

Validity Evidence for an Open-Access Internal Medicine Point-of-Care Ultrasound Image Interpretation Assessment

Brian P. Elliott; Maria Rechtin; Nick Villalobos; Renee Dversdal; Cristin Mount; Robert Browning; Ian Prichard; Kimberly Fabyan

ABSTRACT

Background: Point-of-care ultrasound (POCUS) has expanding utility in internal medicine practice, but there is a paucity of standardized multiple-choice question (MCQ) POCUS assessments with validity evidence for internal medicine residents. We aimed to develop a POCUS assessment that tests major internal medicine concepts and then evaluate its validity. **Methods:** We used published internal medicine consensus recommendations to determine the main concepts to evaluate in our assessment. Three investigators drafted multiple-choice questions for the assessment, which were then distributed to 5 subject-matter experts (SMEs) across 4 institutions, who iteratively evaluated and provided recommendations on content validity until predefined content validity indices were achieved. We distributed these questions to internal medicine physicians trained and untrained in POCUS, and then calculated quality metrics. **Results:** We distributed 59 MCQs to SMEs, and then narrowed them to meet predefined content validity index cutoffs for reliability (r-CVI), usefulness (u-CVI), and clarity (c-CVI). The 36 final questions had high average content ratings (r-CVI: 0.92; u-CVI: 0.92; c-CVI: 0.89). The reliability of these questions was demonstrated by a KR-20 of 0.81.

Evidence for internal structure was shown with an average discrimination index of 0.47. Relationship to other variables was demonstrated by a significantly better average performance by 5 trained versus 16 untrained assessment-takers (trained: 84%, untrained: 60%, $p < 0.01$). **Conclusions:** This MCQ POCUS assessment for internal medicine, with a focus on image interpretation, demonstrated moderate preliminary validity evidence.

INTRODUCTION

Point-of-care ultrasound (POCUS) has expanding utility in internal medicine practice, which has led to a push for standardized POCUS training during internal medicine residency training [1-3]. As POCUS education continues to scale, it is paramount that proper assessments with strong validity evidence scale with it, supporting that adequate competencies are achieved.

There are multiple published POCUS assessments with strong validity evidence, including the ACTS and UCAT assessments [4-6]. These assessments are geared toward testing learners' technical skills and image acquisition, with some image interpretation and clinical integration incorporated. These approaches excel in addressing the multiple

subskills (image acquisition, image interpretation, and clinical integration) in POCUS, but it can be challenging to assess learners' ability to interpret images across a broad range of pathology because of the time-consuming and resource-intensive nature of objective structured clinical examination (OSCE) assessments. Multiple-choice question (MCQ) assessments, on the other hand, can assess image interpretation competency and clinical integration across a broader range of POCUS images. Essentially, while OSCE assessments address the "shows" of Miller's pyramid, an MCQ assessment with strong validity evidence can ensure that the broader requisite knowledge of image interpretation, the "knows", is obtained first [7].

The current literature lacks an open-access POCUS assessment for internal medicine residents with strong validity evidence. Published studies exploring new POCUS curricula often use in-house assessments. The issue with in-house assessments is that the images are usually proprietary or not openly available, making external validation and reproducibility of the curriculum outcomes difficult. The purpose of our study was to create and assess the validity of an open-access MCQ POCUS assessment for internal medicine residents that could be used for curricular benchmarking and low- to moderate-stakes competency assessment [4]. We designed the assessment to cover recommended items for internal medicine resident POCUS curricula, which have previously been evaluated with consensus recommendations [2].

METHODS

Subject Matter Expert (SME) Selection

Since no uniform credentialing for internal medicine POCUS competency exists, we used any of the following as minimal criteria for SME

qualification: 1) research background in POCUS, 2) fellowship training in POCUS, or 3) society certification in ultrasound. We assembled 5 SMEs with these qualifications to participate in validity evaluation: RB, who is board certified in critical care and critical care echocardiography; NV, who is board certified in critical care and critical care echocardiography, RD who is a hospitalist widely published in POCUS and the Chief Medical Officer of a POCUS company, CM who completed an ultrasound fellowship, and IP who completed an ultrasound fellowship. The SMEs are affiliated with 4 separate institutions across the United States with diverse training backgrounds involving both inpatient and outpatient practice. BE served as the SME facilitator and received the Society of Hospital Medicine certificate in POCUS.

Assessment Development

We developed our assessment in parallel with the creation of a POCUS track at an internal medicine residency program. The purpose of the track was to teach a broad range of POCUS exams pertinent to internal medicine. The assessment's purpose was to distinguish internists with a sufficient image interpretation knowledge base from those without a sufficient knowledge base. Track participants took this assessment before and after the track curriculum as a requirement before OSCE assessments. We excluded POCUS related to procedures from our assessment because the assessment's purpose was image interpretation rather than procedural competencies. We followed the sources of validity evidence outlined in the Standards of Educational and Psychological Measurement, and based on Messick's framework for assessment development [8,9].

Content

Previous consensus recommendations for internal medicine POCUS curricula described by

Ma et al. were used as the foundation of our question writing [2]. The study describes final consensus recommendations to include 11 POCUS applications in expanded internal medicine curricula: inferior vena cava (IVC), lung B lines, pleural effusion, abdominal free fluid, internal jugular vein, lung consolidation, pneumothorax, knee effusion, gross left ventricular systolic function, pericardial effusion, and right ventricular strain. Question writers developed an initial blueprint goal, with the final assessment aiming to be around 30 questions total, with around 4 covering knobology/artifacts, 7 cardiac, 4 pulmonary/pleural, 3 evaluation of shock, 2 volume assessment, 2 vascular, 3 skin, soft tissue, and musculoskeletal, 2 abdominal, and 2 renal. These domains were agreed upon by the SMEs and question writers to represent a broad assessment of internal medicine POCUS content, allowing for customization and narrowing to the initial consensus-recommended topics, if desired. BE, KF, and an internal medicine resident (MR) created 1-5 MCQs for each of these topics. The authors followed MCQ best-practice recommendations for content generation, including single-best-answer format, clinically relevant stems, plausible distractors, avoidance of cueing, and three or more answer options [10]. In anticipation that questions would be dropped during content validity ratings and psychometric testing, extra questions were developed in a ratio equal to the desired final blueprint. Since the assessment objective was to focus on image interpretation, question writers were instructed to aim for greater than two-thirds image-based questions, with images obtained from de-identified clinical studies from the question writers or a publicly available educational repositories that allow for redistribution with attribution (thepocusatlas.com), selected to represent high-yield and commonly encountered pathology.

BE, KF, and MR drafted 59 MCQs, 44 (75%)

image-based, for initial distribution to SMEs. The drafted MCQs were distributed to SMEs with the same MCQ best-practice recommendations and a survey form to rank the relevance, usefulness, and clarity of each MCQ on a scale of 1-4 (i.e., 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant), in accordance with previously recommended methods [11]. Ratings were provided in a blinded fashion (SMEs could not see other ratings until aggregated) to prevent group influence via Microsoft Excel files. SMEs could optionally provide comments and suggestions. Only the question writers (none of the SMEs) could view individual SME ratings and feedback, which the facilitators then used to analyze the content validity indices for each characteristic (r-CVI, u-CVI, and c-CVI). We predefined aims in the initial study protocol to achieve recommended CVI parameters of an average rating greater than 0.78 for individual items and a total average rating greater than 0.90 [11]. Items with CVIs less than 0.78 were either revised (typically to add clarity recommended by SMEs) or replaced. Revisions, replacements, and feedback were arbitrated by the initial question writers, BE, KF, and MR. Any time a question was revised or replaced, it required another review by SMEs prior to acceptance. Revised and replaced questions were redistributed to the SMEs in iterative rounds until the CVI parameters were met and at least 40 questions were generated, which lasted 3 rounds.

Response Processes

Authored questions adhered to best-practice recommendations for developing MCQs in health professions education. We used testmoz.com to administer the assessment with a randomized question order and randomized answer choices. Additionally, author MR, an internal medicine resident, served as a target-group reviewer to provide learner-level

feedback regarding item clarity, interpretation, and alignment with intended POCUS concepts. Feedback from this process was incorporated into item revisions to ensure that questions were understood as intended and did not introduce construct-irrelevant difficulty.

Internal Structure

The MCQs that met the CVI content validity criteria were compiled into an online assessment. We invited internal medicine physicians and subspecialists to complete the MCQ assessment. This included 10 residents enrolling in our institution's POCUS track before any educational sessions were completed. All participants were from 3 academic hospitals in the United States and tested in 2024. We analyzed the discrimination indices using a standard 27% top and bottom threshold and point-biserial correlations of their results, which demonstrate how well each item distinguishes high performers from low performers. We set a priori cutoffs following published recommendations for point-biserial indices, specifically to exclude items with a result less than +0.10, include only essential items with results between +0.10 and +0.15, include minimal items with results between +0.15 and +0.20, and include items greater than +0.20 [10]. When conflicts arose between content validity and psychometric performance, items were preferentially retained if they addressed essential content domains. Given the dichotomous (correct/incorrect) scoring of MCQ items, we selected the Kuder-Richardson Formula 20 (KR-20) as an appropriate measure of internal consistency reliability and used this to determine a standard error of measurement.

When choosing questions for a narrowed assessment, we eliminated items based on a point-biserial value of less than +0.10. We then narrowed items to achieve predefined average

CVI item goals of greater than 0.90.

Relationship to Other Variables

Participants were stratified into trained and untrained groups. Participants were categorized a priori as 'trained' or 'untrained' based on level of formal POCUS exposure. The 'trained' group included fellows with at least 6 months of pulmonary/critical care training or attending physicians with regular POCUS use (more frequently than weekly), reflecting sustained exposure to image interpretation. The 'untrained' group included residents without advanced training, although prior limited exposure to POCUS curricula was not an exclusion criterion because these exposures were reportedly brief and not clearly structured or sustained training. We then calculated the discrimination indices between trained and untrained participants and compared overall assessment scores between the two groups.

Consequences/Standard Setting

We used the Angoff method implemented via the 8-step process outlined by Yudkowsky et al. to determine what may be an acceptable standard for passing the MCQ assessment [10, 12]. SMEs were provided with the psychometric data for the final MCQs among both trained and untrained internal medicine physicians. They were instructed to estimate how many test-takers out of 100 who have borderline competency would answer each question correctly. They were briefly trained via email instruction on how the Angoff method works and the definition of a borderline test-taker. After the first round of voting, we calculated means and standard errors, aiming to achieve a standard error of the cut score that is less than half of the standard error of the test [13]. We redistributed the MCQ assessment to SMEs, using the results in an iterative process, until this outcome was achieved.

Statistical Analysis

Descriptive statistics were done with Microsoft

Excel version 2312. Statistical comparisons were done with GraphPad Prism version 10.1.2 for Windows. Continuous variables were compared using the Mann-Whitney test. Categorical variables were compared using Fisher's exact test.

RESULTS

Out of the 59 items drafted for SME review, 22 met acceptable CVI criteria, and we removed 17 for very poor ratings. Twenty questions were edited, and another 18 were added for round two. Of the 60 questions in round two, 42 met acceptable CVI criteria and were compiled for psychometric testing. All content recommended by the framing consensus guidelines was addressed except for using internal jugular vein assessment for volume status, due to low content validity ratings for questions targeting this item.

Twenty-one people took the assessment made for psychometric testing. Of these, 5 were trained, and 16 were untrained. The characteristics of these participants are shown in Table 1. The full psychometric data results, average content validity ratings, and cutoff ratings are shown in Table 2. After analyzing psychometric data, 6 questions did not meet psychometric criteria, yielding 36 final multiple-choice questions. Of the 36 questions, 25 (69%) are image-based. This final assessment demonstrated good internal reliability with a KR-20 of 0.81. Values >0.7 are acceptable, and >0.80 are considered moderate to high internal reliability [10]. Trained participants tested significantly higher than untrained participants (Figure 1), showing a strong relationship between assessment results and other training variables.

Four subject matter experts completed Angoff ratings. The goal standard error of ratings was determined to be less than 1.8 (half the standard error of the test). Initial ratings yielded an

average cut score of 72.8% with a standard error of 1.95. The second round yielded the final average cut score of 72.3% with a standard error of 1.26.

DISCUSSION

In this study, we constructed a validity argument based on Messick's unified framework. Content validity was supported through structured SME review and high content validity indices. Response process validity was supported through adherence to MCQ best-practice guidelines and pilot testing with a representative learner. Internal structure validity was demonstrated by good internal consistency and strong item discrimination. Relationships to other variables were supported by significantly higher performance among trained participants compared to untrained participants. Finally, the validity of consequences was addressed through Angoff standard setting, supporting the use of this assessment for formative and low-stakes evaluation. Collectively, these findings provide preliminary validity evidence supporting the interpretation of assessment scores as a measure of POCUS image interpretation competency.

This assessment could fill a crucial gap in internal medicine POCUS education. While there are several studies detailing the implementation of POCUS curricula among internal medicine residencies, the assessments used for all of them are usually entirely different and lack clear validity evidence [14-17]. Of the published assessments with validity evidence, such as ACTS or UCAT, they are aimed at simulation environments. This assessment is meant to complement them. MCQ-based assessments are particularly suited for evaluating the 'knows' and 'knows how' levels of Miller's pyramid, providing a scalable method to assess foundational knowledge prior to higher-stakes simulation evaluations. In order to truly perform

comparative evaluations of curriculum interventions, we first need a uniform method of measuring efficacy. This assessment not only demonstrates the potential to fill this role but was specifically designed to be open access for easy reproducibility. The assessment is stored on mededassessments.com (direct preview), a free, open-access repository where verified instructors can access the assessment, edit it, reproduce it, collect their own results, and analyze them.

One key advantage of an editable open-access assessment is that POCUS is a broad domain. Clinicians use it very differently based on setting, specialty, and goals, which limits the ability to reach definitive psychometric conclusions. We experienced this during our own psychometric data collection. For example, internists who are primarily hospitalists likely perform much better at cardiac POCUS than musculoskeletal, so obtaining consistent psychometric results across multiple exams was difficult. The confidence survey echoed this point, as most of our expert sample was not confident in their musculoskeletal and skin/soft tissue examinations. This presents a limitation in our data, notably in our definition of an expert, which may not represent true expertise across all exams. Further studies could explore relationships to other variables more closely by splicing exams by expertise. We did not pursue this route because we aimed to develop one broad assessment for internal medicine.

Our assessment development also included an Angoff method that reached an adequate consensus for a recommended passing score of 72.3%. However, this assessment has not yet proved strong enough validity evidence for high-stakes assessment, and this recommendation was informed by a test-taking group that largely (94%) had POCUS exposure, possibly influencing the standards of a borderline test-taker. Evaluating this assessment further

with larger sample sizes that also demonstrate external validity is recommended before implementing it as a high-stakes assessment.

A notable part of our assessment development includes the experience that volume assessment was particularly challenging to reach consensus on. While our initial questions included volume status assessment via the internal jugular vein, questions on this topic failed to meet predefined content validity criteria, likely due to the heterogeneity of this examination [18]. However, volume status assessment via the inferior vena cava performed better and was included, likely because it follows standardized echocardiographic criteria [19]. Further study is needed to specifically determine optimal methods of testing POCUS volume assessment.

Our study had several limitations. We presented an argument for validity evidence aligning with Messick's unified framework, but additional validity evidence, including correlation with external measures of POCUS competency (e.g., OSCE performance), cognitive interviewing techniques to further evaluate response processes, and validation across larger and more diverse learner populations, would help strengthen the evidence. The small sample size limits the internal and external validity of the assessment. Psychometric estimates and reliability testing with KR-20 may also vary with larger and more diverse cohorts. Choosing an acceptable other variable to test for relationships with our assessment was difficult, as there is no gold standard for POCUS competency. We used fellowship or other advanced training for relationships to other variables, but the ability of the characteristic to conclude competent training is limited. The trained participants predominantly had critical care training, which created a limited ability to test non-critical care image interpretation competency. For example, lower SME confidence in musculoskeletal and

skin/soft tissue domains may have reduced content validity in these areas, highlighting the need for domain-specific expert inclusion in future iterations. Our untrained participants also had some formal POCUS education, which may have overestimated what truly untrained internal medicine residents may score on the assessment and have potentially skewed the Angoff ratings. Another limitation is the broad cognitive domain of internal medicine POCUS. Within this domain, each exam type and different clinical setting requires different cognitive abilities, and the ability of an MCQ test to capture this broad domain is limited.

Our findings demonstrate a framework for creating open educational assessment materials that may improve reproducibility and standardization in the assessment of POCUS learners. Future use of this framework may be particularly valuable for multicenter or longitudinal internal medicine POCUS training programs, where a shared open-access assessment could support pre/post curricular evaluation, comparison of educational interventions, and iterative item refinement using pooled learner performance data.

CONCLUSION

We developed an open-access MCQ POCUS assessment for internal medicine residents that demonstrates preliminary validity evidence. While the findings support its potential use for formative assessment and curricular evaluation, further validation in larger and more diverse populations is required before broader implementation.

ARTICLE INFORMATION

Published: 2026-09-01

Corresponding Author: Brian Elliott
brianelliottmd@gmail.com

Author Contributions: All authors contributed to the

initial planning and execution of the study protocol, as well as revisions and editing of the manuscript. Brian Elliott drafted the first version of the manuscript and performed the statistical analysis.

Author Affiliations

Brian Elliott, MD: Uniformed Services University of the Health Sciences, Bethesda, MD, USA; Walter Reed National Military Medical Center, Bethesda, MD, USA

Maria Rechten, MD: Walter Reed National Military Medical Center, Bethesda, MD, USA

Nick Villalobos, MD: University of Texas at McGovern Medical School, Houston, TX, USA

Renee Dversdal, MD: Oregon Health & Science University, Portland, OR, USA

Cristin Mount, MD: Uniformed Services University of the Health Sciences, Bethesda, MD, USA; POCUS Focused, Curriculum Director for Internal Medicine, Chagrin Falls, OH, USA; Saint Joseph Medical Center, Tacoma, WA, USA

Robert Browning, MD: Uniformed Services University of the Health Sciences, Bethesda, MD, USA; Walter Reed National Military Medical Center, Bethesda, MD, USA

Ian Prichard, DO, RDCS: Brooke Army Medical Center, Ft Sam Houston, TX, USA

Kimberly Fabian, MD: Uniformed Services University of the Health Sciences, Bethesda, MD, USA; Walter Reed National Military Medical Center, Bethesda, MD, USA

Acknowledgements:

The full article and any supplementary material are available at rjmedicine.org/publications/e00013

REFERENCES

1. Sabath BF, Singh G. Point-of-care ultrasonography as a training milestone for internal medicine residents: the time is now. *J Community Hosp Intern Med Perspect.* 2016;6(5):33094. 10.3402/jchimp.v6.33094
2. Ma IWY, Arishenkoff S, Wiseman J, Desy J, Ailon J, Martin L, et al. Internal medicine point-of-care ultrasound curriculum: consensus recommendations from the Canadian Internal Medicine Ultrasound (CIMUS) Group. *J Gen Intern Med.* 2017;32(9):1052-1057. 10.1007/s11606-017-4071-5
3. LoPresti CM, Jensen TP, Dversdal RK, Astiz DJ. Point-of-care ultrasound for internal medicine residency training: a position statement from the Alliance of Academic Internal Medicine. *Am J Med.* 2019;132(11):1356-1360. 10.1016/j.amjmed.2019.07.019
4. Kumar A, Kugler J, Jensen T. Evaluation of trainee competency with point-of-care ultrasonography (POCUS): a conceptual framework and review of

- existing assessments. *J Gen Intern Med*. 2019;34(6):1025-1031. 10.1007/s11606-019-04945-4
5. Millington SJ, Arntfield RT, Guo RJ, Koenig S, Kory P, Noble V, et al. The Assessment of Competency in Thoracic Sonography (ACTS) scale: validation of a tool for point-of-care ultrasound. *Crit Ultrasound J*. 2017;9(1):25. 10.1186/s13089-017-0081-0
6. Bell C, Hall AK, Wagner N, Rang L, Newbigging J, McKaigney C. The Ultrasound Competency Assessment Tool (UCAT): development and evaluation of a novel competency-based assessment tool for point-of-care ultrasound. *AEM Educ Train*. 2021;5(3):e10520. 10.1002/aet2.10520
7. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-S67. 10.1097/00001888-199009000-00045
8. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington (DC): American Educational Research Association; 1999.
9. Messick S. Validity. In: Linn RL, editor. *Educational measurement*. 3rd ed. New York: Macmillan; 1989. p. 13-103.
10. Yudkowsky R, Park YS, Downing SM, editors. *Assessment in health professions education*. 2nd ed. New York: Routledge; 2019. 10.4324/9781315166902
11. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459-467. 10.1002/nur.20199
12. Angoff WH. Scales, norms, and equivalent scores. In: Thorndike RL, editor. *Educational measurement*. 2nd ed. Washington (DC): American Council on Education; 1971. p. 508-600.
13. Cohen AS, Kane MT, Crooks TJ. A generalized examinee-centered method for setting standards on achievement tests. *Appl Meas Educ*. 1999;12(4):343-366. 10.1207/S15324818AME1204_2
14. Kuperstein H, Alam W, Paroya A, Patel K, Ahmad S. The design, performance and organizational impact of a point-of-care ultrasound (POCUS) elective for internal medicine residents. *BMC Med Educ*. 2025;25(1):261. 10.1186/s12909-025-06802-x
15. Mellor TE, Junga Z, Ordway S, Hunter T, Shimeall WT, Krajnik S, et al. Not just hocus POCUS: implementation of a point of care ultrasound curriculum for internal medicine trainees at a large residency program. *Mil Med*. 2019;184(11-12):901-906. 10.1093/milmed/usz124
16. Geis RN, Kavanaugh MJ, Palma J, Speicher M, Kyle A, Croft J. Novel internal medicine residency ultrasound curriculum led by critical care and emergency medicine staff. *Mil Med*. 2023;188(5-6):e936-e941. 10.1093/milmed/usab392
17. Boniface MP, Helgeson SA, Cowdell JC, Simon LV, Hiroto BT, Werlang ME, et al. A longitudinal curriculum in point-of-care ultrasonography improves medical knowledge and psychomotor skills among internal medicine residents. *Adv Med Educ Pract*. 2019;10:935-942. 10.2147/AMEP.S220153
18. Drum B, La Course B, Kelly M, York A, Worrall E, Martins J, et al. Does this patient have volume overload? The Rational Clinical Examination. *JAMA*. 2026;335(13):1159-1168. 10.1001/jama.2026.0446
19. Porter TR, Shillcutt SK, Adams MS, Desjardins G, Glas KE, Olson JJ, et al. Guidelines for the use of echocardiography as a monitor for therapeutic intervention in adults: a report from the American Society of Echocardiography. *J Am Soc Echocardiogr*. 2015;28(1):40-56. 10.1016/j.echo.2014.09.009

Table 1. The characteristics of participants taking the POCUS assessment. Data are represented as counts (percentages).

Characteristic	Trained (n=5)	Untrained (n=16)
Level of training		
Resident	0 (0)	14 (88)
Fellow	3 (60)	2 (13)
Attending	2 (40)	0 (0)
Practicing Specialty		
Internal medicine	1 (20)	14 (88)
Pulmonary/Critical Care	3 (60)	1 (6)
Critical Care	1 (20)	0 (0)
Emergency Medicine	1 (20)	0 (0)
Cardiology	0 (0)	1 (6)
Frequency of use		
More frequently than once per week	4 (80)	2 (13)
About once per week	1 (20)	3 (19)
About every 2-3 weeks	0 (0)	2 (13)
About once every 1-2 months	0 (0)	9 (56)
Never	0 (0)	0 (0)
Exams confidently using		
Cardiac	5 (100)	6 (38)
Pulmonary/pleural	5 (100)	5 (31)
Evaluation of shock	5 (100)	4 (25)
Assessment of volume status	3 (60)	9 (56)
Evaluation of skin and soft tissue	1 (20)	2 (13)
Musculoskeletal	1 (20)	1 (6)
Abdominal	2 (40)	0 (0)
Renal	4 (80)	2 (13)
Dedicated training		
None	0 (0)	1 (6)
POCUS elective	2 (40)	2 (13)
Residency-related curriculum	1 (20)	15 (94)
Fellowship-related curriculum	5 (100)	0 (0)
Dedicated certificate, board examination, etc.	1 (20)	0 (0)

Table 2. Psychometric results and characteristics of assessment items.

Item	Proportion Correct	Discrimination Index	Point biserial correlation	Trained/Untrained discrimination index	r-CVI	u-CVI	c-CVI	Subdomain
1	0.38	0.5	0.35	0.29	1	1	1	skin/MSK
2	0.29	0.75	0.61	0.68	0.8	0.8	0.8	volume status
3	0.38	0.5	0.11	0.03	1	1	1	skin/MSK
4	0.62	0	0.13	0.24	0.8	0.8	0.8	knobology/artifacts
5	0.81	0.5	0.55	0.25	1	1	1	knobology/artifacts
6	0.33	0.75	0.52	0.61	1	1	0.8	knobology/artifacts
7	0.95	0.25	0.25	0.06	1	1	1	cardiac
8	0.86	0.25	0.3	-0.08	1	1	1	cardiac
9	0.62	0.75	0.6	0.5	1	1	1	cardiac
10	0.95	0	0.1	0.06	1	1	1	volume status
11	0.48	0.5	0.52	0.43	1	0.8	0.8	cardiac
12	0.48	0.25	0.12	0.16	0.8	1	0.8	cardiac
13	0.86	0.25	0.3	0.19	0.8	0.8	0.8	cardiac
14	0.95	0	0.1	0.06	1	0.8	0.8	cardiac
15	0.62	0.5	0.37	0.5	0.8	0.8	0.8	cardiac
16	0.76	0.25	0.34	0.31	1	0.8	0.8	cardiac
17	0.57	0.75	0.41	0.56	1	1	1	pulmonary
18	0.48	0.25	0.19	0.16	1	1	0.8	pulmonary
19	0.76	0.5	0.52	0.31	1	1	0.8	pulmonary
20	0.95	0.25	0.25	0.06	0.8	0.8	0.8	pulmonary
21	0.76	0.5	0.44	0.31	1	1	0.8	pulmonary
22	0.81	0.5	0.42	0.25	1	1	0.8	pulmonary
23	0.52	1	0.7	0.63	1	1	1	shock
24	0.62	0.75	0.55	0.5	0.8	0.8	1	shock
25	0.52	0.75	0.6	0.63	0.8	0.8	0.8	shock
26	0.57	0.5	0.32	0.04	0.8	0.8	1	volume status
27	0.52	0.75	0.38	0.1	0.8	0.8	0.8	vascular
28	0.38	0.5	0.33	0.29	1	1	0.8	vascular
29	0.71	0.75	0.57	0.38	0.8	0.8	0.8	skin/MSK
30	0.86	0.25	0.18	0.19	1	1	1	skin/MSK
31	0.86	0.5	0.46	0.19	0.8	1	1	abdominal
32	0.52	0.5	0.28	0.36	1	1	0.8	abdominal
33	0.29	0.5	0.15	0.15	1	0.8	1	renal
34	0.43	0.5	0.38	0.23	1	1	1	renal
35	0.81	0.25	0.25	-0.01	0.8	1	0.8	renal
36	0.52	0.75	0.51	0.63	0.8	1	1	cardiac

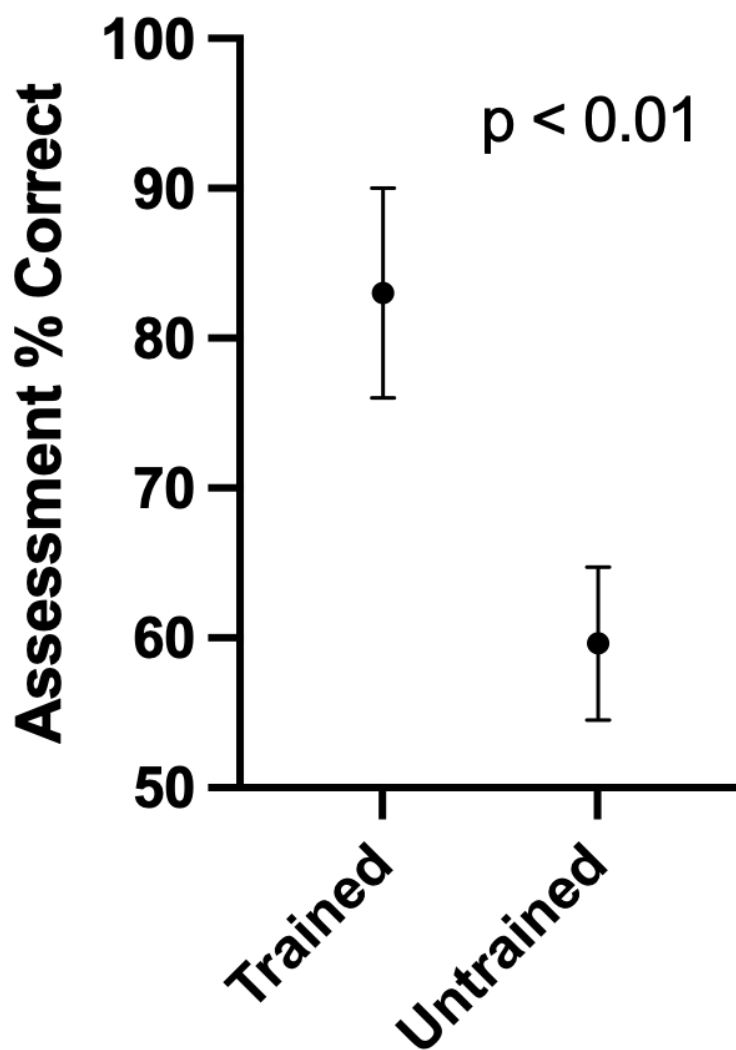


Figure 1. Comparison of complete assessment results between trained (n=5) and untrained participants (n=16). Dots represent the mean, and error bars represent the 95% confidence interval.