

Development and Validation of a Shave Biopsy Training Checklist

Alicia Ludden-Schlatter, MD, MSAM^a; Stephanie Bunt, PhD^b; Kate DuChene Hanrahan, MD, MME^b

AUTHOR AFFILIATIONS:

^aDepartment of Family and Community Medicine, University of Missouri, Columbia, MO

^bDepartment of Family Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA

CORRESPONDING AUTHOR:

Alicia Ludden-Schlatter, Department of Family and Community Medicine, University of Missouri, Columbia, MO, luddena@health.missouri.edu

HOW TO CITE: Ludden-Schlatter A, Bunt S, Hanrahan KD. Development and Validation of a Shave Biopsy Training Checklist. *Fam Med.* 2025;57(X):1-8. doi: [10.22454/FamMed.2025.615731](https://doi.org/10.22454/FamMed.2025.615731)

PUBLISHED: 5 February 2025

KEYWORDS: ACGME competencies, assessment of learner performance, graduate medical education, procedural training

© Society of Teachers of Family Medicine

ABSTRACT

Background and Objectives: Residencies train residents in procedures and assess their competency, but existing assessment tools have demonstrated poor reliability and have not been validated.

Methods: This mixed-methods study validated a shave biopsy checklist with family medicine and dermatology faculty at two academic centers. In each phase of the study, teaching faculty scored a video-recorded simulated procedure using the checklist, and investigators assessed content validity, interrater reliability, and accuracy.

Results: In focus groups of nine family medicine and dermatology faculty, 16 of 18 checklist items met or surpassed 80% interrater reliability. Overall checklist reliability was 74%. Focus group surveys initially revealed insufficient content validity. Lowest performing items were removed, and then the follow-up content validity index (0.76) surpassed the required threshold (0.62). Twenty-one of 70 family medicine faculty completed a final survey, which showed a content validity index of 0.63, surpassing the required threshold of 0.42. Twelve of 70 family medicine faculty viewed and scored a simulated video-recorded procedure. Overall interrater reliability was 91% (Cohen's $d=1.36$). Fourteen of 16 checklist items demonstrated greater than or equal to 90% interrater reliability. Accuracy analysis revealed 67.9% correct responses in focus groups and 84.9% in final testing (simple t test, $P<.001$, Cohen's $d=1.4$).

Conclusions: This rigorously validated checklist demonstrates appropriate content validity, interrater reliability, and accuracy. Findings support use of this shave biopsy checklist as an objective mastery standard for medical education and as a tool for formative assessment of procedural competency.

INTRODUCTION

Procedural training with accurate, objective skills assessment is integral to family medicine residency. The Accreditation Council for Graduate Medical Education (ACGME) requires that residents “must be able to perform all medical, diagnostic, and surgical procedures considered essential for the area of practice.”¹ The American Board of Family Medicine (ABFM) states that graduates must be competent to “perform the procedures most frequently needed by patients in continuity and hospital practices.”² Program directors attest to the ABFM that residents are competent in relevant procedures. This transition to competency-based medical education reinforces the need for validated, objective procedural skills assessments.

Few objective mastery standards for procedure performance or validated tools for assessing procedural competency exist.^{3–5} ACGME milestone evaluations of procedural competency provide narrative descriptions tracking from level 1

(novice) to level 5 (expert) and are intended to be used as a global assessment of all procedures performed rather than specific to one particular procedure (eg, shave biopsy).⁶ The Council of Academic Family Medicine recommends use of the Procedural Competency Assessment Tool (PCAT). The PCAT is a global rating scale based on the Operative Performance Rating System, a tool developed for surgical procedures.^{7–9} The PCAT features vague and unclear wording, and has demonstrated low interrater reliability.¹⁰ To our knowledge, no validation studies regarding the PCAT have been published.

Checklists are a more objective and reliable tool for measuring observable behavior than global rating scales.^{11–14} To improve the consistency and quality of formative faculty feedback and create an objective mastery standard for training, the University of Missouri Department of Family and Community Medicine (MUFCM) residency converted procedure textbook instructions into checklists for resident procedure

workshops.¹⁵ The family medicine residency at the University of Iowa also incorporated these procedure checklists into its training workshops.

Investigators from MUFCM and UI collaborated to rigorously review, refine, evaluate, and validate MUFCM's shave biopsy skills assessment checklist in accordance with commonly used validation frameworks.^{16–18} Here we describe the development and validation process for the shave biopsy procedure checklist. We hypothesized that the shave biopsy checklist would include appropriate content and demonstrate adequate accuracy and interrater reliability for formative assessment of procedure performance.

METHODS

MUFCM created the shave biopsy checklist in 2015 by converting instructions from the *Pfenninger & Fowler's Procedures for Primary Care*¹⁵ textbook into a checklist. MUFCM training workshops use deliberate practice methodology, in which the trainee completes the procedure with direct observation; immediately receives objective, actionable feedback consistent with a mastery standard; and repeats the skill until competency is demonstrated.^{19,20} Accordingly, the checklist was designed to anticipate multiple attempts by the learner until competent. Meetings with MUFCM teaching faculty addressed content, clarity, and stylistic variances. The checklist was used in annual shave biopsy deliberate practice procedure training workshops at MUFCM starting in 2015, with ongoing revisions based on faculty feedback. UI adopted this checklist for its procedure workshops in 2021. Additional minor revisions were made based on UI faculty feedback.

Validation is the process of collecting validity evidence to appraise the appropriateness of the interpretations, uses, and decisions based on assessment results of an evaluation tool.¹⁷ Validation occurred from July 2021 to May 2023 in three phases: (1) primary investigators ([PIs], A.L., K.H., S.B.) and collaborator (J.W.) review checklist content and clarity; (2) focus groups assess checklist content validity (CV) and preliminary interrater reliability (IRR), as well as collect qualitative feedback; and (3) final analysis of checklist CV, IRR, and accuracy is conducted (Figure 1). Over the three phases, this study analyzed CV, IRR, and accuracy of the checklist. This project was determined to be not human subjects research by both the University of Missouri (MU) and the University of Iowa (UI) Institutional Review Boards.

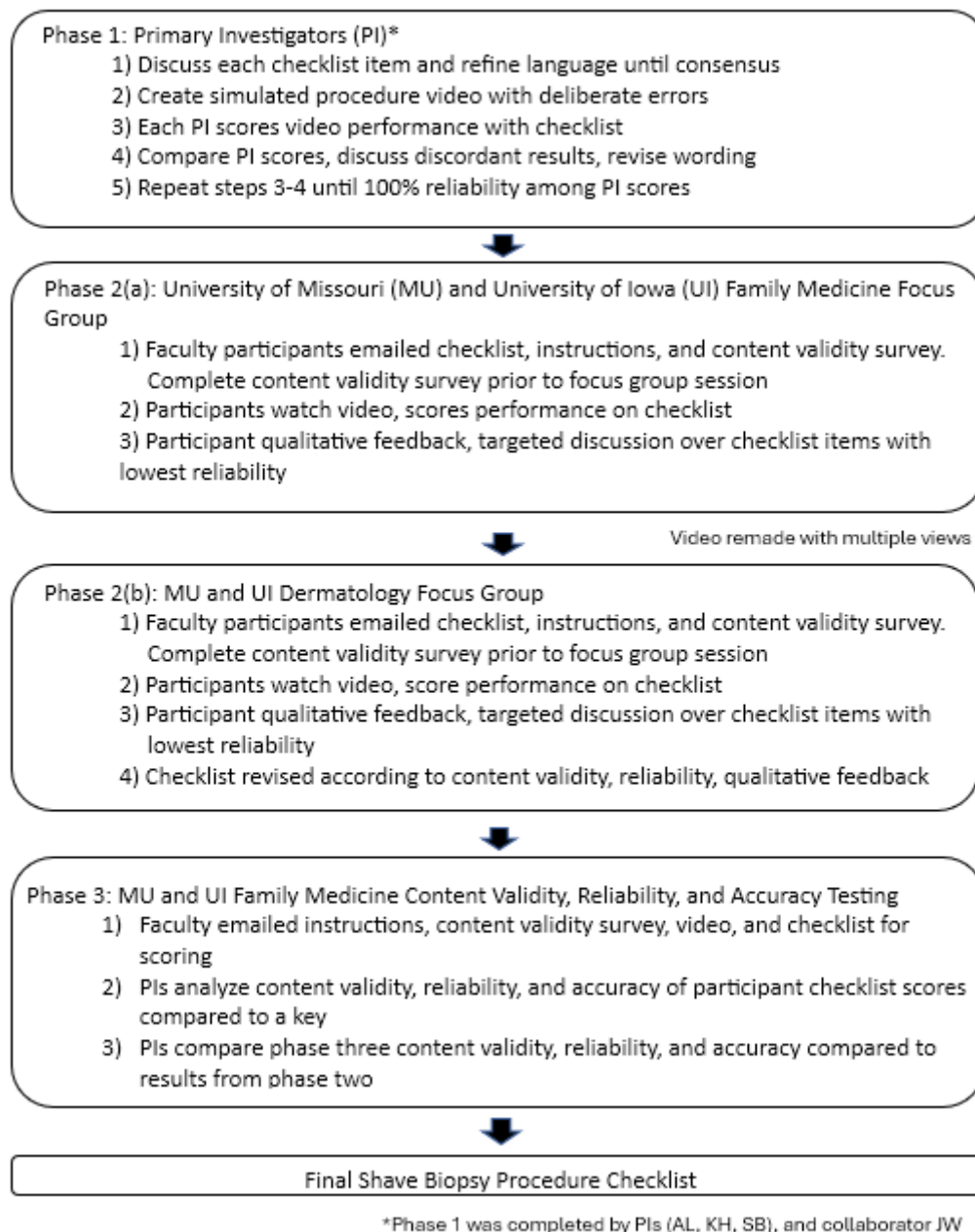
In Phase 1, the PIs reviewed the shave biopsy checklist for content and word clarity. A.L. created a high-resolution video of a simulated shave biopsy with deliberate errors (https://www.youtube.com/watch?v=gNI_WQJGMAY). Checklist items “not done/done incorrectly” on the video included “verbalizes the indication for the procedure,” “states one alternative to the procedure,” and “excises the lesion using dermablade/razor blade in one hand, slightly bowed, and kept parallel to the skin.” All PIs independently watched the video and scored each checklist item as “done” or “not done/done incorrectly.” Any items that were scored differently by PIs were discussed and

revised. This process was repeated until PI “done/not done” scores demonstrated 100% IRR of all checklist items.

Phase 2 comprised two focus groups: first MUFCM and UI family medicine teaching faculty who perform and precept shave biopsies (n=5), followed by MU and UI dermatology faculty (n=4). Prior to each focus group, faculty participants received an email with instructions and completed a CV survey to evaluate each item on the checklist. The CV survey asked participants to evaluate each item on the checklist as “essential,” “useful but not essential,” or “not useful.” A final question asked (with a free-text response) for any items that were missing. After both focus groups concluded, the content validity ratio (CVR) was calculated for each item according to a standard formula that divides the number of participants marking “essential” for each item by the total number of participants.²¹ The CVR for each item was calculated using only ratings of “essential,” excluding “useful but not essential” and “not useful” ratings. A CVR of 1.0 indicates that all participants responded that the item is “essential.”^{21,22} The total checklist content validity index (CVI), a mean of all CVRs, was calculated according to a standard formula described by Lawshe.²¹ Evaluation tools are considered to have adequate CV if the CVI surpasses the critical value, a value determined by a standardized table.²¹

During each focus group, faculty watched a video of a simulated shave biopsy with five errors and scored each item on the checklist as “done” or “not done/done incorrectly” (https://www.youtube.com/watch?v=gNI_WQJGMAY). The video was remade during Phase 2, after the family medicine focus group and prior to the dermatology focus group, to include improved video quality of the procedure (https://www.youtube.com/watch?v=1EdV_Besx50). Any items that had less than 50% IRR were discussed for word clarity and relevance of content. Investigators solicited qualitative feedback and suggestions for improvement. The checklist was revised based on CV data, IRR, and qualitative feedback.

In the final phase, the revised checklist was sent to all MUFCM and UI family medicine teaching faculty that worked regularly with residents who may or may not perform shave biopsy (46 MUFCM, 24 UI family medicine). These 70 faculty received a total of three emails with background information and instructions to (1) complete a CV survey (response options “essential,” “useful but not essential,” or “not necessary”); (2) watch the video of the shave biopsy with errors; and (3) score each item as “done” or “not done/done incorrectly.” PIs calculated final checklist CV according to methods described earlier. PIs analyzed IRR for each checklist item. IRR was determined by calculating the proportion of item responses in agreement with one another for each checklist item. A *t* statistic compared the mean proportion of item responses in agreement (IRR) in Phase 2 and Phase 3 testing. PIs analyzed accuracy for each checklist item. PIs created an answer key reflecting which items were “done” or “not done/done incorrectly” on the video. Accuracy was determined by calculating the proportion of item responses that were correct according to the key.

FIGURE 1. Overview of Shave Biopsy Procedure Checklist Validation Phases

*Phase 1 was completed by PIs (AL, KH, SB), and collaborator JW

A *t* statistic compared the proportion of correct responses for Phase 2 and Phase 3 testing. To determine the effect of improvements to the checklist, Cohen's *d* was calculated comparing Phase 2 and Phase 3 for both IRR and accuracy. A Cohen's *d* value of greater than 0.60 indicates a large effect size. All statistical analysis was completed using the software SAS version 9.4 (SAS Institute). All participation in surveys and focus groups was voluntary, and no incentives were offered. Surveys were conducted via Qualtrics, and all survey responses were anonymous.

RESULTS

Participants

The family medicine focus group had five faculty (3 MUFCM, 2 UI). The dermatology focus group had four UI faculty. One MU dermatology faculty completed the CV survey but could not attend the focus group.

For Phase 3, 21 of 70 (30%) MUFCM and UI family medicine faculty completed the CV survey. Twelve of 70 (17%) MUFCM and UI family medicine faculty scored the video-recorded procedure using the final checklist.

Qualitative Feedback

Both the family medicine and dermatology focus groups provided qualitative feedback revealing that neither specialty had a clear reference standard for the correct performance of dermatologic procedures, including shave biopsy. Many educational resources exist, but training largely follows an apprenticeship model in which residents are trained according to the stylistic preferences of their supervising attending.

The family medicine focus group indicated that faculty liked the checklist because it was thorough and “black and white, they either did it or did not do it.” Faculty members identified items that were vague or had subjective wording such as “correct technique.” Faculty suggested that some items were missing, including “appropriate return precautions for the patient.”

The dermatology focus group highlighted the importance of including a time-out. Dermatologists unanimously agreed that considering whether epinephrine is mixed with lidocaine for local anesthesia is not important because no evidence exists for necrosis in noses or digits anesthetized with epinephrine.²³ Dermatologists were unanimous that aspiration prior to injection of anesthetic is not useful due to the superficial nature of shave biopsy. Dermatologists also were unanimous that direction of the bevel is not necessary for this superficial procedure; fanning the needle underneath the skin is acceptable, and minimizing needle sticks is more important. The item “reviews contraindications with the patient” was originally intended to prompt learners to consider contraindications; however, dermatologists advocated to remove this item from the checklist because informed consent would not be discussed if the patient had a disqualifying contraindication and would therefore be ineligible for the procedure.

Content Validity

Table 1 summarizes CV, reliability, and accuracy data.

Ten faculty (including MU and UI family medicine and dermatology) completed the CV surveys prior to their respective focus groups. The standard CVI for 10 participants is 0.62; and the CVI for the checklist prior to focus groups was 0.58, not meeting the threshold for adequate CV.²¹ The checklist items with the lowest CVRs (indicating that the fewest participants rated the item as “essential”) were “states one alternative” (CVR=0.2), “verbalizes absence of contraindications” (CVR=0.2), “aspirates before injection” (CVR=-0.2), “inserts with needle bevel up” (CVR=0), and “applies Band-Aid” (CVR=0). The latter four of these items were removed, which increased the checklist’s CVI to 0.76, therefore demonstrating adequate CV. In Phase 3, the overall CVI was 0.63, which exceeded the CVI threshold of 0.42, indicating that the content of the checklist measures the intended constructs and has adequate CV.

Interrater Reliability

In the family medicine focus group (n=5), 10 of 18 checklist items demonstrated 100% IRR and six items had 80% IRR. The 22 items with the lowest IRR of 60% were “describes the

indication for procedure” and “inserts and withdraws needle in one plane.”

The dermatology focus group (n=4 UI) had eight items with 100% reliability, seven items with 75% reliability, and three items with 50% reliability (although note that one faculty member responded that all items were “done”). The three items with 50% IRR were “verbalizes absence of contraindications for the procedure,” “describes the indication for procedure,” and “inserts and withdraws needle in one plane.”

In the third phase, 12 of 70 (17%) invitees completed the “done/not done” reliability testing. Fourteen out of 16 items had over 90% IRR. “Cleans site with alcohol” had 63% IRR, and “verbalizes indication for procedure” had 50% IRR.

The overall checklist IRR improved from 74% in Phase 2 to 91% in Phase 3. The effect size, as measured by Cohen’s *d*, was *d*=1.36, indicating a large effect size.

Accuracy

The proportion of correct rater “done/not done” responses compared to the key was 67.9% in Phase 2 and 84.9% in Phase 3 (simple *t* test, *P*<.001, Cohen’s *d*=1.4, indicating a very large effect size). These results demonstrate improvement in the checklist’s ability to facilitate accurate evaluator assessment of trainee behavior.

Summarization of Changes to Checklist During Validation Process

PIs revised the checklist based on focus group qualitative feedback and results of CV and IRR testing. The shave biopsy checklist at the beginning of Phase 2 had 18 items. Two items were removed: “verbalizes one contraindication for the procedure” and “applies Band-Aid.” The items “verbalizes return for care precautions” and “confirms patient name, date of birth, procedure and allergies” were added. We combined three items to become “injects anesthetic into the dermis underneath the lesion to elevate the lesion.” Wording was clarified on six items. The final checklist had 16 items. Changes to the checklist in response to findings are summarized in **Table 2**. The final published checklist is provided in **Figure 2**.

DISCUSSION

This study addressed gaps in current competency-based residency training. Despite ACGME requirements for objective competency-based assessments, few tools exist to support objective evaluation or to standardize minimum procedural competency metrics. Focus groups of both family medicine and dermatology faculty revealed that neither specialty has universally accepted standards for cutaneous procedures. The Council of Academic Family Medicine recommends the PCAT for assessing procedural competence, but this global rating scale is too subjective to facilitate timely, objective, and actionable feedback and has not been validated.¹⁰

Few validated tools exist for assessment of medical trainee’s performance of procedures, especially in the ambulatory setting. Published literature supports the use of validated checklists and global rating scales for infant lumbar

TABLE 1. Comparison of Shave Biopsy Checklist Performance Between Phases Two and Three

	Phase 2 (focus groups) checklist n=# of faculty	Phase 3 checklist n=# of faculty
Number of items in checklist	18	16
Content validity index	Critical value=0.62 Initial CVI=0.58 CVI after revisions=0.76 (n=10)	Critical value=0.42 Initial CVI=0.63 (n=21)
Interrater reliability ^a	74% (n=9)	91% (n=12)
Accuracy ^b	67.9% (n=9)	84.9% (n=12)

^aCohen's $d=1.36$ ^bSimple t test, $P<.001$, Cohen's $d=1.4$

Abbreviation: CVI, content validity index

TABLE 2. Comparison of Original and Final Shave Biopsy Checklist

Original checklist (N=18 items)	Final checklist (N=16 items)
Informed consent	Informed consent
	Confirms patient name, date of birth, procedure, and allergies
Verbalizes absence of contraindications for the procedure	
Describes the indication for the procedure	Verbalizes the indication for the procedure
States one potential risk	States one potential risk (eg, pain, bleeding, infection)
States one potential benefit	States one potential benefit (eg, cosmetic removal, confirm diagnosis)
States one alternative to the procedure	States one alternative to the procedure
Local anesthesia	Local anesthesia
Dons nonsterile gloves	Dons nonsterile gloves
Uses appropriate lidocaine (with or without epinephrine)	Uses appropriate anesthetic
Draws lidocaine with red drawing needle	Draws anesthetic with large bore needle
Cleans site with alcohol in outward concentric circles prior to injecting local anesthesia	Cleans site with alcohol in outward concentric circles prior to injecting local anesthesia
Injects with needle bevel up	
Aspirates before injection	
Inserts and withdraws needle in one plane	Injects anesthetic into the dermis underneath the lesion to elevate the lesion
Procedure flow	Procedure flow
Uses dermatome/razor with correct technique	Excises the lesion using dermatome/razor blade in one hand, slightly bowed, and kept parallel to the skin
Obtains specimen and places specimen in preservative	Obtains specimen and places specimen in preservative
Manages bleeding via gauze and/or silver nitrate/aluminum chloride	Manages bleeding via gauze and/or silver nitrate/aluminum chloride
Verbalizes wound care instructions	Verbalizes wound care instructions
Applies Band-Aid	
Throughout procedure, at all times, sharps were capped (using no touch technique) or disposed in sharps container.	Throughout procedure, at all times, sharps were capped (using no touch technique) or disposed in sharps container.
	Verbalizes return for care precautions

puncture,^{24,25} paracentesis,²⁶ and pediatric laceration repair.²⁷ Our project addressed the paucity of validated assessment tools for outpatient dermatologic procedural competency, specifically shave biopsy.

Commonly used validation frameworks developed by Kane and Messick (in Cook & Hatala) have emphasized the importance of considering and scientifically testing assumptions and properties of the assessment tool.^{16–18} The evidence collected

may support or refute the validity argument, which is ultimately a judgment that the aforementioned evidence supports using the tool for the purpose it was intended and within the context under which it was studied. The quality and quantity of validity evidence needed correlates to the implications for the learner being assessed.¹⁷ The revisions made to this shave biopsy checklist in response to evidence collected during the validation process improved CV, increased IRR, and signif-

FIGURE 2. Shave Biopsy

Shave Biopsy

Resident Name: _____ Evaluator Name: _____



Date: _____

Instructions for Evaluators:
Only mark "done" if completed correctly and directly observed by you.

	Attempt #1		Attempt #2		Attempt #3	
	DONE	NOT DONE/DONE INCORRECTLY	DONE	NOT DONE/DONE INCORRECTLY	DONE	NOT DONE/DONE INCORRECTLY
INFORMED CONSENT						
Verbalizes the indication for the procedure						
Confirms patient name, DOB, procedure, and allergies						
States one potential risk (i.e, pain, bleeding, or infection)						
States one potential benefit (i.e., cosmetic removal, confirm diagnosis)						
States one alternative to the procedure						
LOCAL ANESTHESIA						
Dons non-sterile gloves						
Uses appropriate anesthetic						
Draws anesthetic with large bore needle						
Cleans site with alcohol in outward concentric circles prior to injecting local anesthesia						
Injects anesthetic into the dermis underneath the lesion to elevate the lesion						
PROCEDURE FLOW						
Excises the lesion using Dermablade/razor blade in one hand, slightly bowed, and kept parallel to the skin						
Obtains specimen and places specimen in preservative						
Manages bleeding via gauze and/or silver nitrate/aluminum chloride						
Verbalize wound care instructions						
Verbalizes return for care precautions						
Throughout procedure, at all times, sharps were capped (using no touch technique) or disposed in sharps container						

Reference: Fowler G. Pfenninger and Fowler's Procedures for Primary Care. 4th ed. Canada: Elsevier; 2020.

Developed by: Alicia Ludden-Schlatter, MD, MSAM; Jack Wells, MD, MHA, CHSE; Stephanie Bunt, PhD; Kate DuChene Hanrahan, MD, MME, FFAFP

icantly increased accuracy with large effect size. The final shave biopsy procedure checklist demonstrated acceptable CV, IRR, and accuracy for assessment and feedback of trainee performance in procedure workshop training sessions. This checklist may serve as a national competency standard for teaching the shave biopsy procedure in dermatology, family medicine, and other primary care training programs.

MUFCM and UI have used earlier variants of the checklist for simulation workshops in lieu of the PCAT since 2015 and 2021, respectively. At these institutions, these checklists both demonstrate trainee competency and facilitate achieving competency as an integral component of deliberate practice training methodology. In deliberate practice, educators observe the trainee’s focused performance of a skill, compare performance to well-defined objectives or tasks, and provide correction. The learner repeats and improves performance until competency is demonstrated, as compared to a mastery standard.¹⁹ At MUFCM and UI, this checklist serves as the

mastery standard for resident shave biopsy competence.

Our study had several potential limitations. Response rates for participation invitations were low, potentially leading to response bias. Some items continue to demonstrate mild discordance in faculty assessment of items “done/not done.” Whether this is due to limitations of the checklist tool, limitations of capturing performance on video, or the need to train faculty in using the checklist is unclear. The video used in Phase 2 was remade to include improved video quality, which may have impacted Phase 2 IRR. Observer limitations are a challenge for both recorded and live procedure assessment. One dermatology faculty responded “done” to all checklist items, which may have been due to misunderstanding the directions and may have impacted Phase 2 IRR. Although the item “states one alternative” had a low CVR, investigators included it in the final checklist because although few focus group participants identified it as “essential,” all other participants identified it as “useful.” Considering that this tool’s purpose is to be used

in the context of a training workshop, investigators determined that this item is important to procedural training and left it in the checklist.

Finally, though validation was conducted with physicians representing two specialties at two institutions, results may not be generalizable to practice patterns outside of the United States.

Next Steps

Validated checklists like the one described here may increase accuracy of competency-based skills assessment, improve quality and promote consistency of feedback, and serve to standardize the quality of procedural training nationally. This checklist may inform residency program directors' attestation of graduating residents' procedural competency. The checklist also may serve as a shave biopsy competency standard for physicians and providers in dermatology and primary care fields.

This study validated the shave biopsy checklist in the context of training with formative feedback; additional testing is recommended to validate the checklist in the context of procedure performance in the clinical care of live patients. We recommend additional validity testing, such as correlating scores of video-recorded procedure performance on live patients done by novice, intermediate, and expert physicians to validate this tool in the context of proof of competency for hospital privileges, board certification, or licensure.

Next steps will include validity testing for other checklists developed at MUFCM and in use at MUFCM and UI. The use of validated checklists may carry implications for reducing implicit bias of evaluators working with trainees of underrepresented groups.

CONCLUSIONS

Validated checklists like the one described here may serve to standardize the quality of procedural training nationally. This shave biopsy checklist is based on textbook instructions and was refined over several years of use in family medicine resident procedure training workshops at two institutions. Multispecialty, multi-institution focus group feedback and content validity analysis yielded a thorough yet concise procedural competency standard. Statistical analysis demonstrated acceptable content validity, reliability, and accuracy in evaluator scores for this checklist to serve as a validated tool for competency-based assessment of shave biopsy skills in a formative setting.

Presentations

- ▶ Works in Progress Poster, Society of Teachers of Family Medicine (STFM) Annual Spring Conference 2022, Indianapolis, Indiana.
- ▶ Selected methods and results discussed in 60-minute seminar, STFM 2023, Tampa, Florida.
- ▶ Selected methods discussed in Round Table Discussion, STFM 2023, Tampa, Florida.

- ▶ Selected methods and results discussed in 60-minute workshop, International Meeting on Simulation in Healthcare 2024, San Diego, California.

ACKNOWLEDGMENTS

The authors thank Jack Wells, MD, MHA, CHSE-A, who contributed to Phase 1 of the project (primary investigator revisions for content and clarity); and Clarence Kreiter, PhD, who assisted with statistical analysis.

REFERENCES

1. ACGME Program Requirements for Graduate Medical Education in Family Medicine. Accreditation Council for Graduate Medical Education. 2024. https://www.acgme.org/globalassets/pfassets/programrequirements/120_familymedicine_2024.pdf.
2. Newton W, Cagno CK, Hoekzema GS, Edje L. Core outcomes of residency training 2022 (provisional). *Ann Fam Med*. 2023;21(2):191-194.
3. Kedian T, Gussak L, Savageau JA. An ounce of prevention: how are we managing the early assessment of residents' clinical skills? A CERA study. *Fam Med*. 2012;44(10):723-726.
4. Whitehead CR, Kuper A, Hodges B, Ellaway R. Conceptual and practical challenges in the assessment of physician competencies. *Med Teach*. 2015;37(3):245-251.
5. Sawyer T, White M, Zaveri P. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med*. 2015;90(8):33.
6. Family Medicine Milestones. Accreditation Council for Graduate Medical Education. 2019. <https://www.acgme.org/globalassets/pdfs/milestones/familymedicinemilestones.pdf>.
7. Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL. Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. *Surgery*. 2005;138(4):640-647.
8. A User's Manual for the Operative Performance Rating System (OPRS). American Board of Surgery. 2012. https://program.absurgery.org/xfer/assessment/oprs_user_manual.pdf.
9. Benson A, Markwell S, Kohler TS, Tarter TH. An operative performance rating system for urology residents. *J Urol*. 2012;188(5):877-878.
10. Wells J, Ludden-Schlatter A, Kruse RL, Cronk NJ. Evaluating resident procedural skills: faculty assess a scoring tool. *PRiMER*. 2020;4:4.
11. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ*. 2015;49(2):161-173.
12. Schoepp K, Danaher M, Kranov A, A. An effective rubric norming process. *Pract Assess Res Eval*. 2018;23(11):1-12.
13. Turner K, Bell M, Bays L. Correlation between global rating scale and specific checklist scores for professional behavior of physical therapy students in practical examinations. *Educ Res Int*. 2014;2014(7):1-6.
14. Kogan JR, Hess BJ, Conforti LN, Holmboe ES. What drives faculty ratings of residents' clinical skills? the impact of faculty's own clinical skills. *Acad Med*. 2010;85(10):25-28.
15. Pfenninger JL, Fowler GC. *Pfenninger and Fowler's Procedures for Primary Care, 3rd ed*. Saunders; 2010.

16. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560–575.
17. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)*. 2016;1(1):31.
18. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50(1):1–73.
19. Mcgaghie WC, Issenberg SB, Cohen ER, Barsuk JH, Wayne DB. Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? a meta-analytic comparative review of the evidence. *Acad Med*. 2011;86(6):706–711.
20. Wiggins G. Seven keys to effective feedback. *Educ Leadersh*. 2012;70(1):10–16.
21. Lawshe CH. A quantitative approach to content validity. *Pers Psychol*. 1975;28(4):563–575.
22. Gilbert G, Prion S. Making sense of methods and measurement: Lawshe's content validity index. *Clin Simul Nurs*. 2016;12(12):530–531.
23. Denkler K. A comprehensive review of epinephrine in the finger: to do or not to do. *Plast Reconstr Surg*. 2001;108(1):114–124.
24. Braun C, Kessler DO, Auerbach M, Mehta R, Scalzo AJ, Gerard JM. Can residents assess other providers' infant lumbar puncture skills? validity evidence for a global rating scale and subcomponent skills checklist. *Pediatr Emerg Care*. 2017;33(2):80–85.
25. Gerard JM, Kessler DO, Braun C, Mehta R, Scalzo AJ, Auerbach M. Validation of global rating scale and checklist instruments for the infant lumbar puncture procedure. *Simul Healthc*. 2013;8(3):148–154.
26. Riesenber LA, Berg K, Berg D. The development of a validated checklist for paracentesis: preliminary results. *Am J Med Qual*. 2013;28(3):227–231.
27. Seo S, Thomas A, Uspal NG. A global rating scale and checklist instrument for pediatric laceration repair. *MedEdPORTAL*. 2019;15:10806.