

## Interrater Reliability Study Considerations for a Suture Skills Assessment Tool

Alicia Ludden-Schlatter, MD, MSAM | Jane A. McElroy, PhD

PRiMER. 2026;10:14.

Published: 4/9/2026 | DOI: 10.22454/PRiMER.2026.161102

### To the Editor:

We read with great interest the recent article by Patel et al, "Interrater Reliability of a Suture Assessment Tool in Family Medicine Training."<sup>1</sup> The authors highlight an important challenge in competency-based medical education (CBME): ensuring that procedural assessment tools are appropriately validated for the context in which they are used. Their study highlights the value of conducting validation work specifically within family medicine, whereas many assessment tools have historically been adapted from other specialties without sufficient evidence of applicability. We commend the authors for undertaking this careful and much-needed work.

Validation is the process of collecting evidence to appraise an assessment tool.<sup>2</sup> Validation is critical in CBME to maximize consistency in instructor feedback and align training programs toward a mastery standard. We encourage PRiMER readers to validate additional assessment tools for CBME, and Patel et al's study may serve as a model. However, their findings illustrate several methodological considerations that are relevant whenever investigators adapt or validate an instrument for new learner populations, evaluators, or educational settings.

Patel et al analyzed the interrater reliability (IRR) of a suturing skills checklist adapted from the tool originally developed by Sundhagen et al.<sup>3</sup> Patel et al asked 15 family medicine faculty to rate 20 resident procedure videos. Using Light's  $\kappa$  analysis, they found item agreement ranged from "no agreement" to "fair agreement." This finding contrasts with the original checklist, which demonstrated good reliability when analyzed using intraclass correlation (ICC).

Several factors may explain these differing results. First, modifying the Sundhagen et al checklist effectively created a new instrument requiring its own validation, even if changes seemed minor. Second, the populations and contexts differed substantially: the original tool was used by medical students and was reviewed by plastic surgery specialists in Norway, whereas Patel et al's study applied a revised checklist to family medicine residents and faculty in the United States. Assessment tools may perform differently across settings, learner levels, and evaluator expertise. Finally, the analytical methods may contribute to reported differences. Light's  $\kappa$  and ICC assess agreement in distinct ways. Light's  $\kappa$  is more sensitive to category imbalance and chance agreement, whereas ICC is better suited for continuous or ordinal rankings. This difference makes direct comparisons difficult.<sup>4</sup>

Patel et al discuss opportunities to strengthen their checklist, which apply to any instrument readers may consider using: clarifying language, providing structured evaluator training sessions, and adding competency anchors. We also encourage using anchored rating descriptors in assessment tools, because this was shown

to improve the IRR of the Procedural Competency Assessment Tool for shave biopsy.<sup>5</sup>

When adopting any assessment tool, educators should consider whether it is appropriate with respect to its intended purpose. A tool with limited reliability, such as in this study, may still provide value for formative feedback during workshops but is less suitable for high-stakes decisions such as board eligibility.<sup>2</sup> Readers should also consider the limitations of context extrapolation, that is, a tool validated on a task trainers may not perform similarly in real-world patient care.<sup>2</sup> Additional validation assessment for other domains such as content and accuracy would provide insights regarding the utility of this and other checklists.

We commend Patel et al for undertaking the demanding and necessary work of instrument validation. We encourage readers to pursue similar work in validating tools to assess the performance of other procedures required by the American Board of Family Medicine. Patel et al's findings reinforce two important messages for medical educators: first, that even small modifications to an assessment tool or its use with a new population require renewed validation to ensure accuracy, reliability, and educational value; and second, that such validation should be comprehensive, with careful methodological planning and statistical rigor. We hope their study encourages continued development of robust, family medicine-specific assessment tools.

## Corresponding Author

Alicia Ludden-Schlatter, MD, MSAM

University of Missouri Department of Family and Community Medicine, Columbia, MO

[luddena@health.missouri.edu](mailto:luddena@health.missouri.edu)

## Author Affiliations

Alicia Ludden-Schlatter, MD, MSAM - Department of Family and Community Medicine, University of Missouri, Columbia, MO

Jane A. McElroy, PhD - Department of Family and Community Medicine, University of Missouri-Columbia

## References

1. Patel K, Dargel C, Aguayo-Ortega RA, Boswell CL, Stacey SK. Interrater reliability of a suture assessment tool in family medicine training. *PRiMER*. 2026;10:2. doi:10.22454/PRiMER.2026.615323
2. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1(1):31. doi:10.1186/s41077-016-0033-y
3. Sundhagen HP, Almeland SK, Hansson E. Development and validation of a new assessment tool for suturing skills in medical students. *Eur J Plast Surg*. 2018;41(2):207-216. doi:10.1007/s00238-017-1378-8
4. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23-34. doi:10.20982/tqmp.08.1.p023
5. Wells J, Ludden-Schlatter A, Kruse RL, Cronk NJ. Evaluating resident procedural skills: faculty assess a scoring tool. *PRiMER*. 2020;4:4. doi:10.22454/PRiMER.2020.462869

Copyright © 2026 by the Society of Teachers of Family Medicine