

Original Paper

# Digital Simulation–Based Ultrasound Training for Physiotherapy Students: Prospective Randomized Controlled Trial

Samuel Fernández-Carnero<sup>1</sup>, PhD; Belen Diaz-Pulido<sup>1</sup>, PhD; Jorge Méndez-Rodríguez<sup>2</sup>, MSc; Daniel Pecos-Martin<sup>1</sup>, PhD; Santiago Garcia-Miguel<sup>1</sup>, PhD; Alexander Achalandabaso-Ochoa<sup>3</sup>, PhD; Nicolas Cuenca-Zaldívar<sup>4</sup>, PhD; Fermín Naranjo-Cinto Sr<sup>5</sup>, PhD

<sup>1</sup>Universidad de Alcalá, Facultad de Medicina y Ciencias de la Salud,, Departamento de Enfermería y Fisioterapia, Grupo de Investigación Fisioterapia y Dolor, 28801 Alcalá de Henares, Spain

<sup>2</sup>Development and Programming Department, Amazonic Global, s.l, Madrid, Madrid, Spain

<sup>3</sup>Department of Health Sciences, Universidad de Jaen, Jaen, Spain

<sup>4</sup>Research Group in Nursing and Health Care, Puerta de Hierro Health Research Institute—Segovia de Arana (IDIPHISA), Majadahonda, Madrid, Spain

<sup>5</sup>Facultad de Medicina, Benemerita Universidad Autonoma de Puebla, Puebla, Mexico

**Corresponding Author:**

Nicolas Cuenca-Zaldívar, PhD

Research Group in Nursing and Health Care, Puerta de Hierro Health Research Institute—Segovia de Arana (IDIPHISA)

C. Joaquín Rodrigo, 1

Majadahonda, Madrid 28222

Spain

Phone: 34 639962935

Email: [nicolas.cuenca@salud.madrid.org](mailto:nicolas.cuenca@salud.madrid.org)

## Abstract

**Background:** Ultrasound education has traditionally relied on on-site training, but scalable digital solutions are increasingly needed.

**Objective:** This study aimed to evaluate the effectiveness of a simulation-based online ultrasound platform compared with traditional on-site training.

**Methods:** A prospective randomized blinded study was conducted with 68 physiotherapy students (n=34 per group). Participants were assigned to a simulation-based online training platform (WAZO) or traditional on-site instruction. Both groups completed identical theoretical and practical assessments. Item response theory using a Rasch model was applied to evaluate item difficulty and student ability.

**Results:** No significant differences were found between the online and on-site groups in theoretical scores (mean 4.94, SD 1.47 vs mean 5.08, SD 1.14;  $P=.65$ ) or practical performance variables, including probe handling (26/34, 76.5% vs 28/34, 82.4%;  $P=.37$ ) and structure identification (24/34, 70.6% vs 22/34, 64.7%;  $P=.19$ ). Measurement outcomes also showed no significant differences, including structure diameter (mean 3.78, SD 0.79 mm vs mean 3.98, SD 1.27 mm;  $P=.46$ ) and structure surface distance (mean 3.94, SD 1.97 mm vs mean 3.24, SD 0.64 mm;  $P=.06$ ). Item response theory analysis identified image optimization, structure diameter, and structure surface distance as the most difficult items, while probe handling and structure identification were the most informative. The model demonstrated high discriminative performance (area under the curve=0.93), with sensitivity of 87% and specificity of 80%.

**Conclusions:** Simulation-based online ultrasound training provides comparable theoretical and practical outcomes to traditional on-site instruction, supporting its use as a scalable and accessible educational alternative.

*JMIR Med Educ* 2026;12:e87897; doi: [10.2196/87897](https://doi.org/10.2196/87897)

**Keywords:** ultrasound education; simulation-based training; physiotherapy education; digital health education; randomized controlled trial; item response theory

## Introduction

Training in ultrasound for health science professionals for clinical use is based on the traditional classroom model [1] due to the empirical need to learn this tool, which has demonstrated its usefulness and suitability. Innovation in higher education is an opportunity to provide better educational experiences that allow health care professionals to provide a greater benefit to society [2]. Point-of-care ultrasound has demonstrated high reliability in soft-tissue scanning, being completely harmless, low cost, and high value for bedside use [3,4]. These conditions mean that ultrasound enjoys greater consideration than other imaging tests because the clinician can extract relevant information that can be directly correlated with clinical history, making diagnosis more powerful, because imaging tests are not always indicative of dysfunction or pain [5]. The training pathway in some health sciences disciplines includes ultrasound competencies from the second year of training [6], an approach that has also been adopted by other disciplines to better prepare professionals for clinical practice.

The use of simulators in certain disciplines such as aeronautics [7], automotive engineering, or robotic surgeries is widely accepted and proven to be a fundamental basis for a lower carbon footprint. Online training systems in the health sciences have been shown to enable professionals to achieve higher success rates, lower costs, and greater efficiency and to work more quickly [8].

Online and digital materials have been previously used to explore the possibility of ultrasound education in different specialties, such as obstetrics [9], anesthesia residents for lung assessment [10], physiotherapy [11,12], postdegree physician echocardiography [13], and other allied health sciences, demonstrating the potential of these resources and the need to improve this line of education to achieve the necessary skills and knowledge for clinical settings.

To the best of our knowledge, the current models of training in ultrasound are based on videos of varying complexity, leaving aside simulation and training in aspects of image optimization or the use of tools such as gain, dual image, or calipers.

Item response theory (IRT) was used to evaluate the psychometric properties of the assessment. Detailed methodological assumptions and model specifications are provided in [Multimedia Appendix 1](#) [14,15].

One of the main arguments in favor of IRT methods is that they do not use, as in classical test theory, the standard error of measurement (SEM) of the tested population to estimate the parameters, so the calculated parameters are not influenced by the characteristics of the population. This allows the creation of questionnaires whose properties are invariant regardless of the population to which they are applied [16].

The use of the Rasch analysis requires the fulfillment of several assumptions: unidimensionality, that is, that all the items measure the same construct; local independence, that is, the items are not related to each other; and invariance

of the parameters, that is, the items present the same item characteristic curves across subsamples of the participants. Finally, the existence of differential item functioning must be verified, that is, whether students with the same level of  $\theta$  belonging to different subgroups of the sample have the same probability of responding correctly to the items [17].

The purpose of this study was to evaluate whether online training with a simulator platform called WAZO (WAZO Technologies Lab, sl) [18] (“the platform” hereafter) could be a viable alternative to traditional training, given that point-of-care ultrasound is becoming essential, and health care professionals need the knowledge necessary to provide better health care.

## Methods

### Overview

A prospective longitudinal experimental study was designed in which all second-year students enrolled in the Assessment in Physiotherapy course of the e Degree in Physiotherapy program at the University of Alcalá (Spain) were randomized in 2 groups: online (experimental) and on-site (control).

The study design followed the CONSORT (Consolidated Standards of Reporting Trials) standard [19]. This study was not prospectively registered in a clinical trials registry, as it was conducted as an educational intervention within an academic setting.

The inclusion criterion was successful completion of previous subjects related to this course (such as anatomy), and the exclusion criterion was the presence of vision and/or mobility difficulties in the upper limbs.

The on-site group received the basic musculoskeletal ultrasound training provided each year as part of the General Procedures II and Assessment in Physiotherapy courses through face-to-face lectures followed by practical sessions using ultrasound scanners at a ratio of 1 ultrasound scanner for every 2 students, with teacher support provided in small groups of 12 students. The online group was given access to the online ultrasound training platform without the opportunity to ask questions to instructors or handle an ultrasound scanner until 2 hours before the examination, when they were allowed contact with the device.

For the evaluation of the competencies acquired, all students took the same tests. The study variables were extracted from these tests, consisting of two parts: (1) theoretical (10 multiple-choice questions with 4 answers, only 1 of which was valid), and (2) practical examination to identify an anatomical structure (selected at random) using ultrasound in a partner (selected at random) where the examiner collected the following items on a rubric (success or failure, time to recognize it, position of the assessor, position of the patient, skill in handling the probe, and image optimization). Measurements in a training phantom (Venipuncture IV Injection Training Pad Model, VTurboWay, Ltd) of the intragel structures (structure diameter and structure surface distance) were taken with the ultrasound calipers.

The examiners who assessed the students were different from the investigator who performed the first randomization for online and on-site assignment to avoid bias at this point and differed from the instructor who taught the face-to-face classes. The phantom measurements were taken by an ultrasound expert on the same day prior to the student assessments to provide a reference standard for subsequent comparative analysis.

Statistical analysis was performed with R software (version 4.1.3: R Foundation for Statistical Computing). The significance level was set at  $P < .05$ . The distribution of the quantitative variables was tested with the Shapiro-Wilk test. Quantitative variables are expressed as mean (SD) and qualitative variables as absolute and relative values (%).

The presence of significant differences in the responses between the online and on-site groups was verified using Fisher exact test in the case of qualitative variables, and with the Mann-Whitney  $U$  test or 2-tailed Student  $t$  test for independent samples (after checking the assumption of homogeneity of variances with the Levene test) in the quantitative variables based on their distribution.

For structure diameter and structure surface distance measurements, the intraclass correlation coefficient (ICC [3,1]) was calculated between student measurements and the reference values using a 2-way mixed single-measure model (consistency and absolute agreement). Relative reliability was defined as poor ( $<0.5$ ), moderate (0.5-0.75), good (0.75-0.9), or excellent ( $>0.9$ ). As absolute reliability, the SEM was calculated. The minimum detectable change was also calculated.

IRT analysis using a Rasch model was conducted to estimate item difficulty and student ability. Detailed assumptions, model fit procedures, and statistical tests are provided in [Multimedia Appendix 1](#).

To minimize potential bias related to conflicts of interest, several safeguards were implemented. Group allocation was performed by an independent researcher not involved in teaching or assessment, ensuring allocation concealment. Examiners responsible for evaluating both theoretical and practical performance were independent of the intervention and blinded to group assignment. These examiners were not involved in the development of the platform and had no access to allocation data.

All assessments were conducted under standardized conditions using identical evaluation criteria for both groups. Although participants could not be blinded due to the nature of the intervention, no explicit indicators of group allocation were provided to examiners during the evaluation process.

## Ethical Considerations

The study was approved by the Vicerrectorado de Innovación Docente from the Universidad de Alcalá (UAHEV/1434). An informed consent form was signed by all participants. In addition, data were analyzed after anonymization to further reduce the risk of bias.

## Results

### Overview

The sample consisted of 68 students, mostly women ( $n=42$ , 61.8%). No significant differences were observed in the demographic characteristics and examination outcomes between the groups ([Table 1](#)).

In the analysis of the structure diameter and structure surface distance vs reference measurements, the ICC (3,1) was not significant with poor values and high and precise SEMs ([Table 2](#)).

**Table 1.** Demographic characteristics and examination outcomes of the participants.

Variables	Overall (n=68)	Online (n=34)	On-site (n=34)	<i>P</i> value
Sex, n (%)				>.99
Female	42 (61.8)	21 (61.8)	21 (61.8)	
Male	26 (38.2)	13 (38.2)	13 (38.2)	
Patient position, n (%)				.37
Correct	56 (82.4)	27 (79.4)	29 (85.3)	
Incorrect	7 (10.3)	3 (8.8)	4 (11.8)	
Missing	5 (7.4)	4 (11.8)	1 (2.9)	
Sonographer position, n (%)				.24
Correct	62 (91.2)	30 (88.2)	32 (94.1)	
Incorrect	1 (1.5)	0 (0)	1 (2.9)	
Missing	5 (7.4)	4 (11.8)	1 (2.9)	
Probe handling, n (%)				.37
Correct	54 (79.4)	26 (76.5)	28 (82.4)	
Incorrect	9 (13.2)	4 (11.8)	5 (14.7)	
Missing	5 (7.4)	4 (11.8)	1 (2.9)	
Image optimization, n (%)				.31

Variables	Overall (n=68)	Online (n=34)	On-site (n=34)	P value
Correct	17 (25)	7 (20.6)	10 (29.4)	
Incorrect	46 (67.6)	23 (67.6)	23 (67.6)	
Missing	5 (7.4)	4 (11.8)	1 (2.9)	
Structure identification, n (%)				.19
Correct	46 (67.6)	24 (70.6)	22 (64.7)	
Incorrect	17 (25)	6 (17.6)	11 (32.4)	
Missing	5 (7.4)	4 (11.8)	1 (2.9)	
Test duration (seconds), mean (SD)	101.44 (66.41)	102.52 (61.55)	100.46 (71.48)	.90
Structure diameter (mm), mean (SD)	3.88 (1.06)	3.78 (0.79)	3.98 (1.27)	.46
Structure surface distance (mm), mean (SD)	3.57 (1.47)	3.94 (1.97)	3.24 (0.64)	.06
Theoretical exam score, mean (SD)	5.01 (1.30)	4.94 (1.47)	5.08 (1.14)	.66

**Table 2.** Interobserver reliability between student and reference measurements.

Group and variable	ICC <sup>a</sup> (95% CI)	Mean (SD; 95% CI)	SEM <sup>b</sup> (95% CI)	MDC <sup>c</sup>
Online				
Structure diameter (mm)	0.253 (-0.052 to 0.514)	3.771 (0.606; 3.59-3.951)	0.537 (0.408-0.666)	1.488
Structure surface distance (mm)	0 (-0.301 to 0.301)	4.304 (1.453; 3.923-4.684)	1.457 (1.057-1.856)	4.038
On-site				
Structure diameter (mm)	0.009 (-0.278 to 0.295)	3.863 (0.923; 3.629-4.097)	0.923 (0.712-1.134)	2.558
Structure surface distance (mm)	0 (-0.287 to 0.287)	3.959 (0.921; 3.84-4.078)	0.67 (0.454-0.885)	1.857

<sup>a</sup>ICC: intraclass correlation coefficient.

<sup>b</sup>SEM: SE of measurement.

<sup>c</sup>MDC: minimal detectable change.

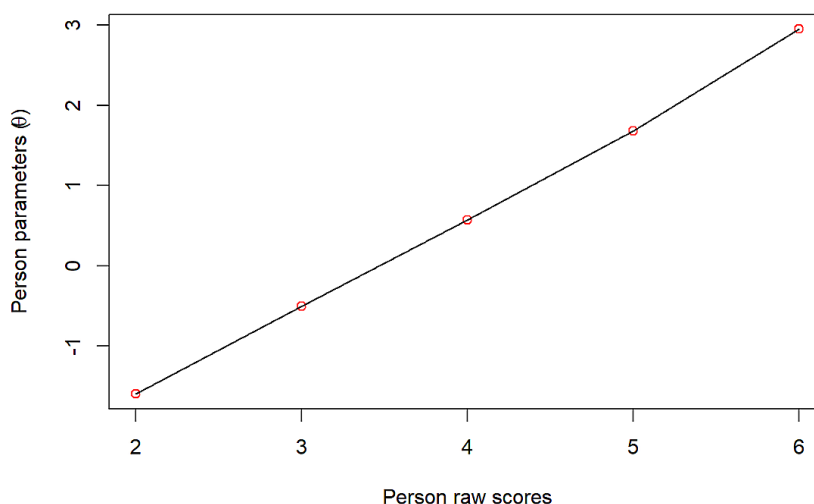
## IRT Analysis

### Overview

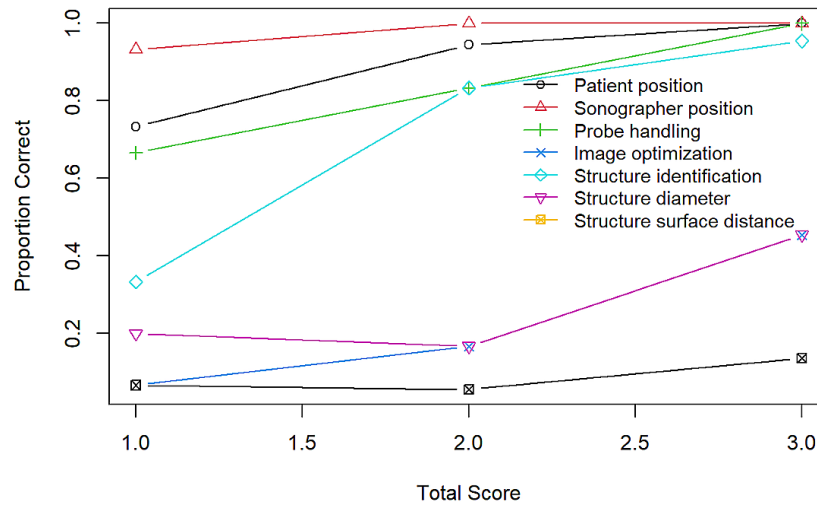
The average level of success of the items was 59.637%, with the questions on image optimization, structure diameter, and structure surface distance having a hit rate of less than 50%. A total of 15 students answered 3 questions correctly, 18 answered 4 questions correctly, and 22 answered 5 questions correctly (Figure 1). Only 5 students answered 6 questions correctly and 3 students answered 2 questions correctly; there were no students who answered all the questions correctly or

who failed all of them. It can be seen that in the questions on patient position, sonographer position, probe handling, image optimization, and structure identification, the proportion of correct answers increases as the final score obtained increased, while in the questions on structure diameter and structure surface distance, the proportion of hits decreased up to a score of 2 and then increased slightly up to the maximum score. The probability of correct responses increased consistently with total score for most items, supporting the expected hierarchical structure of the assessment (Figure 2).

**Figure 1.** Plot of person parameters showing the monotonic mapping of total raw scores to the estimated latent ability parameters. Higher raw scores correspond to higher levels of latent competence, providing a linear-like transformation within the observed score range (2-6).



**Figure 2.** Test information plot showing the proportion of correct responses for each assessment criterion across overall score levels (1.0-3.0). Steeper lines (eg, structure identification) indicate criteria with high discriminative power. Lines near the top (eg, sonographer position) represent easier tasks, whereas lines near the bottom (eg, structure surface distance) represent highly difficult tasks across all proficiency levels.

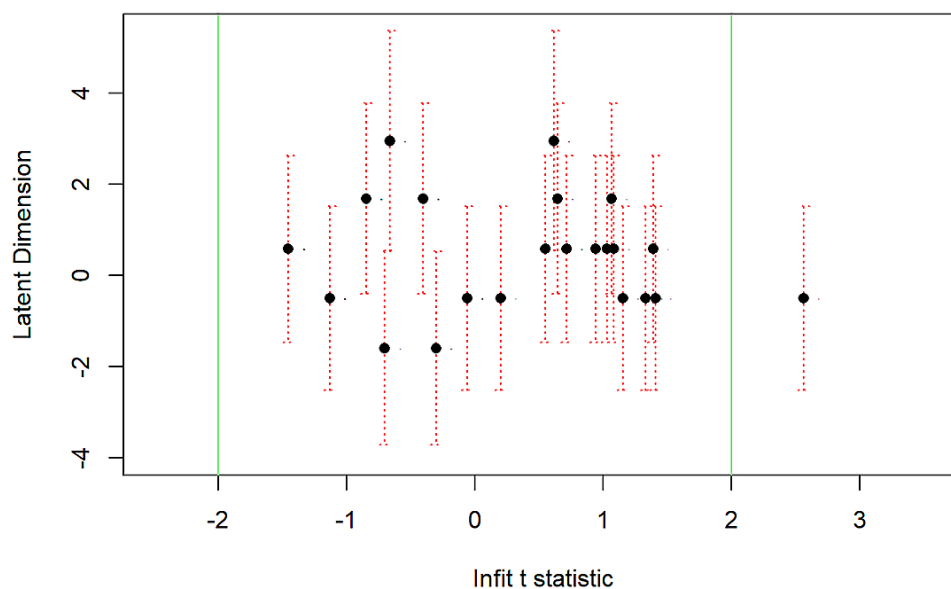


### Model Assumptions

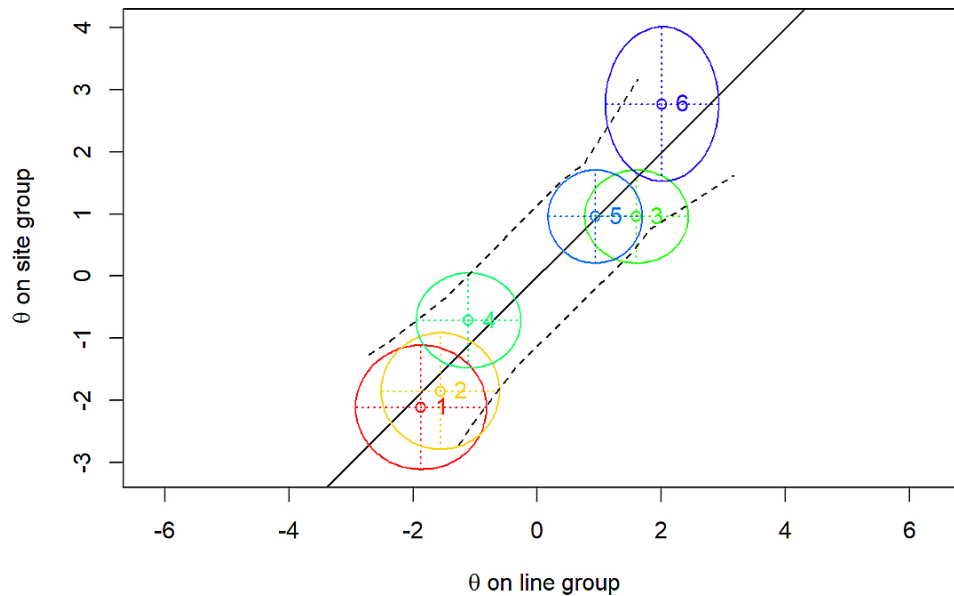
The nonsignificant T1m statistic in all pairs of questions ( $P>.05$ ) indicated that the practical examination was 1D. Pairwise analysis of item responses (Table S1 in [Multimedia Appendix 2](#)) showed no significant violations of unidimensionality, with only a minimal proportion of item pairs displaying residual dependence. The distribution of person fit statistics also showed acceptable fit across participants ([Figure 3](#)). The nonsignificant Andersen likelihood ratio test ( $\chi^2_3=1.2$ ;  $P=.75$ ) confirmed the invariance of the model parameters ([Figure 4](#)). Regarding local independence,

Ponocny T1 statistic revealed that only the pair “structure identification–patient position” showed a significant residual correlation ( $P=.02$ ). This dependency represented only 4.76% of the total item pairs, a figure within the expected margin of error (<5%) that did not compromise the model’s unidimensionality. This specific correlation was likely due to the procedural nature of the ultrasound examination, in which proper patient positioning serves as a technical prerequisite for successful structure identification. Therefore, this minor violation was considered clinically coherent and statistically negligible for the overall validity of the trait estimates ( $\theta$ ).

**Figure 3.** Person vs infit t-statistic plot displaying person ability parameters on the vertical axis and their respective infit t-statistics on the horizontal axis. The green vertical lines at -2 and 2 represent the thresholds for acceptable item fit.



**Figure 4.** Goodness of fit plot. The numbered colored circles (1-6) represent the location parameters (difficulty) of the 6 items, with circle sizes reflecting the confidence ellipses. The solid diagonal line represents perfect invariance (equal difficulty across groups). The dashed reference lines indicate the 95% confidence bounds. Because all item circles either cross or fall within these dashed boundaries, the item parameters are considered invariant, indicating no significant differential item functioning between the online and onsite student groups.



### Item Parameters

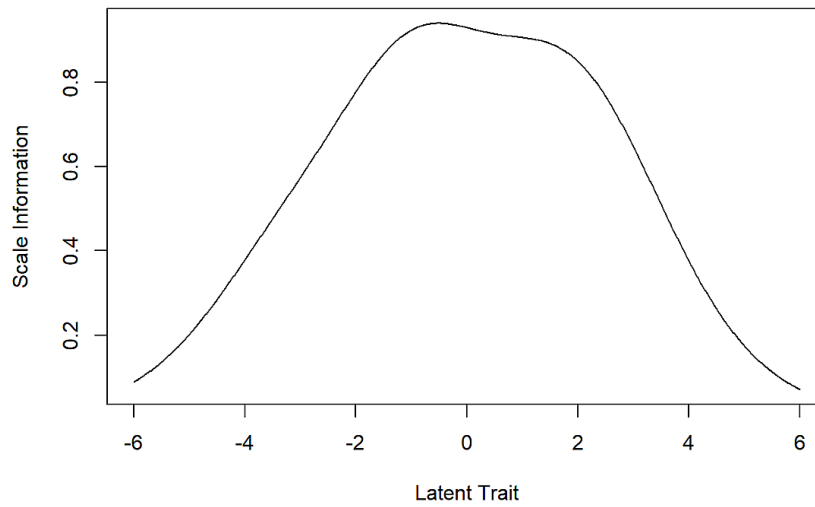
The examination presented an average level of difficulty  $\delta$  of 0 (SD 2.213) with a higher difficulty level in the questions on image optimization, structure diameter, and structure surface distance and lower for the questions on patient position, sonographer position, and probe handling (Table 3). It was observed that the test provided the maximum information with the minimum error at levels of  $\theta$  around  $-2$  to  $2$  (Figure 5). A clear division of the items was observed according to the level of difficulty; the easiest questions were on patient position and sonographer position; the questions with

an intermediate level of difficulty were on probe handling and structure identification; and the most difficult questions were on image optimization, structure diameter, and structure surface distance. The Wright map (Figure 6) illustrated a clear separation between lower- and higher-difficulty items, confirming the hierarchical progression of skill acquisition. The item characteristic curves for the seven assessment criteria are presented in Figure 7, showing the observed response patterns and the model-predicted probabilities across the latent ability continuum.

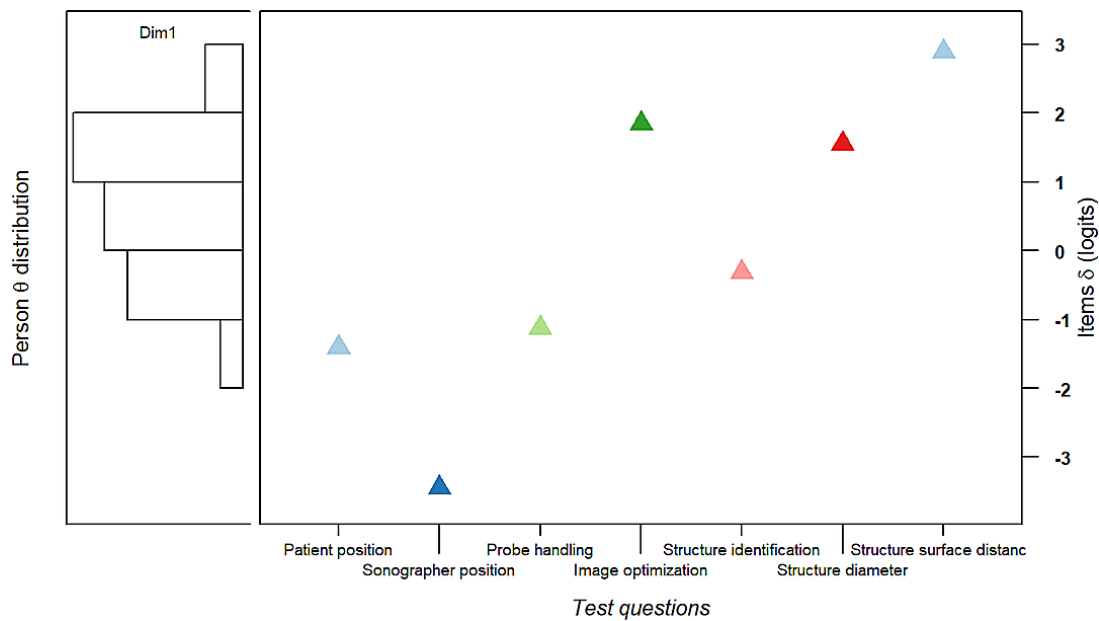
**Table 3.** Item difficulty parameters.

Item	Difficulty (SE; 95% CI)
Patient position	-1.407 (0.395; -2.181 to -0.634)
Sonographer position	-3.446 (0.876; -5.163 to -1.729)
Probe handling	-1.121 (0.366; -1.837 to -0.405)
Image optimization	1.845 (0.319; 1.22 to 2.469)
Structure identification	-0.312 (0.314; -0.928 to 0.303)
Structure diameter	1.548 (0.308; 0.944 to 2.152)
Structure surface distance	2.894 (0.399; 2.113 to 3.675)

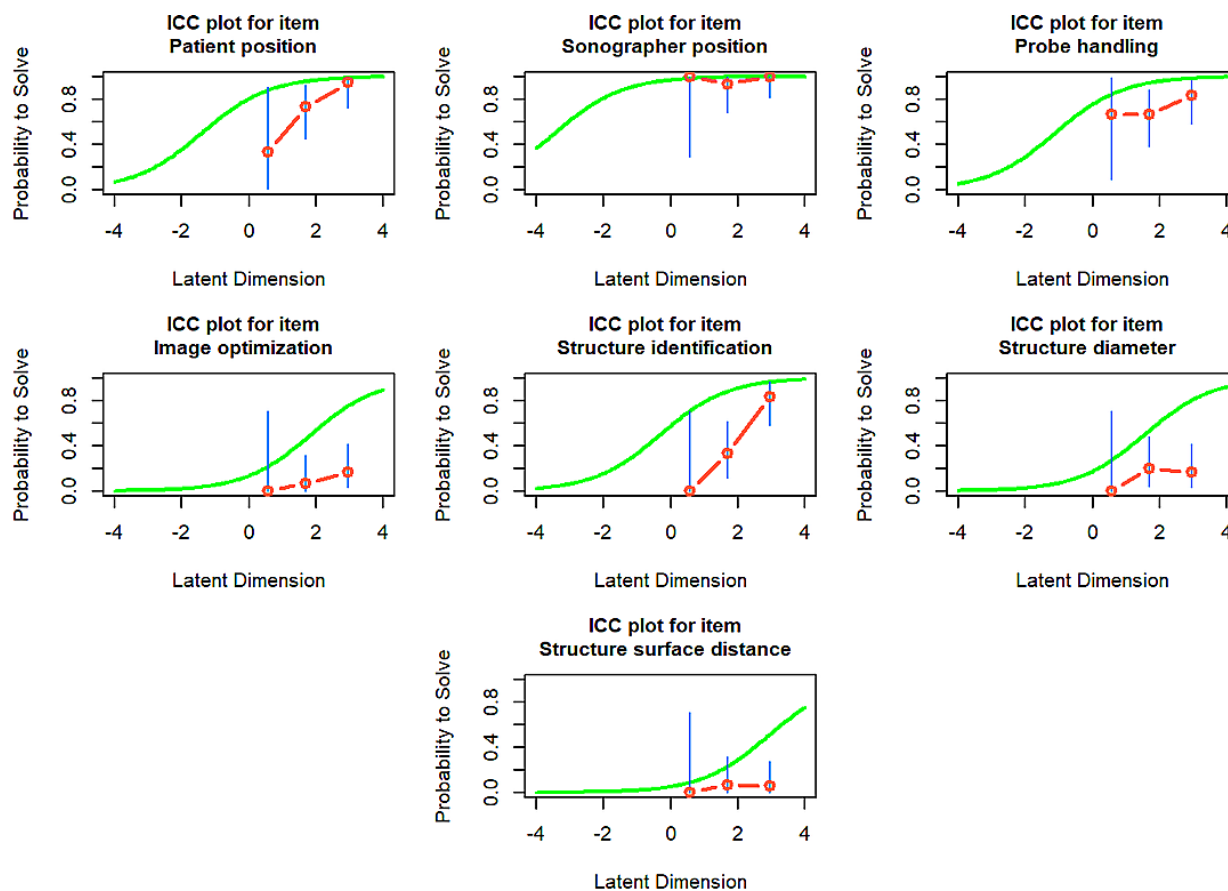
**Figure 5.** Test information plot illustrating the amount of psychometric information provided by the test across different levels of student ability (latent trait). The curve peaks and remains high between approximately -2 and +2, indicating that the assessment instrument achieves maximum measurement precision and reliability when evaluating students with average or moderate proficiency levels.



**Figure 6.** Wright map. The left panel displays a histogram of the participants' estimated latent proficiency ( $\theta$  distribution) along the primary latent dimension (Dim1). The right panel plots the difficulty parameters ( $\delta$  in logits) of the 7 assessment criteria, where each colored triangle represents a specific test question or criterion. Both panels share the central logit scale. Items located higher on the vertical axis (eg, "structure surface distance") represent greater difficulty, whereas triangles near the bottom (eg, "sonographer position") indicate easier tasks. This alignment allows direct evaluation of targeting, demonstrating that the test items cover a broad range of the students' proficiency distribution.



**Figure 7.** Item characteristic curves. The green solid lines represent the theoretical intraclass correlation coefficient (ICCs; model-predicted probability of a correct response across the latent dimension). The red squares and dashed lines show the observed empirical proportions for each score group, with blue vertical lines indicating the 95% CIs. Closer alignment between the empirical red points and the theoretical green curve reflects better fit of the item to the measurement model.



## Discussion

### Principal Findings

This randomized study found that simulation-based online ultrasound training produced comparable theoretical and practical outcomes to traditional on-site instruction. IRT analysis revealed a structured hierarchy of item difficulty and identified key competencies for assessment, while both training modalities showed limited reliability relative to expert performance, reflecting the early stage of skill acquisition.

The theoretical and practical assessment outcomes revealed comparable knowledge acquisition across both groups, with no statistically significant differences between the online and on-site training modalities. Notably, the online group demonstrated nearly half the error rate in image identification compared with the on-site cohort, suggesting enhanced visual-cognitive integration when using simulation-based e-learning.

A noteworthy finding was the poor absolute reliability (ICC) and the relatively high SEM and minimal detectable change values observed in both groups. We contextualize these results as a transparent reflection of the learning curve and the expert-novice gap inherent in high-precision technical measurements.

Because the ICC compared student performance against a gold standard expert, these low values underscore that a single training session—regardless of the modality (online or on-site)—is insufficient for achieving clinical-grade proficiency. These data are crucial for future pedagogical designs, as they highlight that while online training might slightly reduce measurement error (SEM), both formats require extended supervised practice to transition from conceptual understanding to technical mastery.

The IRT analysis enabled a detailed understanding of item difficulty within the practical assessment. The practical implications of these findings are 2-fold. First, the item difficulty hierarchy ( $\delta$ ) provides a roadmap for curriculum design: training should prioritize “image optimization” and “metric precision” (diameter and distance), as these were identified as the most challenging competencies to acquire ( $\delta > 1.5$ ). Second, the test information function reveals that the current assessment is highly reliable for discriminating between novice and intermediate students (latent trait between  $-2$  and  $+2$ ) but may require more challenging items to effectively assess advanced practitioners. Finally, the disparity between observed responses and the ICC curves (Figure 7) in high-difficulty items suggests that metric accuracy is a specific technical bottleneck that persists despite general procedural knowledge, highlighting the need for targeted feedback in ultrasound-guided measurements.

The insights provided by the IRT model allow for a data-driven optimization of the training program, focusing on 3 key areas.

First, regarding curriculum design, because “sonographer position” and “patient position” are low-difficulty items ( $\delta < -1.4$ ), they should be introduced as foundational self-learning modules. This allows classroom or simulation time to be shifted toward high-difficulty tasks such as “structure surface distance” ( $\delta = 2.89$ ), which require expert-led, hands-on guidance.

Second, assessment weighting and IRT difficulty parameters justify a nonlinear scoring system. We suggest that “image optimization” and “metric measurements” should carry higher weight in final assessments, as they are the primary discriminators of technical proficiency, whereas “patient positioning” serves only as a prerequisite.

Third, regarding targeted skill remediation, the item characteristic curves identified specific failure points. For instance, students who failed “structure identification” should not proceed to “structure diameter” training, as the model confirmed a hierarchical dependency. Remediation can thus be personalized based on each student’s position on the latent trait ( $\theta$ ) scale.

No significant differences were found between online and on-site groups in item difficulty patterns. Although ICCs between the 2 groups were nonsignificant and associated with relatively high SEMs, the online group demonstrated slightly better reliability, a finding consistent with previous research [12].

Prior studies have investigated the feasibility of online ultrasound education across various health professions, including obstetrics [9], anesthesia [10], physiotherapy [11, 12], and postgraduate echocardiography [13]. These works generally reported equivalent or superior outcomes for knowledge acquisition and image interpretation in comparison to traditional face-to-face methods. These findings are consistent with this body of evidence, supporting the educational validity of simulation-based e-learning for ultrasound training.

However, few previous investigations have objectively assessed psychomotor and image acquisition skills using psychometrically robust frameworks. To our knowledge, no earlier ultrasound education studies have used an IRT model to examine item difficulty and learner ability distribution. This methodological innovation provides a degree of measurement precision rarely applied outside high-stakes simulation domains, such as aviation [20]. Consequently, our results extend the current understanding of how simulation-based training can be objectively evaluated and standardized in medical education.

The findings suggest that digital simulation tools such as the platform can democratize ultrasound training by increasing accessibility and scalability, enabling a broader range of health care professionals to achieve comparable

levels of competence. As highlighted by Einstein’s well-known maxim, “doing the same thing and expecting different results” limits innovation [21]; the integration of digital solutions into medical curricula represents a necessary evolution aligned with modern health care and societal needs.

IRT-derived ability mapping indicated that most students clustered at ability levels 1 to 2 within a  $-3$  to  $+3$  range, a valuable reference point for designing adaptive assessment systems. Future studies may leverage these psychometric profiles to personalize training progression and feedback in simulation-based e-learning environments.

Finally, previous evidence demonstrated that online experiential learning models offer significant advantages in efficiency and cost reduction [8]. Combined with comparable learning outcomes, these results reinforce simulation-based digital education as a sustainable and effective approach for future ultrasound and clinical skills training.

Because this study was not prospectively registered, transparency may be limited. Future studies should ensure prior trial registration, use a multi-institution design, and examine a broader cohort.

## Conclusions

Online ultrasound training provides the same theoretical and practical results as face-to-face training, with all the economic, academic, and professional benefits that this entails. The most recommendable questions (“probe handling” and “structure identification”) for the evaluation of this training have been identified. It will be interesting in future studies to evaluate the capacity achieved according to the number of hours, as well as to conduct a cost-effectiveness study.

This study provides a preliminary validation of a practical assessment tool for ultrasound-guided procedures using IRT. The application of the 1-parameter Rasch model proved to be a robust methodological choice for this cohort ( $n=68$ ); its mathematical parsimony allowed for stable and reliable estimates of item difficulty and student proficiency even in samples of modest size, where more complex models might falter [22,23]. The nonsignificant Andersen likelihood ratio test and the low SEs further reinforced the internal validity of the findings within this specific educational context. However, given the single-site nature of the study and the specific sample size, these results should be interpreted as a foundational pilot validation. While the identified difficulty hierarchy offers a clear road map for curriculum design and targeted remediation, the generalizability and global scalability of the instrument remain to be confirmed. Future research should focus on multicenter studies with larger, more diverse populations to establish definitive normative data and ensure cross-cultural parameter invariance. In conclusion, this work represents a significant step toward a data-driven, standardized evaluation in ultrasound training, providing a reliable framework for initial competency assessment.

---

## Acknowledgments

The authors thank the participating physiotherapy students and the faculty members who facilitated data collection. The authors also thank Dr Josue Fernández-Carnero from King Juan Carlos University for his helpful contributions to this publication.

---

## Funding

This research received no external funding. The study was conducted without financial support from public agencies, commercial entities, or nonprofit organizations.

---

## Authors' Contributions

Conceptualization: SF-C

Data curation: BD-P, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Formal analysis: BD-P, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Funding acquisition: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Investigation: BD-P, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Methodology: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Project administration: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Resources: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Software: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Supervision: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Validation: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Visualization: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Writing—original draft: BD-P, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

Writing—review and editing: SF-C, BD-P, JM-R, DP-M, SG-M, AA-O, NC-Z, FN-C Sr

All authors read and approved the published version of the manuscript.

---

## Conflicts of Interest

The authors declare no commercial, financial, or personal relationships that could be perceived as potential conflicts of interest. The platform provided access to the digital ultrasound simulation platform used in this study; however, the company had no role in the study design, data collection, statistical analysis, interpretation of the results, or the decision to submit the manuscript for publication. No author received financial compensation for participating in this research. Several authors participated in the development of the ultrasound simulation tool. To minimize bias, all researchers involved in its design or programming were blinded to group allocation, performance metrics, and outcome data until all analyses were completed.

---

## Multimedia Appendix 1

Detailed item response theory analysis and model diagnostics.

[\[DOCX File \(Microsoft Word File\), 24 KB-Multimedia Appendix 1\]](#)

---

## Multimedia Appendix 2

Psychometric analyses and graphical outputs from the item response theory (IRT) Rasch model, including item characteristic curves, Wright maps, goodness-of-fit plots, and test information functions comparing online and on-site ultrasound training groups.

[\[DOCX File \(Microsoft Word File\), 32 KB-Multimedia Appendix 2\]](#)

---

## References

1. Minimum training requirements for the practice of medical ultrasound in Europe. European Federation of Societies for Ultrasound in Medicine and Biology. URL: <http://s834315022.websitehome.co.uk/wp-content/uploads/2020/12/2009-04-14apx12.pdf> [Accessed 2026-5-27]
2. Montesinos L, Salinas-Navarro DE, Santos-Diaz A. Transdisciplinary experiential learning in biomedical engineering education for healthcare systems improvement. *BMC Med Educ*. Apr 3, 2023;23(1):207. [doi: [10.1186/s12909-023-04171-x](https://doi.org/10.1186/s12909-023-04171-x)] [Medline: [37013525](https://pubmed.ncbi.nlm.nih.gov/37013525/)]
3. Andersen CA, Holden S, Vela J, Rathleff MS, Jensen MB. Point-of-care ultrasound in general practice: a systematic review. *Ann Fam Med*. Jan 2019;17(1):61-69. [doi: [10.1370/afm.2330](https://doi.org/10.1370/afm.2330)] [Medline: [30670398](https://pubmed.ncbi.nlm.nih.gov/30670398/)]
4. Chawla TP, Cresswell M, Dhillon S, et al. Canadian Association of Radiologists position statement on point-of-care ultrasound. *Can Assoc Radiol J*. Aug 2019;70(3):219-225. [doi: [10.1016/j.carj.2019.06.001](https://doi.org/10.1016/j.carj.2019.06.001)] [Medline: [31376885](https://pubmed.ncbi.nlm.nih.gov/31376885/)]
5. Nakashima H, Yukawa Y, Suda K, Yamagata M, Ueta T, Kato F. Abnormal findings on magnetic resonance images of the cervical spines in 1211 asymptomatic subjects. *Spine (Phila Pa 1976)*. Mar 15, 2015;40(6):392-398. [doi: [10.1097/BRS.0000000000000775](https://doi.org/10.1097/BRS.0000000000000775)] [Medline: [25584950](https://pubmed.ncbi.nlm.nih.gov/25584950/)]

6. Bacon DR, Cowles K, Thapa D, et al. Creating an ultrasound scholarly concentration program for medical students. *Adv Med Educ Pract.* 2021;12:1103-1110. [doi: [10.2147/AMEP.S330771](https://doi.org/10.2147/AMEP.S330771)] [Medline: [34594147](https://pubmed.ncbi.nlm.nih.gov/34594147/)]
7. Sommer KJ. Pilot training: what can surgeons learn from it? *Arab J Urol.* Mar 2014;12(1):32-35. [doi: [10.1016/j.aju.2013.08.011](https://doi.org/10.1016/j.aju.2013.08.011)] [Medline: [26019919](https://pubmed.ncbi.nlm.nih.gov/26019919/)]
8. Maloney S, Haas R, Keating JL, et al. Breakeven, cost benefit, cost effectiveness, and willingness to pay for web-based versus face-to-face education delivery for health professionals. *J Med Internet Res.* Apr 2, 2012;14(2):e47. [doi: [10.2196/jmir.2040](https://doi.org/10.2196/jmir.2040)] [Medline: [22469659](https://pubmed.ncbi.nlm.nih.gov/22469659/)]
9. Shaw-Battista J, Young-Lin N, Bearman S, Dau K, Vargas J. Interprofessional obstetric ultrasound education: successful development of online learning modules; case-based seminars; and skills labs for registered and advanced practice nurses, midwives, physicians, and trainees. *J Midwife Womens Health.* Dec 2015;60(6):727-734. [doi: [10.1111/jmwh.12395](https://doi.org/10.1111/jmwh.12395)] [Medline: [26769384](https://pubmed.ncbi.nlm.nih.gov/26769384/)]
10. Krishnan S, Kuhl T, Ahmed W, Togashi K, Ueda K. Efficacy of an online education program for ultrasound diagnosis of pneumothorax. *Anesthesiology.* Mar 2013;118(3):715-721. [doi: [10.1097/ALN.0b013e31827f0979](https://doi.org/10.1097/ALN.0b013e31827f0979)] [Medline: [23291625](https://pubmed.ncbi.nlm.nih.gov/23291625/)]
11. Fernández-Lao C, Cantarero-Villanueva I, Galiano-Castillo N, Caro-Morán E, Díaz-Rodríguez L, Arroyo-Morales M. The effectiveness of a mobile application for the development of palpation and ultrasound imaging skills to supplement the traditional learning of physiotherapy students. *BMC Med Educ.* Oct 19, 2016;16(1):274. [doi: [10.1186/s12909-016-0775-1](https://doi.org/10.1186/s12909-016-0775-1)] [Medline: [27756288](https://pubmed.ncbi.nlm.nih.gov/27756288/)]
12. Fernández-Carnero S, Cuenca-Zaldivar JN, Pecos-Martin D, Achalandabaso-Ochoa A, Ferragut-Garcias A, Gallego-Izquierdo T. Validity and reliability of methods for sonography education in physiotherapy: onsite vs. online. *Appl Sci.* 2020;10(18):6456. [doi: [10.3390/app10186456](https://doi.org/10.3390/app10186456)]
13. Vegas A, Meineri M, Jerath A, Corrin M, Silversides C, Tait G. Impact of online transesophageal echocardiographic simulation on learning to navigate the 20 standard views. *J Cardiothorac Vasc Anesth.* Jun 2013;27(3):531-535. [doi: [10.1053/j.jvca.2012.09.006](https://doi.org/10.1053/j.jvca.2012.09.006)] [Medline: [23127696](https://pubmed.ncbi.nlm.nih.gov/23127696/)]
14. DeMars C. *Item Response Theory.* Oxford University Press; 2010. ISBN: 9780199847341
15. Kruyen PM. Using short tests and questionnaires for making decisions about individuals: when is short too short [PhD thesis]. Tilburg University; 2012. URL: <https://research.tilburguniversity.edu/en/publications/using-short-tests-and-questionnaires-for-making-decisions-about-i/> [Accessed 2026-05-14]
16. Hambleton RK. Item response theory: introduction and bibliography. *Psicothema.* 1990;2:97-107. URL: <https://reunido.uniovi.es/index.php/PST/article/view/7061> [Accessed 2026-06-17]
17. Ponocny I. Nonparametric goodness-of-fit tests for the Rasch model. *Psychometrika.* Sep 2001;66:437-459. [doi: [10.1007/BF02294444](https://doi.org/10.1007/BF02294444)]
18. WazO. URL: <https://wazotechnologies.com/> [Accessed 2023-07-05]
19. Moher D, Schulz KF, Altman DG, Lepage L. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet.* Apr 14, 2001;357(9263):1191-1194. [Medline: [11323066](https://pubmed.ncbi.nlm.nih.gov/11323066/)]
20. Arjoni DH, de Souza Rehder I, Pereira Figueira JM, Villani E. Augmented reality for training formation flights: an analysis of human factors. *Heliyon.* Mar 2023;9(3):e14181. [doi: [10.1016/j.heliyon.2023.e14181](https://doi.org/10.1016/j.heliyon.2023.e14181)] [Medline: [36950645](https://pubmed.ncbi.nlm.nih.gov/36950645/)]
21. Noseworthy T. A flight simulator for careers in health services and policy research. *Healthc Policy.* Oct 2019;15(SP):6-9. [doi: [10.12927/hcpol.2019.25984](https://doi.org/10.12927/hcpol.2019.25984)] [Medline: [31755855](https://pubmed.ncbi.nlm.nih.gov/31755855/)]
22. Linacre JM. Sample size and item calibration [or person measure] stability. *Rasch.org.* URL: <https://www.rasch.org/rmt/rmt74m.htm> [Accessed 2026-05-14]
23. Stone M, Yumoto F. The effect of sample size for estimating Rasch/IRT parameters with dichotomous items. *J Appl Meas.* 2004;5(1):48-61. [Medline: [14757991](https://pubmed.ncbi.nlm.nih.gov/14757991/)]

## Abbreviations

**CONSORT:** Consolidated Standards of Reporting Trials

**ICC:** intraclass correlation coefficient

**IRT:** item response theory

**SEM:** standard error of measurement

*Edited by Lorainne Tudor Car; peer-reviewed by Pelumi Adedigba, Rafael Tezza; submitted 17.Nov.2025; final revised version received 21.Mar.2026; accepted 22.Apr.2026; published 02.Jul.2026*

*Please cite as:*

*Fernández-Carnero S, Diaz-Pulido B, Méndez-Rodríguez J, Pecos-Martin D, Garcia-Miguel S, Achalandabaso-Ochoa A, Cuenca-Zaldívar N, Naranjo-Cinto Sr F*  
*Digital Simulation–Based Ultrasound Training for Physiotherapy Students: Prospective Randomized Controlled Trial*  
*JMIR Med Educ 2026;12:e87897*  
URL: <https://mededu.jmir.org/2026/1/e87897>  
doi: [10.2196/87897](https://doi.org/10.2196/87897)

©Samuel Fernández-Carnero, Belen Diaz-Pulido, Jorge Méndez-Rodríguez, Daniel Pecos-Martin, Santiago Garcia-Miguel, Alexander Achalandabaso-Ochoa, Nicolas Cuenca-Zaldívar, Fermín Naranjo-Cinto Sr. Originally published in JMIR Medical Education (<https://mededu.jmir.org>), 02.Jul.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Education, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.