 2019 Paul et al. This is an open-access publication distributed under the terms of the Creative Commons Attribution-NonCommercial-Share Alike license.

Abstract

Introduction: There is an increasing call for developing validity evidence in medical education assessment. The literature lacks a practical resource regarding an actual development process. Our workshop teaches how to apply principles of validity evidence to existing assessment instruments and how to develop new instruments that will yield valid data. **Methods:** The literature, consensus findings of curricula and content experts, and principles of adult learning guided the content and methodology of the workshop. The workshop underwent stringent peer review prior to presentation at one international and three national academic conferences. In the interactive workshop, selected domains of validity evidence were taught with sequential cycles of didactics, demonstration, and deliberate practice with facilitated feedback. An exercise guide steered participants through a stepwise approach. Using Likert-scale items and open-response questions, an evaluation form rated the workshop's effectiveness, captured details of how learners reached the objectives, and determined participants' plans for future work. **Results:** The workshop demonstrated generalizability with successful implementation in diverse settings. Sixty-five learners, the majority being clinician-educators, completed evaluations. Learners rated the workshop favorably for each prompt. Qualitative comments corroborated the workshop's effectiveness. The active application and facilitated feedback components allowed learners to reflect in real time as to how they were meeting a particular objective. **Discussion:** This feasible and practical educational intervention fills a literature gap by showing the medical educator how to apply validity evidence to both existing and in-development assessment instruments. Thus, it holds the potential to significantly impact learner and, subsequently, patient outcomes.

Keywords

Evaluation, Editor's Choice, Assessment, Curriculum Development, Validity Evidence, Quantitative Research

Educational Objectives

By the end of this activity, learners will be able to:

1. Apply recognized frameworks for gathering validity evidence to medical education instruments in a stepwise manner.
2. Begin to analyze evidence of validity for existing medical education instruments.
3. Develop roadmaps for domains of validity for an instrument of their choosing.

Introduction

Validity evidence has been well described in education as crucial to the development of scholarly teaching interventions and assessments.¹⁻³ However, medical education has lagged behind other educational professions in vetting the rigor of assessment methods. As a result, there has been a recent recognition of the need for validity evidence specifically in medical education.⁴ Academic journals increasingly require validity evidence in assessment instruments utilized in pertinent studies.^{2,4}

Appendices

- A. 3-Hour Session Facilitator Guide.docx
- B. 4-Hour Session Facilitator Guide.docx
- C. Learner Exercise Guide .docx
- D. Instructor Exercise Guide .docx
- E. Evaluation Form.doc
- F. Assessment Instrument Didactic Slides.pptx
- G. Glossary of Terms.docx

All appendices are peer reviewed as integral parts of the Original Publication.

With the increased scholarly focus on assessment and evaluation, there is a call for validity evidence with the instruments used to assess learners at all levels of medical education.⁵ Validity evidence can be considered central to the tiers of Kirkpatrick's hierarchy of evaluation as one moves from participant feedback to eventual impact on patient outcomes.⁶

As educators recognize the need for validity evidence, studies of instruments with validity evidence have emerged in the medical literature.⁷⁻¹¹ These works are varied and traverse specialties, exemplifying the need and opportunity for the demonstration of validity evidence in medical education in all clinical areas. For example, a study on validity evidence regarding a checklist for pediatric otoscopy has shown how an evaluation instrument that contains proven validity evidence can be used in direct patient care settings.⁸ Overall, there is greater emphasis in medical education itself on the need for validity evidence as stakes of assessment and evaluation have greater impact on the career potential of medical students and as the importance of summative and formative feedback is increasingly recognized.^{12,13}

However, the literature is lacking in descriptions of how to apply principles of validity evidence to medical education assessment instruments. There are no practical resources in the cited literature regarding either the actual process of developing instruments with validity evidence or how to apply validity evidence principles to existing medical education assessment instruments.

Medical educators need to learn the basic principles of validity evidence and how to apply these principles and core tenets to the application of their work as clinical educators. For them to learn how to do this, a practical and feasible approach needs to be developed.

Using peer-reviewed models of validity evidence, we have developed a practical model to apply principles of validity evidence to existing medical education assessment instruments and to the development of such instruments. Our model teaches medical educators the core tenets of validity evidence while providing a practical approach on how to apply these tenets to their educational activities.

Methods

Curriculum Context and Target Audience

After conducting a needs assessment of educators in a working-group manner at the Council on Medical Student Education in Pediatrics (COMSEP) Annual Meeting in 2013, we recognized the need to develop this workshop. We based the content, methodology, and mode of delivery on literature findings, on consensus findings of pediatric medical educators, and on principles of adult learning.

We did not require prior knowledge or skills of our learners. We did inform learners that the workshop was a practical one. We sought to attract participants with an interest in developing and applying validity evidence to outcome measures and instruments. Learners would gain maximum benefit if they were able to bring an instrument in any stage of development to the workshop. In the meeting brochures, we asked attendees to bring an existing instrument example or an idea to develop into an instrument that would contain validity evidence. Thus, the hands-on, practical approach of this curriculum started with the workshop brochure description itself.

We chose facilitators to lead the workshop. All facilitators were medical educators who had some experience in assessment and in teaching adult learners. Most did not have formal education in validity evidence but were able to deliver the workshop after an orientation process. The facilitators participated in a detailed orientation regarding the overall content and instructional techniques of the workshop during two 1-hour conference-call meetings. The facilitators used facilitator guides (Appendices A and B) during the orientation meetings and while leading the workshop. Individual facilitator guides were developed for 3-hour and 4-hour workshop time slots, which were needed to adapt to different conference-scheduling requirements. Both workshops contained the same content, but the 4-hour workshop provided more time for individual work, large-group sessions, and breaks.

Ideal facilitators for the workshop would be medical educators with moderate experience in assessment and in teaching adult learners. However, educators with more modest backgrounds could also serve as facilitators. Those who presented the original workshops included individuals with significant expertise as well as those who were novices. For the original presentations, we provided the aforementioned orientation regarding the overall content and instructional techniques of the workshop. Included in this publication are facilitator guides (Appendices A and B) that mirror the content of the orientation sessions provided to our original facilitators.

Implementation

We used peer-reviewed and evidence-based findings from the literature and expert content sources to guide the content of the workshop. We based all instructional methods and strategies on standard curriculum processes. All learning principles were introduced in a brief didactic session and then were demonstrated to the learners using an example from the literature. The learners were given time to apply the intended principle to their own assessment instruments, with workshop facilitators available to offer guidance and feedback.

Framework

We delivered the highly interactive workshop during either a 3-hour or a 4-hour session, as illustrated in the facilitator guides (Appendices A and B). Using adult learning theory, we designed the workshop with built-in opportunities for individualized work and with group report-out activities both to designated partners, using the think-pair-share approach, and to the entire learner group.

We started the workshop with an introduction of the facilitators. We deliberately assigned the instructors to tables for facilitated work with the learners. We then performed a needs assessment to explore learners' needs and perceived level of experience in validity evidence.

The teaching session began with a brief didactic presenting an overview of validity, using many examples to demonstrate key principles. The content was largely based on the *Standards for Educational and Psychological Testing*.¹ While newer frameworks for establishing validity evidence of educational assessments, such as the method described by Kane,¹⁴ have been introduced since Messick first proposed standards of construct validity, an ideal gold-standard framework has not yet been firmly established. We chose the traditional methods of the *Standards* because they continue to be applied in a wide range of validation studies recently published in medical education.

Using a learner exercise guide (Appendix C) we developed, we quickly engaged the learners in beginning their work. Our first exercise asked learners to work independently on their own idea or an existing instrument that they intended to use to gather validity evidence.

We introduced and taught other selected domains of validity evidence, including content, response process, relation to other variables, and consequences. Each domain was taught using the following framework:

1. Brief didactic session that included definitions and examples from the literature to simplify the concept, followed by an exercise.
2. Demonstration with examples.
3. Practice opportunities described in the learner exercise guide (Appendix C) to help guide learners' work in a stepwise manner: Learners answered the prompts in the exercise guide as appropriate to their idea or instrument and their level of progression.

We developed a PowerPoint slide set (Appendix F) that contained an introduction, all of the teaching content and speaker notes for the didactic sessions, and instructions for each exercise that the learners performed after receiving each didactic. We used the PowerPoint slide set in conjunction with the learner exercise guide (Appendix C) throughout the workshop. The PowerPoint slide set contained several

teaching examples and a reference to a published instrument⁸ that could be used as an example throughout the workshop.

Preparation of the workshop space included the following equipment:

1. Round tables for four to six participants.
2. Audiovisual support, including an LCD projector.
3. Two flip charts at the front of the room.
4. A table containing a printed learning exercise guide for each learner.

As noted previously, we invited participants to bring an idea, an instrument in any stage that they were developing, or an already existing instrument to further develop with evidence of validity. We ensured that there were enough facilitators to establish a good ratio of instructors to learners at each table. To stay on schedule, we designated one workshop leader to serve as the timekeeper.

Learner Exercise Guide, Instructor Exercise Guide, and Facilitator Guide

We provided materials formatted for a 3-hour or a 4-hour workshop. Alternatively, the materials could be broken up into four 1-hour workshops. However, for such a version, we recommend that participants commit to attending all of the shorter workshops so that they can continue to build on the work of their chosen instruments.

In addition to the learner exercise guide described above, we developed an instructor exercise guide (Appendix D). This guide contained the learner exercise guide plus instructions and specific teaching guidance for the instructors. These guides were developed over a 6-month period and underwent several revisions. They were pilot tested at the lead author's institution. We also developed the facilitator guides (Appendices A and B) to provide direction to the instructors leading the workshop.

The guides and references included in this publication were designed to provide typical medical educators with the background needed to meaningfully serve as facilitators. We encourage facilitators to engage in self-study of the materials, including the references, and follow up with group discussions to ensure that all facilitators have similar and sufficient understanding prior to the implementation of the workshop. It would also be helpful to have at least one educator with more advanced training in educational assessment to both facilitate at the workshop and participate in preworkshop facilitator orientation and discussions. Workshop facilitators could each present one topic (e.g., construct, content, or response process). This approach would work well with facilitators who have less expertise in validity evidence as they could focus their learning about the topic and build their expertise.

The learner exercise guide (Appendix C) steered learners step by step through the development of an assessment instrument with validity evidence and the application of validity evidence to existing instruments, directly linking all practical work to the workshop's education objectives. The guide aimed to meet a fundamental curriculum principle that all learning strategies, including demonstration of principles and practice opportunities for learners, should link to the educational objectives. The guide was divided into sections. It began with a didactic section highlighting key section points. Learners then actively used the guide. Further adhering to curriculum process tenets, each phase of the guide stressed the particular learning point and provided learners the opportunity to receive a demonstration and then critically apply the learning point to their actual practical work. Thus, immediately following the demonstration of a particular content point, learners essentially received time to practice and apply the point to their own work. Facilitated expert feedback from the instructors followed the practice time. This process not only allowed for effective learner strategies but also provided an opportunity for learners and instructors to evaluate whether the learners were meeting the workshop's education objectives. Designed to balance appropriate white space, visuals, texts, and terminology to allow for optimal practical work during the session, the guide also contained a glossary to use during instruction time (Appendix G).

Evaluation Form

An anonymous paper evaluation form (Appendix E) was distributed in person at the conclusion of the workshop. The evaluation form used a 5-point Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*) to rate the overall effectiveness of the workshop and the degree to which the objectives had been met. We developed the form to address each educational objective and to meet the unique needs of our learners, who were medical educators themselves. For example, we asked learners after completion of the workshop about their ability to apply principles of validity of evidence to instruments and to formulate plans for further development. Also, we created open-response questions to allow for detailed answers about how learners met the educational objectives. The evaluation form overall was intended to capture learner outcomes achieved during the actual workshop as well as learners' plans for further work outside of the workshop. Both outcomes related directly to the education objectives.

Open responses also guided us in further iterative work. We were able to learn more about all learners' ideas and their instrument development following the workshop. Questions such as the one relating to future plans at learners' institutions were intended both to provide feedback for the instructors and to serve as an opportunity for learners to reflect on the knowledge and skills gained during the workshop.

Results

Implementation

The curriculum demonstrated feasibility, adaptability, and generalizability by being implemented as a 4-hour workshop at the COMSEP Annual Meeting in St. Louis, Missouri, in 2016 and at the Educators Across the Healthcare Spectrum (EAHCS) Symposium: Assessment in Health Professional Education in Doha, Qatar, in 2016, as well as a 3-hour workshop at the Pediatric Academic Societies (PAS) Meetings in San Francisco, California, in 2017 and in Toronto, Ontario, Canada, in 2018. Postworkshop evaluations were submitted by a total of 65 learners who attended one of the four workshops. The learners included 12 participants at COMSEP 2016, 35 participants at EAHCS 2016, 11 participants at PAS 2017, and seven participants at PAS 2018.

The learners at each of the three national conferences (COMSEP 2016, PAS 2017, and PAS 2018) were all pediatric clinician-educators, many of whom served in educational leadership roles (e.g., clerkship director, program director). Specifically, the COMSEP learners were pediatric educators with a focus on medical student education. The PAS learners were academic pediatricians whose foci varied in medical education (across the learner continuum, primary vs. specialty), quality improvement, and clinical research. Learners at the national conferences chose to attend the workshop from a list of other, concurrent workshops. In contrast, learners at the one international conference (EAHCS 2016) were a more diverse audience. They were health professionals from a variety of disciplines, as attendance at the workshop was required as part of a 2-day symposium conducted for licensure renewal. Since the audience of the three national workshops was different from that of the international workshop, we analyzed the overall evaluation data across all workshops and compared the pooled evaluation data from the three national workshops with the data from the international workshop.

Overall, learners indicated that they were relative novices with respect to their experience in collecting validity evidence. Across all workshops, 69% of learners reported they were at the beginning or idea stage of developing a project. Those who participated in the international workshop endorsed less overall experience in this area (83% reporting that they were at the beginning stage) compared with those who participated in the national workshops (69% reporting that they were at the beginning stage). No learner reported experience at the middle stage of collecting validity evidence for an instrument.

Learners rated the workshops favorably for each prompt on the evaluation form. The highest-rated prompt was "I learned new knowledge and skills from this activity," which received a mean rating of 4.64 (with 5 = *Strongly Agree*). We used a Mann-Whitney *U* test to compare ratings from the national workshops with

those from the international workshop. Learners in the national workshops provided significantly higher ratings ($p < .001$) across all prompts. A summary of the evaluation results is provided in the [Table](#).

Table. Results of Evaluations at Three National Workshops and One International Workshop Using a 5-Point Likert Scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*)

Prompt	M (SD)		
	All (N = 65)	National (n = 30)	International (n = 35)
Overall, the workshop was effective.	4.42 (0.63)	4.80 (0.63)	4.11 (0.57)
Overall, the speakers were effective.	4.44 (0.56)	4.77 (0.56)	4.17 (0.45)
The format of this activity was appropriate for its content.	4.44 (0.73)	4.87 (0.73)	4.08 (0.73)
This activity was a worthwhile investment in my professional development.	4.47 (0.71)	4.87 (0.71)	4.14 (0.72)
I learned new knowledge and skills from this activity.	4.64 (0.57)	4.93 (0.57)	4.39 (0.64)

Learners were prompted to provide narrative commentary on workshop strengths, opportunities for improvement, and their plans to apply the material. Examples of the narratives included the following:

- How can we improve the activity to make it more relevant?
 - “Provide glossary of terms.”
 - “Provide pre-reading materials.”
 - “Additional workshop with a stats person who could help with psychometric issues.”
- Describe one important thing you plan to apply.
 - “I had not heard of propositions. It helped me clarify my aims, which I will refine further. I will collect more evidence with focus groups and interviews with users.”
 - “Assess my tool according to those domains. I have to search the literature better to see if there are tools I can use.”
 - “I am trying to develop and validate a tool, so all of it was important!”
 - “Thinking about unintended consequences.”
 - “Ensuring I have adequate construct representation.”
- How did the workshop specifically help you with your idea?
 - “Idea of using a gold standard, thinking about unintended consequences.”
 - “Workshop provided me with an idea to assess the response process.”
 - “It helped develop a plan for implementation.”
 - “Detailed guidance on what I need to do next to systematically evaluate our tool.”
 - “Definitely gave me clear goals to go forward with my project.”
 - “Very helpful for getting me to think about refinement of the tool, convergent variables, and response process.”
 - “Narrow the focus. Think systematically about validity evidence to collect.”

Among the national workshops, 25 learners identified one important thing that they planned to apply from the workshop. While the breadth of workshop content was mentioned, learners most commonly planned to consider the unintended consequences of their instrument and response process earlier in the development of the instrument. There were no identified thematic differences from one iteration of the workshop to another. Twenty-four national workshop learners answered the question “How did the workshop specifically help you with your idea?” The most common theme was that the workshop helped the learner better identify next steps for moving the project forward.

Thirteen national workshop learners answered the question “How can we improve the activity to make it more relevant?” Two themes emerged: (1) preference for one consistent example used throughout the workshop to demonstrate novel concepts and (2) a glossary of terms to which participants could refer throughout the workshop. Learners at COMSEP 2016 suggested that we assign reading in advance. COMSEP 2016 was the only national conference for which learners signed up in advance and, thus, the only one for which this suggestion was applicable.

Discussion

Evidence of validity is essential to the assessment and evaluation instruments used in medical education. Aiming for such validity evidence promotes improved outcomes for learners with the eventual goal of impacting patient care outcomes. Medical educators need to be able to gather such evidence for their own instruments and for instruments adapted from another contexts or populations of learners.¹⁵ However, the literature currently lacks a practical approach to explain and apply validity evidence to assessment instruments used in medical education. Our workshop fills this gap. Based on core and peer-reviewed tenets of validity evidence, we have developed a teaching model that has been demonstrated to be effective in meeting educational objectives, feasible, and generalizable.

Summary of Impact

Overall, learners' self-reported responses to the Likert-scale items, as well as their narrative comments, indicated that the workshop was effective. The active application exercises and facilitated feedback during the workshop allowed learners to reflect in action (i.e., in real time) on whether they were meeting a particular educational objective. Because of the high ratio of instructors to learners, coleaders were able to evaluate and reflect in action on the real-time gains of the learners and progress toward meeting the workshop's educational objective.

The exercises provided immediate evaluation feedback as they guided learners to immediately and directly apply their newly gained knowledge to their tool or idea during the workshop and then to their future plans. The evaluation form allowed learners to further reflect on any gains in the scaled and open-ended questions.

Our results fall in Kirkpatrick's tiers of learner reaction and self-reported gains in knowledge and skills.⁶ While these tiers are lower in the hierarchy of the Kirkpatrick pyramid, our results ranked high in their respective tiers, suggesting overall curricular impact in terms of generalizability (variance in educators, geographical locale, institutions, and educator foci), adaptability (national vs. international workshop, 3-hour vs. 4-hour session, and medical educator vs. clinician learners), and feasibility. This speaks to the rigor and quality of the curriculum, suggesting that it can be implemented widely at the present time and evaluated in the higher-ranked tiers of the Kirkpatrick pyramid in the future.

While feedback from the workshop was generally positive across sites, we observed a difference in perceptions when comparing the national workshops conducted in North America and the international workshop conducted in Qatar. We suspect that the difference in perception was associated with two factors. First, participants in the national workshops selected our workshop from a menu of concurrent options, whereas participants in the international workshop were obligated to participate in it as part of a large core curriculum. Evidence suggests that attendees select CME activities based on their interest and needs in a particular topic.¹⁶ Therefore, we hypothesize that those who participated in the national workshops were predisposed to greater interest in the workshop material. In addition, participants in the national workshops were educational leaders and/or educational researchers, whereas participants in the international workshop were clinician-educators who may have had less interest in educational research methods. Validity evidence may be viewed as less integral to the day-to-day work of a clinician-educator. Collectively, these findings suggest some caution in both requiring this workshop as part of a faculty development program and providing it to those without specific interest in educational assessment and scholarship.

Further implementation and evaluation phases should focus on widespread implementation of the work initiated during the workshop. This would add more rigor to the evaluation of the workshop, aiming for the ultimate outcome: development of instruments of validity evidence for use in real-time teaching and clinical care. Thus, we recommend that workshop leaders, in future iterations, consider tracking the work of their learners especially with regard to feasibility and use of their assessment instruments, as we do. Successful implementation of the workshop at different national and international conferences attended

by different types of medical educators indicates that the workshop can be widely disseminated. We hope that further dissemination in *MedEdPORTAL* will lead to more widespread implementation of the workshop and allow further evaluation of its success at higher tiers of the Kirkpatrick pyramid.⁶

Limitations include the evaluation process of the workshop itself. The workshop was evaluated only by learners' self-reported gains of knowledge and skills. Further longitudinal data are needed to track learners' actual work at their institutions following completion of the workshop. Additionally, the workshop has thus far targeted medical educators primarily in North America. It has been implemented and evaluated largely by self-report by medical educators participating in traditional academic conferences. Finally, the evaluation form is generic and is intended to be adapted. Future workshop leaders should fine-tune the evaluation form to meet the specific content and strategies of their workshop.

Implications

While our resource has been evaluated in 3-hour and 4-hour formats, we recognize that needs will vary with different learner populations (e.g., time constraints, cognitive overload) and different conference types (e.g., local to international forums). This workshop can be adapted to meet such demands and has already been adapted into a 1-hour format as a workshop series. Indeed, a 1-hour introductory session followed by two or three follow-up sessions may better suit implementation of the workshop at an institutional level. Follow-up could also occur one on one, with workshop facilitators serving as mentors for individual faculty members to complete small-group exercises using a specific assessment tool. Overall, we urge future workshop facilitators to consider adapting this workshop to suit their own teacher and learner needs. We will continue to consider such adaptation in our own future iterations. Second, we stress the need to consider the evaluation process specific to the learner population. We encourage others to adapt the provided evaluation form to meet their own needs and to consider longitudinal evaluation of their learners. Third, we have delineated the background needed to facilitate this workshop. We note that not all facilitators need to be an expert in a particular content domain. We, the authors, have grown more experienced in particular domains over time, and we continue to learn from each other. Finally, we performed formal debriefings after each workshop to collect immediate reflections regarding actual implementation, overall tone, and response of the learners. We encourage future workshop facilitators to partake in such debriefing sessions, as they added greatly to our own learning. In the future, we also plan to use a focus group to collect data as an additional assessment of our learners.

The need for rigorous assessment and evaluation instruments has risen to prominence. Validity evidence is part of the development process of these instruments in both medical evaluation and clinical care. We provide a practical model that teaches the medical educator how to apply validity evidence to existing instruments and how to develop new instruments that contain validity evidence. Once gained, such knowledge and skills can optimize the future development of instruments used in medical education, with the ultimate goal of leading to eventual improvements in patient care.

Caroline R. Paul, MD: Assistant Professor, Department of Pediatrics, University of Wisconsin School of Medicine and Public Health

Michael S. Ryan, MD, MEHP: Associate Professor, Department of Pediatrics, Virginia Commonwealth University School of Medicine

Gary L. Beck Dallaghan, PhD: Research Associate Professor, Department of Pediatrics, University of North Carolina School of Medicine

Thanakorn Jirasevijinda, MD: Associate Professor, Department of Pediatrics, Weill Cornell Medicine

Patricia D. Quigley, MD, MME, EdS: Assistant Professor, Department of Pediatrics, Johns Hopkins University School of Medicine

Janice L. Hanson, PhD: Professor, Department of Pediatrics, University of Colorado School of Medicine

Amal M. Khidir, MD: Associate Professor, Department of Pediatrics, Weill Cornell Medical College in Qatar

Jean Petershack, MD: Professor, Department of Pediatrics, University of Texas Health Science Center at San Antonio

Joseph Jackson, MD: Assistant Professor, Department of Pediatrics, Duke University Hospital

Linda Tewksbury, MD: Professor, Department of Pediatrics, New York University School of Medicine

Mary Esther M. Rocha, MD: Associate Professor, Department of Pediatrics, Baylor College of Medicine

Disclosures

None to report.

Funding/Support

Dr. Jirasevijinda reports personal fees from the *Journal of Communication in Healthcare* outside the submitted work. Dr. Hanson reports support from Weill Cornell Qatar during the conduct of the study.

Ethical Approval

Reported as not applicable.

References

1. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. 14th ed. Washington, DC: American Educational Research Association; 2014.
2. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. *Acad Med*. 2016;91(6):785-795. <https://doi.org/10.1097/ACM.0000000000001114>
3. Wetzel AP. Factor analysis methods and validity evidence: a review of instrument development across the medical education continuum. *Acad Med*. 2012;87(8):1060-1069. <https://doi.org/10.1097/ACM.0b013e31825d305d>
4. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis. *Med Educ*. 2018;52(6):641-653. <https://doi.org/10.1111/medu.13574>
5. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods, and reporting quality. *Acad Med*. 2013;88(6):872-883. <https://doi.org/10.1097/ACM.0b013e31828ffdcf>
6. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels*. 3rd ed. San Francisco, CA: Berrett-Koehler Publishers; 2006.
7. Kessler CS, Kalapurayil PS, Yudkowsky R, Schwartz A. Validity evidence for a new checklist evaluating consultants, the 5Cs model. *Acad Med*. 2012;87(10):1408-1412. <https://doi.org/10.1097/ACM.0b013e3182677944>
8. Paul CR, Keeley MG, Rebella G, Frohna JG. Standardized Checklist for Otoscopy Performance Evaluation: a validation study of a tool to assess pediatric otoscopy skills. *MedEdPORTAL*. 2016;12:10432. https://doi.org/10.15766/mep_2374-8265.10432
9. Till H, Ker J, Myford C, Stirling K, Mires G. Constructing and evaluating a validity argument for the final-year ward simulation exercise. *Adv Health Sci Educ Theory Pract*. 2015;20(5):1263-1289. <https://doi.org/10.1007/s10459-015-9601-5>
10. Lafave M, Katz L, Butterwick D. Development of a content-valid standardized orthopedic assessment tool (SOAT). *Adv Health Sci Educ Theory Pract*. 2008;13(4):397-406. <https://doi.org/10.1007/s10459-006-9050-2>
11. Andreatta PB, Marzano DA, Curran DS. Validity: what does it mean for competency-based assessment in obstetrics and gynecology? *Am J Obstet Gynecol*. 2011;204(5):384.e1-384.e6. <https://doi.org/10.1016/j.ajog.2011.01.061>
12. Sargeant J, Mann K, Sinclair D, Van der Vleuten C, Metsemakers J. Challenges in multisource feedback: intended and unintended outcomes. *Med Educ*. 2007;41(6):583-591. <https://doi.org/10.1111/j.1365-2923.2007.02769.x>
13. Didier T, Kreiter CD, Buri R, Solow C. Investigating the utility of a GPA institutional adjustment index. *Adv Health Sci Educ Theory Pract*. 2006;11(2):145-153. <https://doi.org/10.1007/s10459-005-0390-0>
14. Kane M. Validation. In: Brennan R, ed. *Educational Measurement*. 4th ed. Westport, CT: Prager Publishers; 2006.
15. Colbert-Getz JM, Ryan M, Hennessey E, et al. Measuring assessment quality with an assessment utility rubric for medical education. *MedEdPORTAL*. 2017;13:10588. https://doi.org/10.15766/mep_2374-8265.10588
16. McLeod PJ, McLeod AH. If formal CME is ineffective, why do physicians still participate? *Med Teach*. 2004;26(2):184-186. <https://doi.org/10.1080/01421590310001643136>

Received: October 25, 2018 | **Accepted:** January 28, 2019 | **Published:** April 12, 2019