ORIGINAL ARTICLE

Psychometric Properties of Self-report Questionnaires in Evaluating Blended Learning in Health Science University Students: A Systematic Review

Valentin C. Dones III, PhD, MSPT, PTRP,[1]* Maria Teresita B. Dalusong, MHPEd, PTRP,[2]*
Donald G. Manlapaz, PhD, MSPT, PTRP,[1] Juan Alfonso S. Rojas,[3] Ma. Bianca Beatriz P. Ballesteros,[3]
Ron Kevin S. Flores,[3] Kaela Celine C. Ho,[3] Jose Angelo D. Monreal,[3] Audrey Marie A. Narcelles,[3]
Jose Joaquin R. Reyes[3] and Lianna Andrea B. Sangatanan[3]

[1]*Center for Health Research and Movement Science – A JBI-Affiliated Group, College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*
[2]*College of Physical Therapy, Pamantasan ng Lungsod ng Maynila, Manila, Philippines*
[3]*College of Rehabilitation Sciences, University of Santo Tomas, Manila, Philippines*

ABSTRACT

Background. Due to the COVID-19 outbreak, schools had to switch online. The sudden transition to blended teaching and learning (BTL) poses challenges for students and teachers, especially for health science programs that require hands-on practical experience. The validity, reliability, and responsiveness of these self-report questionnaires (SRQs) should be established to ensure the accuracy of the results as intended by the SRQ.

Objectives. This study critically appraised, compared, and summarized the psychometric properties of SRQ evaluating BTL among health science university students. This review determined the SRQ's reliability, internal consistency, various forms of validity (content, criterion, construct), and responsiveness.

Methods. Following a 10-step procedure based on COSMIN guidelines, we conducted a systematic review of SRQs used by health science university students to evaluate blended teaching and learning. Studies were eligible if they reported psychometric properties of SRQs related to blended learning among university health science students; exclusions included studies focusing on perceptions, attitudes, self-efficacy, and satisfaction, as well as articles such as biographies, editorials, and conference materials. Searches covered multiple electronic databases until April 26, 2023, including PubMed, EMBASE, Web of Science, MEDLINE (OVID), PsycInfo, CINAHL, EBSCOHOST, ERIC, Scopus, Science Direct, Google Scholar, JSTOR, Acta Medica Philippina, Philippine Journal of Health Research and Development, and HERDIN, managed through Zotero. Two independent reviewers performed database searches, title and abstract screening, and full-text evaluations, with a third reviewer resolving any disputes. The COSMIN Risk of Bias Checklist was employed to evaluate included studies on the development and various measurement properties of SRQs. The reviewers assessed SRQ standards, including validity, reliability, internal consistency, measurement error, responsiveness, interpretability, and feasibility. Data extraction and result tabulation were independently completed, with content comparison by two health education experts. This evaluation categorized the SRQs into three quality and validity levels.

** Professor Dones and Associate Professor Dalusong share first authorship for this manuscript.*

Paper presentation – World Physiotherapy Congress, June 3, 2023, Dubai.

Corresponding author:
Valentin C. Dones, III, PhD, MSPT, PTRP
Center for Health Research and Movement Science –
A JBI-Affiliated Group, College of Rehabilitation Sciences
University of Santo Tomas
España Boulevard, Sampaloc, Manila 1015, Philippines
Email: vcdones@ust.edu.ph
ORCiD: https://orcid.org/0000-0002-0048-4995

Results. The study examined five articles; four were rated as 'doubtful' and one as 'inadequate' in the overall development of SRQ. All four 'doubtful' studies demonstrated questionable content validity when university students were asked about the questionnaire's relevance, comprehensiveness, and comprehensibility. Only half of these studies achieved an 'adequate' rating for content validity based on expert opinions on relevance and comprehensiveness. All but one study scored from 'very good' to 'adequate' in structural validity. Three out of the four studies scored a very good rating for internal consistency, while one was deemed 'inadequate' in internal consistency, cross-cultural validity, and reliability. Three out of four studies scored 'very good' on construct validity, but all overlooked criterion validity and responsiveness. Conducted in various locations, including Australia, Romania, Turkey, and Taiwan, these studies highlighted both common characteristics and limitations in questionnaire development according to the COSMIN guidelines. Four studies were deemed reliable and valid for BTL constructs (Category A); Wu et al. requires further validation (Category B). Study limitations included heterogeneity in populations, settings, and questionnaire versions, potential subjective bias in SRQ content comparison, and the evolving nature of SRQs in blended learning contexts.

Conclusion. The systematic review reports the development and evaluation of SRQs for BTL while identifying gaps in their applicability to health science programs. The Blended Learning Scale (BLS) of Lazar et al. and the Blended Learning Questionnaire (BLQ) of Ballouk et al. showed an 'adequate' rating for content validity. BLS revealed very good structural validity, internal consistency, and adequate content validation. Although the BLQ lacked Confirmatory Factor Analysis, it yielded valuable constructs for evaluating health sciences students' experiences in BTL. Both tools require improvements on recall period, completion time, interpretability, and feasibility. The review underscores the necessity for continuous assessment and enhancement of such instruments in BTL, advocating a rigorous scale development process. Furthermore, it encourages the customization of teaching and learning evaluation tools to suit specific institutional contexts while promoting further validation of these questionnaires across different populations in future research.

Keywords: psychometrics, checklist, self report, universities, health education

INTRODUCTION

Amidst the COVID-19 pandemic, blended teaching and learning (BTL) has become a prevailing educational model for allied health science university students worldwide. BTL, characterized as the intentional integration of traditional face-to-face classroom instruction with online learning experience, has provided educational institutions with a flexible approach to adapting to the challenges posed by the pandemic.[1–3] This approach ensures the safety of allied health science students, faculty, and staff by minimizing in-person interactions.[4] Albeit, BTL has implications for health science university students who rely on direct patient care for professional practice.[5,6]

BTL effectively provide quality education that facilitates easy access, flexibility, and self-paced and cost-efficient learning.[7] Integrating blended systems overcomes time and space, encourages independence in academic exploration, improves health science university student participation, and reaches more health science university students without increasing resource demands.[8–11] However, the abrupt shift to BTL without prior preparation is associated with adaptability struggles, computer literacy problems, connectivity problems, and difficulty practicing hands-on skills.[7,12]

SRQs are essential to measure the effectiveness and quality of BTL. SRQs, completed by students as the direct recipients, can effectively assess BTL's effectiveness through valid and reliable measures.[13] Assessing students' perceptions of BTL's ease of use, satisfaction, and effectiveness in achieving learning outcomes is crucial. This evaluation enables teachers to gauge BTL's impact on students' knowledge, skills, and attitudes.[14] The validity, reliability, and responsiveness of these SRQs should be established to ensure the accuracy of the results as intended by the SRQ. High-quality SRQS should have robust psychometric quality.[15]

To optimize the effectiveness of BTL in health professional education, accurate assessment of students' perceptions of BTL is crucial. Although SRQs such as the Classroom Environment Questionnaire, Dental Clinical Learning Environment Instrument, and Student Course Experience Questionnaire have assessed students' perceptions of learning environments, they were originally designed for traditional face-to-face settings.[16–19] In BTL, students alternate between in-person and online learning, significantly influencing their learning approach, use of online technologies, and engagement with materials.[13] BTL resulted in various changes in instructional methods, such as asynchronous and synchronous modes.

The intricacies of the BTL environment should be considered when evaluating the quality of course and teaching because the traditional self-report questionnaire used in the classroom does not adequately reflect the essentials of blended and online teaching practices since these tools do not consider the online aspect incorporated in BTL.[20,21] Unlike teaching evaluation in a classroom setting, there are

few underdeveloped SRQs for online courses and teacher quality in an online or BTL environment in health science programs in a university setting.[22–26] The SRQ to evaluate quality teaching delivery and performance must be reviewed, modified, or created as health science programs adopt BTL.[27–29] No synthesized studies have been performed on the quality of their psychometric properties, especially on SRQs on BTL in health science programs.

We critically appraised, compared, and summarized the psychometric property scores of SRQ evaluating the quality of BTL delivery among health science university students. Specifically, this review determined these measurement properties of the SRQs used to assess BTL:

1. Reliability, measurement error, internal consistency,
2. Content validity, criterion validity, construct validity, and
3. Responsiveness.

METHODS

Study Design

The study is a systematic review and meta-analysis of SRQs used by health science university students to evaluate BTL. The review followed the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. This research was exempted from ethics review by the University of the Philippines – Manila with protocol number: UPMREB 2022-0259-EX. The study has a PROSPERO registration ID number of CRD42022372362. The published protocol is accessible in BMJ Open.[30]

The researchers followed a 10-step procedure to systematically review SRQs using Consensus-based Standards for the selection of Health Measurement Instruments (COSMIN).[31] Two independent reviewers evaluated the self-report questionnaire used by health science university students to evaluate BTL. They reviewed its content validity, the internal structure, including structural validity, internal consistency, and cross-cultural validity. Furthermore, they examined its reliability, measurement error, criterion validity, construct validity (through hypotheses testing), and responsiveness. Figure 1 presents the COSMIN 10-step procedure, outlining the sequential steps from formulating the research question to reporting findings. Each stage in the flowchart provides a concise overview of the systematic review process, illustrating the methodological approach employed to synthesize evidence on the psychometric properties of self-report questionnaires in blended learning for health science education.

Eligibility Criteria

Studies that met the following inclusion criteria were included in the systematic review:

a. Study population consisting of health science university students (i.e., medicine, physical therapy, occupational therapy, speech language pathology, psychology, nutrition, or nursing) engaged in BTL;
b. Studies reporting the development of SRQs for health science university students evaluating the quality of the BTL self-report questionnaire in all aspects (perspective, satisfaction, or delivery). We considered all scale development studies of self-reported questionnaires that measure specific aspects of BTL;

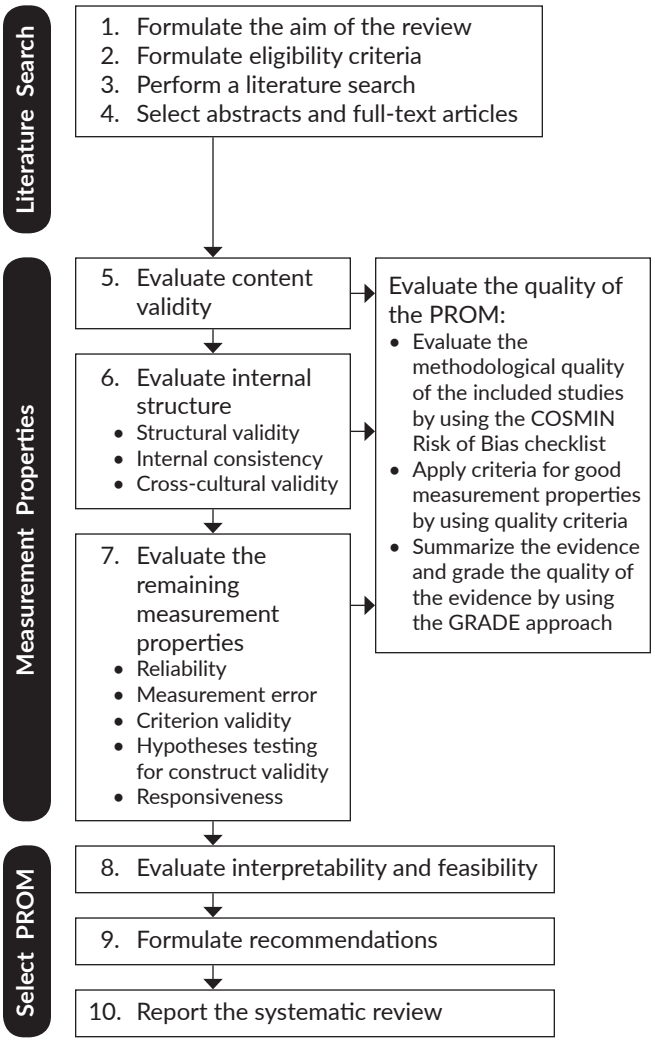

Figure 1. The 10-step procedure for conducting a systematic review and meta-analyses of patient-reported outcome measures by COSMIN.[31]

COSMIN – Consensus-based Standards for the selection of Health Measurement Instruments, GRADE – Grading of Recommendations Assessment, Development and Evaluation, PROMs – patient-reported outcome measures

c. Studies determining the psychometric properties of SRQs for health science university students on BTL; and
d. Studies reporting the distribution of scores, percentage of missing items, floor and ceiling effects, the availability of scores and change of scores, or the minimal important difference of self-reported questionnaires used in evaluating BTL delivery among health sciences university students.
e. Studies in any language were included in the review.

These were the exclusion criteria: a. studies reporting student perception, attitudes, learning experience, self-efficacy, satisfaction, and learning outcomes in BTL without the use of any SRQs; and b. biographies, case reports, editorials, newspaper articles, handout, consensus development conference, practice guidelines, short communications, abstracts, and meetings.

Literature Search

These databases were searched: PubMed, EMBASE, Web of Science, MEDLINE (OVID), PsycInfo (via ProQuest), CINAHL, EBSCOHOST, ERIC, Scopus, Science Direct, Google Scholar, JSTOR, Acta Medica Philippina, Philippine Journal of Health Research and Development, and HERDIN, including all dates up to April 26, 2023. Zotero software was used to manage the references in each database.

We used the following search terms for these three important concepts:

a. Context (Blended Teaching and Learning): blended learning OR delivery blended teaching OR learning flexible learning.
b. Population (Health Science University Students): Students, Medical OR medical student OR medicine student OR intern OR interns.
c. Construct (Psychometric Properties including reports on validity and reliability): instrumentation OR methods OR validation studies OR comparative study OR psychometrics.

The concepts were sequentially combined using the Boolean operator "AND". Study designs that were part of our exclusion criteria were also included in the search using the Boolean operator 'NOT'.

Search terms were adapted from the COSMIN search filter to find studies on measurement properties and search blocks of the Biomedische Informie (BMI) group.[31,32] COSMIN offers methods and practical tools to find and select the proper outcome measurement instrument to evaluate BTL.[31] Search blocks are a compilation of search strategies developed by Dutch medical information specialists. Search strings make it easier to begin a search, identify relevant terms and variations, and perform searches efficiently.[32]

An initial pilot search was conducted across various databases to identify relevant articles for our systematic review. This search yielded a significant number of hits: 272 from PubMed with 55 potentially relevant papers, 54 from EBSCO with eight of interest, three from ProQuest with one pertinent article, two from Google Scholar with one potentially relevant, and 569 from ScienceDirect with 55 of interest. These results underscore the feasibility of our systematic review by confirming the availability of a satisfactory volume of articles that meet our inclusion criteria for thorough evaluation.

Abstract and Full-text Article Selection

Two independent reviewers initially searched databases for studies with relevant titles and abstracts. Using eligibility criteria, two other independent reviewers screened the titles and abstracts of the studies and agreed on possibly relevant articles for this systematic review. Another two independent reviewers evaluated the relevance of the original articles included in the full-text articles. At all stages of the review process, the reviewers reached a consensus through discussion. A third independent reviewer was available for arbitration.

The COSMIN Risk of Bias Checklist[4]

Two reviewers independently assessed the methodological quality of each included study using the COSMIN Risk of Bias Checklist. The COSMIN Risk of Bias Checklist has ten boxes comprising 1. Standards for self-report questionnaire development, 2. Content validity, 3. Structural validity, 4. Internal consistency, 5. Cross-cultural validity, 6. Reliability, 7. Measurement error, 8. Criterion validity, 9. Hypothesis testing for construct validity, and 10. Responsiveness. Content validity assesses whether the SRQ's content adequately reflects the intended construct. Structural validity examines how well SRQ scores represent the construct's dimensionality, often using factor analysis. Internal consistency reviews the interrelatedness of items within a scale or subscale, typically measured by Cronbach's alpha. Cross-cultural validity, or measurement invariance, checks if the SRQ functions similarly across various cultural groups. Reliability focuses on score consistency over time when the construct remains stable. Measurement error assesses both systematic and random errors in individual scores, which may hinder accurate detection of real changes. Criterion validity evaluates how closely SRQ scores align with a recognized "gold standard," while construct validity involves testing hypotheses about the SRQ's relationship with other measures. Lastly, responsiveness gauges the SRQ's ability to detect meaningful changes over time in the construct being measured. It was not necessary to complete the checklist in the study as boxes to be evaluated were solely based on the measurement properties that the study authors evaluated. Independent reviewers evaluated all possible measurement properties extracted from the SRQ investigated in each study.[31]

Evaluation of the Content Validity of the Self-report Questionnaire

Two independent reviewers in sequence evaluated the standards for the development of the self-report questionnaire, assessed the quality of content validity studies of the SRQ, determined the overall content validity rating per SRQ, and assessed the quality of evidence per SRQ. The standards for the development of the SRQ considered general design, concept elicitation, cognitive interview, or pilot test of the self-report questionnaire. The quality of the content validity studies of the SRQ included an assessment of the relevance, comprehensiveness, and comprehensibility of the revised SRQ.[33] Ratings were assigned based on evidence of quality: 'Very good' for adequate quality, 'Adequate' if quality was assumable, 'Doubtful' if quality was unclear, 'Inadequate' for poor quality, and 'Not applicable' if not required. The "worst score counts" method is a key principle in the COSMIN checklist, where the overall quality rating for a study's measurement property is determined by its lowest score across all standards. This approach prevents methodological weaknesses from being offset by strengths in other areas; thus, a single "fatal flaw" can lower a study's rating. For example, a reliability study lacking independent administrations would receive an "inadequate" rating despite other strengths. Consistently applied across COSMIN boxes, this method enforces quality by addressing even minor flaws. Certain standards, however, have nuanced thresholds like "doubtful" or "adequate," reflecting their relative importance, especially in complex properties like construct validity. By requiring consistently high standards, the "worst score counts" method promotes trust in study results used in healthcare decisions.[33]

Using the guide to give a sufficient (+) rating for the 10 criteria for good content validity of a SRQ, the reviewers determined the overall content validity rating per SRQ.[33] Each criterion was rated as sufficient (+), insufficient (-), indeterminate (?) or inconsistent (±). A sufficient rating was given when ≥85% of the questionnaire items met the criterion, insufficient when <85% of the items met the criterion, and indeterminate when the information was inadequate, or the quality was poor. The guide to determine the relevance rating, the comprehensive rating, and the comprehensive rating per study, was used to evaluate the quality of each study included in the systematic review. The overall rate per study were determined using the guide for determining the content validity rating per study.[32]

The Modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Approach measures the trustworthiness of the overall content validity ratings. Quality was rated high, moderate, low, or very low based on study design, risk of bias, inconsistency, and indirectness. The modified GRADE approach evaluates the quality of evidence for PROM measurement properties in systematic reviews, providing a confidence level that pooled or summarized results are reliable. This system categorizes evidence quality as high, moderate, low, or very low, with potential downgrading based on risk of bias, inconsistency, imprecision, and indirectness. Starting from an assumption of high quality, downgrades are applied for "serious" (one level) or "very serious" (two levels) concerns. In extreme cases, a single inadequate study may reduce the rating by three levels. Downgrades for each factor follow specific guidelines: risk of bias is evaluated with the COSMIN checklist, inconsistency with unexplained contradictions in study results, imprecision by sample size, and indirectness by the relevance of study populations or settings. Accumulating downgrades across factors determines the final grade. While the modified GRADE approach provides a structured guide, reviewers apply their judgment in its use.[33]

If there was high-quality evidence that the content validity of an SRQ was insufficient, the SRQ was not further considered to evaluate the internal structure, the remaining measurement properties, the interpretability, and feasibility of the self-report questionnaire. Independent reviewers could directly draw a recommendation 'C' for this self-report questionnaire (that is, the self-report questionnaire with high evidence of insufficient content validity). Questionnaires with inadequate content validity were excluded, while those with at least 'doubtful' content validity were further evaluated.[31]

Evaluation of the Internal Structure of the Self-report Questionnaire

Reviewers used COSMIN to evaluate the measurement properties of SRQs, marking the same box multiple times if necessary for different sample populations. They verified structural validity, internal consistency, and cross-cultural validity using a rating system ranging from "very good quality" to "not applicable" based on the "worst score counts" method. Once the risk of bias was assessed, independent reviewers extracted and compared the results of the measurement properties of the SRQs used by health science university students to evaluate BTL against established good measurement property criteria. The evidence was then summarized for each measurement property of each questionnaire.[33]

Evaluation of the Remaining Measurement Properties of the Self-report Questionnaire

In COSMIN, certain measurement properties were assessed for the overall quality of a scale or subscale, rather than the individual items. These properties include reliability, measurement error, criterion validity, construct validity through hypothesis testing, and responsiveness. The reviewers decided whether there was a gold standard to measure the construct of interest in the target population. If there was no gold standard, the reviewers did not use the box for criterion validity and the box for the criterion approach for responsiveness. Instead, the reviewers formulated a set of hypotheses about the expected direction (i.e., positive or negative) and magnitude (i.e., absolute or relative) between the self-report questionnaire under review and other well-defined comparator questionnaires used in assessing BTL.

The study results were compared with the hypotheses of the reviewers.[33]

Evaluation of Interpretability and Feasibility of the Self-report Questionnaire

Interpretability and feasibility are critical, albeit informal, aspects of SRQs. Interpretability involves understanding the meaningfulness of scores, which can be improved by analyzing the score distribution, the minimum important change values, and the details of the response change. Feasibility assesses the practicality of administering these questionnaires considering factors such as time, cost, and ease of use. Although these are not official measurement attributes of a self-report questionnaire, these elements guide the selection of the most suitable instrument for a specific purpose.[33]

Data Extraction

To avoid missing relevant data, two independent reviewers, who critically evaluated the articles, extracted these data from the studies: a) characteristics of the included SRQs of health science university students evaluating BTL (i.e., constructs, target population, recall period, number of items, response options, scoring, original language, available translations), b) characteristics of health science university students (i.e., number of samples, age, sex, setting, country, language, response rate). The characteristics of the study samples had all information important for the generalizability of the results, emphasizing the similarity or dissimilarity of the study samples.

Two independent reviewers extracted, if available, the minimal important change values and the response shift. They extracted, if available, from the studies, the student's comprehensibility, type and ease of administration, length of the instrument, completion time, ease of standardization, ease of scoring calculation, copyright, cost, availability in different settings, and the approval requirement of the regulatory agency.

Summary of Findings Table

The two independent reviewers tabulated all the results by measurement property of the SRQ. The table has the summary result, overall rating (+/-/±/?), and the quality of the evidence (i.e., high, moderate, low, very low). The summary of findings table recommended the most appropriate SRQs for use by health science university students in evaluating BTL.

Formulation of Recommendations

Using the COSMIN guideline for recommendations, the SRQs were divided into three categories. The Category A SRQ showed satisfactory content validity and sufficient internal consistency, making them reliable and valid. The Category B SRQs, which did not fit either Category A or Category C, need further validation. Temporary recommendations can be made for these SRQs, subject to additional evidence. The Category C self-report questionnaire, characterized by insufficient measurement properties, should be excluded from the recommendations. The categorization process is non-arbitrary, requiring justification and possible validation guidelines. The recommendations are based on evaluations of measurement properties, interpretability, and feasibility, with the aim of approving suitable SRQs.[33]

RESULTS

Our search strategy yielded 205 studies from the electronic literature search. After duplicates were removed, 190 titles and abstracts were selected. One hundred seventy-one (171) studies were excluded after screening the titles and abstracts, leaving 19 studies for full-text review. Of the 19 full-text articles screened, five studies met our inclusion criteria. The study selection procedure is depicted in the PRISMA flow diagram (Figure 2).

Characteristics of Included Self-report Questionnaires

Table 1 compares SRQs on diverse BTL constructs for university students.[34–38] Akkoyunlu and Yilmaz-Soylu assesses BTL with 50 items on a 10-point Likert scale.[34] Ballouk

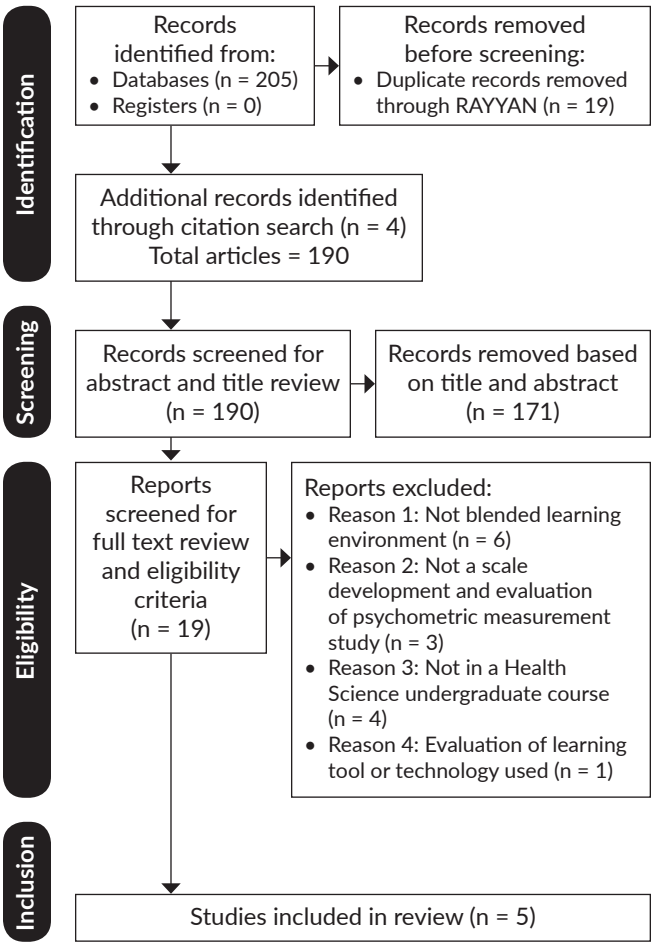

Figure 2. PRISMA Flowchart.

Table 1. Characteristics of the Included Self-report Questionnaires

Authors (Year) and Self–reported Instrument	Construct	Subscale/ Domain(s)/ Factors/ Dimensions (Items)	SRQ Purpose	No. of Items	Target Population	Recall Period	Response Options	Time To Administer (Minutes)	Original Language	Available Translations
Ballouk et al. (2022)[35] Blended Learning Questionnaire	Students' self-regulated learning in blended environment	**4 Subscales/ Factors:** 1: Resources: access and guidance (4) 2: Learning behavior: social and contextual (5) 3: Resources: delivery of content online (accessibility) (6) 4: Motivation: intrinsic and extrinsic (4)	Explore how medical students utilize self-regulated learning in blended environment	19	120 Medical Students	NR	7-point Likert Scale (ordinal) 1 = Not at all true of me 7= Very true of me	NR	English	English
Lazar et al. (2020)[37] Blended Learning Scale	Factors that influence students' use of digital tools in blended learning in HEI	**7 Dimensions:** 1. Familiarity with high-tech digital tools (modern) and familiarity with classical digital tools (traditional) (11) 2. Perceived barriers (5) 3. Computer anxiety (4) 4. Perceived usefulness (5) 5. Perceived ease of use (9) 6. Behavioral intention of use (3)	Determine students' perception and intention, and factors that influence their decision regarding digital tools in blended learning environment	36	University undergraduate students: Pilot Testing: 412 Engineering EFA: 352 Education Science Studies CFA: 414 Psychology studies Final: 511 Educ Science and Environmental Science	NR	5-point Likert Scale 1= Completely disagree 5= Completely agree	NR	English	English
Karadenyz (2012)[36] Student Satisfaction with the Blended Learning Questionnaire	Perceived usability Perceived usefulness of e-content and course structure Instructor support	**3 Dimensions:** 1. Perceived usability of the learner interface (5) 2. Perceived usefulness of e-content and course (8) 3. Instructor support (5)	To determine student satisfaction in blended learning environments where e-learning is supported by face-to-face learning	18	760 out of 1058 students from Bahçeşehir University enrolled in the "history of revolution" course	NR	5-point Likert Scale 1 = Strongly Disagree 5 = Totally Agree	NR	Turkish	None
Wu et al. (2010)[38] Blended e-Learning Scale	Student satisfaction in blended environment	**7 Dimensions:** 1. Computer self-efficacy; (3) 2. Performance expectations (3) 3. System functionality (3) 4. Content feature (2) 5. Interaction (3) 6. Learning climate (2) 7. Learning satisfaction (4)	Examine the determinants of student learning satisfaction in blended e-learning environment	21	212 students from different universities and colleges in Taiwan offering blended learning	NR	7-point Likert Scale (ordinal) 1= Strongly disagree 7= Strongly agree	NR	English	English
Akkoyunlu and Yilmaz-Soylu (2008)[34] Learner's View on Blended Learning and Its Implementation Instrument	Learner's view on blended learning and its implementation process	**2 Subscales:** 1: View of Blended Learning Implementation: Ease of use for the web environment (7) Online environment (6) Content module (11) Face-to-face sessions (7) Assessment (4) 2: View of blended learning in GENERAL (15)	Evaluate the view on blended learning and its implementation process	50	Students from the different universities in Turkey (level not specified)	NR	10-point Likert Scale: (ordinal) 1 = not at all 10 = totally true	NR	Turkish	English

CFA – Confirmatory Factor Analysis, NR – not reported, SRQ – self-report questionnaire

et al. targets medical students using 19 items on a 7-point scale.[35] Lazar et al. analyzes digital tool effects on BTL with 36 items.[37] Karadenyz evaluates BTL satisfaction through 18 items on a 5-point scale.[36] Wu et al. measures e-learning satisfaction with 21 items on a 7-point scale.[38] Responses are uniformly quantified using 5 to 10-point Likert scales. Most are in English[35,37,38], with two studies in Turkish[34,36] showcasing diverse BTL nuances. Despite the shared focus on various blended learning aspects, these SRQs differ in item number, response options, target populations, and administration languages, illustrating different facets of BTL such as student satisfaction, self-regulated learning, digital tool usage, and perception on the BTL environment. However, the recall periods and administration times were not specified for these instruments.

Table 2 details study characteristics from five studies. Studies in Turkey[34,36], Australia,[35] Romania,[37] and Taiwan[38] targeted university students,[34–38] including those in health sciences[35,37]. The questionnaires were predominantly in English.[34,35,37,38] Varied sample sizes underline scale development. All but Ballouk et al.[35] had adequate Confirmatory Factor Analysis (CFA) samples[34,36–38]. Age representation differs, with some studies omitting age details.[35,36] The sex distribution also varied, Wu et al. displaying a balanced distribution,[38] whereas Lazar et al. exhibited fluctuating male representation across different stages,[37] suggesting diverse sex dynamics. Response rates varied, with some studies not reporting them.[35–37]

Table 3 evaluates SRQ design quality across studies. Ballouk et al., Lazar et al. , and Wu et al. achieved 'very good' ratings in several criteria,[35,37,38] while Akkoyunlu and Yilmaz-Soylu and Karadenyz were rated 'doubtful' in some areas.[34,36] Four studies were marked 'doubtful' in overall SRQ design,[34,35,37,38] with Karadenyz rated "inappropriate" for lacking a pilot test.[36] Due to the 'inadequate' rating, the questionnaire by Karadenyz was not further evaluated.[36]

Table 4 examines content validity in four studies, focusing on relevance, comprehensibility, and comprehensiveness

Table 2. Characteristics of the Included Study Populations

Authors (Year)	Population			Self-report questionnaire administration			Response rate (%)
	N	Age Mean (SD, range) yr*	Gender (% male)*	Setting	Country	Language	
Ballouk et al. (2022)[35]	FGD: 15 medical students Expert validation (face, content and context): 14 experts Pilot testing: 120 Medicine students	NR	NR	Medical students at the university	Australia	English	NR
Lazar et al. (2020)[37]	Expert Content Validation: 12 experts Stage 1 Pre-testing: 250 of 412 Stage 2 Dimensionality: 206 of 352 Stage 3 CFA: 262* of 414 Stage 4 CFA: 310* of 511	Stage 1: Mean age = 24.03 years old Stage 2: Mean age = 25.04 years old Stage 3: Mean age = 22.95 years old Stage 4: Mean age = 30.53 years old	Stage 1: 60% males Stage 2: 1.5% males Stage 3: 10.1% males Stage 4: 17.5% males*	Stage 1: Sciences, Vasile Alecsandri University of Bacau Stage 2: Educational Sciences, Bucharest University Stage 3: Psychology, Bucharest University Stage 4: Educational Sciences and Environmental Science, Vasile Alecsandri University of Bacau	Romania	English	NR
Karadenyz (2012)[36]	FGD: 6 students Expert validation (content validity): 3 experts CFA: 760 of 1058 students	NR	NR	28 departments at Bahçeşehir University	Turkey	Turkish	NR
Wu et al. (2010)[38]	**Pre-testing**: Interview to four instructors and five students Final administration of cross-sectional survey: 212 university students*	Number of respondents/age group 18-30: 101 31-40: 82 41-50: 23 51-60: 6 Over 60: 1	Male: 106 Female: 106	Universities	Taiwan	English	56
Akkoyunlu and Yilmaz-Soylu (2008)[34]	Pilot testing: 82 university students Final tool administration: 463 of 488 university students*	Range = 19-21*	48*	Different universities in Turkey	Turkey	English	95

CFA – Confirmatory Factor Analysis, FGD – focus group discussion, NR – not reported, SD – standard deviation

**obtained result from final tool administration*

for university students.[34,35,37,38] Ballouk et al. and Lazar et al. showed 'adequate' content validity.[35,37] All had commendable structural validity.[34,35,37,38] Three studies displayed 'very good' internal consistency,[35,37,38] but Akkoyunlu and Yilmaz-Soylu fell short.[34] Cross-cultural validity and reliability were inadequate in the latter study.[34] Only Lazar et al. showed 'very good' known group validity.[37] All studies lacked responsiveness evaluation.[34,35,37,38]

Results on Interpretability and Feasibility of the Self-report Questionnaire

Akkoyunlu and Yilmaz-Soylu, Ballouk et al., and Wu et al. did not report the distribution of scores in the study population, percentage of missing items and total scores, floor and ceiling effects, scores and change scores for relevant subgroups, minimal important change or minimal important difference, and information on response shift.[34,35,38] Remarkably, the study by Lazar et al. distinguished itself by providing detailed information on variability in response scores across the seven domains of the Blended Learning Scale.[37] These encompassed familiarity with high-tech (modern) and classical digital tools (traditional), with mean scores ranging from 2.86 to 3.8 and 3.47 to 3.86, respectively. Perceived barriers and computer anxiety domains demonstrated mean values between 2.93-3.21 and 2.06-3.78, respectively. Meanwhile, perceived usefulness and ease of use showed a relatively high score range of 3.63-3.78 and 3.42-3.79, respectively. Lastly, the intention to use domain reflected a strong behavioral inclination with mean values from 3.47 to 3.85.[37]

All studies did not report the participants' completion time.[34,35,37,38] Only Lazar et al. and Wu et al. provided information on the mental and physical ability requirements for participants.[37,38] In the former, a basic level of familiarity with digital tools and technology was necessary,[37] while the latter required that students be capable of reading, understanding the questions, and responding accordingly.[38] All SRQs demonstrated high ease of standardization and easy score calculation.[34,35,37,38] Information about copyrights and costs remained unreported in all studies.[34,35,37,38] The required equipment was uniformly a questionnaire in all studies.[34,35,37,38] The students' comprehensibility of the questionnaires was

Table 3. Quality of the Self-report Questionnaire Development

Authors (Year)	SRQ design							Cognitive interview study				Total SRQ Development
	General design requirements					Concept elicitation[1]	Total SRQ design	General design requirements	Comprehensibility	Comprehensiveness	Total CI/pilot study	
	Clear construct	Clear origin of construct	Clear target population for which SRQ was developed	Clear context of use	SRQ developed in a sample representing target population			CI/pilot study performed on a sample representing target population				
Ballouk et al. (2022)[35]	V	V	V	V	V	V	D	V	D	D	D	D
Lazar et al. (2020)[37]	V	V	V	V	V	V	D	V	D	D	D	D
Karadenyz (2012)[36]	V	D	V	V	V	D	D				I	I
Wu et al. (2010)[38]	V	V	V	V	V	D	D	V	D	D	D	D
Akkoyunlu and Yilmaz-Soylu (2008)[34]	V	D	V	D	V	D	D	V	D	D	D	D

Ratings: V – very good, A – adequate, D – doubtful, I – inadequate, N – not applicable

CI – cognitive interview, SRQ – self-report questionnaire

[1] When the SRQ was not developed in a sample representing the target population, the concept elicitation was not further rated.

Table 4. Methodological Quality of Studies on Measurement Properties

Authors (Year)	Content validity					Structural validity	Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
	Asking university health science students			Asking experts								Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison between subgroups	Comparison before and after intervention
	Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness												
Akkoyunlu and Yilmaz-Soylu (2008)[34]	D	D	D	D	D	A	I	I	I	0	0	0	0	0	0	0	0
Ballouk et al. (2022)[35]	D	D	D	A	A	A	V	0	0	0	0	V	0	0	0	0	0
Lazar et al. (2020)[37]	D	D	D	A	A	V	V	0	0	0	0	V	V	0	0	0	0
Wu et al. (2010)[38]	D	D	D	D	D	V	V	0	0	0	0	V	0	0	0	0	0

V – very good, A – adequate, D – doubtful, I – inadequate, N – not applicable, 0 – not evaluated

Empty cells indicate that a CI study (or part of it) was not performed.

classified as doubtful in all studies.[33,34,36,37] Although most studies were available in English, Akkoyunlu and Yilmaz-Soylu was also available in Turkish.[34]

Measurement properties

Table 5 presents the psychometric properties of SRQs used by university students to assess aspects of BTL. Most evidence for these properties was rated as 'moderate.' However, Wu et al. was the outlier with a 'very low' quality rating.[38] This

Table 5. Results of Measurement Properties of Self-report Questionnaires Used by Health Sciences University Students in Evaluating Blended Teaching and Learning in a University Setting

Psychometric Properties / Authors (Year)	Results	Overall Rating	Quality of Evidence
Structural Validity			
Akkoyunlu and Yilmaz-Soylu (2008)[34]	Two components extracted: Component 1 (35 items) and Component 2 (15 items) Component 1 divided into 5 sub-components The rotated component matrix for component 1 reports eigenvalues over 1.5.	?	Moderate
Ballouk et al. (2022)[35]	Kaiser Meyer Olkin = 0.655 (p <0.001) Barlett's test of sphericity = (x2(546), p <0.001) Eigenvalues: F1-F4 (3.83, 2.343, 2.086, 1.569)	?	Moderate
Lazar et al. (2020)[37]	Undergraduates enrolled in the psychology study program (N = 262): χ2/Df = 3.157 TLI = 0.948 CFI = 0.978 RMSEA = 0.091 SRMR = 0.048 Graduates (master's students) enrolled in Education Science and Environmental Science study programs (N = 310): χ2/Df = 1.678 TLI = 0.984 CFI = 0.993 RMSEA = 0.047 SRMR = 0.047	+	Moderate
Wu et al. (2010)[38]	AVE Score/Category Computer self-efficacy: 0.605 System functionality: 0.761 Content feature: 0.802 Interaction: 0.782 Performance expectations: 0.838 Learning climate: 0.807 Learning satisfaction: 0.849 The model explained 67.8% of the variance in learning satisfaction.	?	Very low
Internal Consistency			
Akkoyunlu and Yilmaz-Soylu (2008)[34]	Cronbach's alpha = 0.72	+	Moderate
Ballouk et al. (2022)[35]	Cronbach's alpha = 0.764 (0.764 to 0.770)	+	Moderate
Lazar et al. (2020)[37]	Cronbach's alpha = 0.901 (range= 0.736 to 0.917)	+	Moderate
Wu et al. (2010)[38]	Composite reliability = 0.821 to 0.957	+	Very low
Cross-cultural Validity/ Measurement Invariance			
Akkoyunlu and Yilmaz-Soylu (2008)[34]	Cronbach's alpha (Turkish) = 0.86; Cronbach's alpha (English) = 0.88	?	Moderate
Lazar et al. (2020)[37]	χ2/df = 2.418, CFI = 0.986, TLI = 0.967, RMSEA = 0.050, SRMR = 0.048	+	Moderate
Reliability			
Akkoyunlu and Yilmaz-Soylu (2008)[34]	alpha reliability coefficient = 0.74	?	Moderate
Construct Validity			
Ballouk et al. (2022)[35]	BLQ-16: Spearman's rho correlation coefficient = 0.314 to 0.530; BLQ 7-12: Spearman's rho correlation coefficient = 0.300 to 0.519; BLQ 13-19: Spearman's rho correlation coefficient = 0.308 to 0.544	-	Moderate

AVE – Average Variance Extracted, CFI – Comparative Fit Index, RMSEA – Root Mean Square Error of Approximation, RMSEA – Root Mean Square Error of Approximation, SRMR – Standardized Root Mean Square Residual, TLI – Tucker Lewis Index

+ – sufficient, - – insufficient, ? – indeterminate

variation underscores differing robustness in evidence across the studies, highlighting concerns with Wu et al.[38]

Structural validity

Studies by Akkoyunlu and Yilmaz-Soylu, Ballouk et al., and Wu et al. yielded a '?' rating, indicating a degree of ambiguity due to their partial reporting.[34,35,38] These studies employed methods using eigenvalues,[34,35] Kaiser Meyer Olkin,[35] Bartlett's sphericity test,[35] or reporting of average variance extracted scores[38]. However, they did not provide complete confirmatory measures, such as the Comparative Fit Index (CFI), Tucker Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), or Standardized Root Mean Square Residual (SRMR).[34,35,38]

In contrast, Lazar et al. achieved a '+' rating, reporting a higher degree of structural validity through their detailed reporting, including key confirmatory measures. Lazar et al. has a bifurcation in results based on the participant group. For the undergraduate group, despite their TLI of 0.948 and RMSEA of 0.091 falling short of the COSMIN standards, their CFI of 0.978 and SRMR of 0.048 adhered to the cut-off point. In contrast, the graduate group exhibited strong structural validity, with their $\chi 2/Df$ of 1.678, TLI of 0.984, CFI of 0.993, RMSEA of 0.047, and SRMR of 0.047 all exceeding the COSMIN cut-off point.[37]

Internal consistency

The studies by Akkoyunlu and Yilmaz-Soylu, Ballouk et al., Lazar et al., and Wu et al. have all demonstrated sufficient reliability, as evidenced by their respective Cronbach alpha or composite reliability scores, all of which meet or surpass the COSMIN cut-off of 0.70 for each unidimensional scale or subscale.[34,35,37,38] Akkoyunlu and Yilmaz-Soylu reported a Cronbach alpha of 0.72,[34] while Ballouk et al. had a slightly higher alpha of 0.76[35]. Instead, the study by Wu et al. reported composite reliability of 0.82 to 0.96, with both scoring above the COSMIN threshold, indicating acceptable internal consistency.[38] Lastly, Lazar et al. reported a Cronbach's alpha of 0.90, again indicating acceptable internal consistency.[37]

Cross-cultural validity/ Measurement invariance

Lazar et al. received a '+' rating, indicative of the robust cross-cultural validity of their study. Their results, including $\chi 2/df = 2.418$, CFI = 0.986, TLI = 0.967, RMSEA = 0.050, SRMR = 0.048, suggest that no important differences were found between group factors in multiple-group factor analysis, meeting the COSMIN criteria for a '+' rating.[37] Akkoyunlu and Yilmaz-Soylu's study in 2008 received a '?' rating, indicating that some aspects of their research were unclear.[34] Despite reporting Cronbach's alpha for both Turkish (0.86) and English (0.88) versions of the scale - a statistic indicating moderate reliability - the study failed to conduct a multiple-group factor analysis. This type of analysis is crucial for determining measurement invariance, or the degree to which the same underlying constructs are being measured across different groups.[34]

Reliability

Akkoyunlu and Yilmaz-Soylu (2008) reported an alpha reliability coefficient of 0.74, indicating internal consistency within the measurement tool.[34] However, the COSMIN rating was '?', due to the lack of reported ICC or weighted Kappa, which are crucial indicators for the reliability assessment per COSMIN guidelines.[34]

Construct validity

The study by Ballouk et al. reported Spearman's rho correlation coefficients that ranged between 0.30 and 0.54 across various versions of their instrument.[35] However, these correlation coefficients are lower than the benchmark set by the COSMIN guidelines, which require a correlation of ≥0.70. Therefore, the study by Ballouk et al. was given a '-' rating.[35]

Recommendation

In line with the COSMIN guideline, SRQs in Category A were sourced from research conducted by Akkoyunlu and Yilmaz-Soylu, Ballouk et al., and Lazar et al.[34,35,37] A solitary study by Wu et al. provided the questionnaire for Category B.[38]

DISCUSSION

This systematic review determined the quality of SRQs' psychometric properties developed for BTL. Only SRQ development and evaluation studies were obtained for BTL administered to university students, including health science students. Using COSMIN guidelines, the SRQs obtained from this systematic review classified Learner's View on Blended Learning and Its Implementation Instrument by Akkoyunlu and Yilmaz-Soylu, Blended Learning Questionnaire (BLQ) by Ballouk et al., and Blended Learning Scale (BLS) by Lazar et al. as Category A.[34,35,37] The only SRQ for Category B was the Blended e-Learning Scale by Wu et al..[38]

Among the SRQs evaluated, the BLS by Lazar et al. takes precedence in terms of its methodological and psychometric rigor.[37] While exhibiting commendable structural validity and internal consistency, this scale underscores its superiority by meeting sample size requirements for CFA, a critical analytical tool often bypassed, as seen with studies by Akkoyunlu and Yilmaz-Soylu and Ballouk et al..[34,35] A lack of CFA integration, as noted in scale development processes, is a recurrent psychometric shortcoming. Moreover, the BLS by Lazar et al.[37] demonstrated the best methodological quality compared to the other three SRQs[34,35,38]. It obtained an adequate content validation process that involved both target participants and experts in assessing the relevance and comprehensiveness of the items.

A meticulous analysis of the data posits the BLQ by Ballouk et al. in a favorable light.[35] Its design is rated as 'very

good', attributed to its comprehensive concept elicitation, which distinguishes it from other studies. The inductive methodology, incorporating focus group discussions with experts and potential respondents, is pivotal for item generation and dimension determination in SRQs. Such an approach garners invaluable pragmatic insights that refine the construct definition and augment content validity. The BLQ by Ballouk et al. holds a 'moderate' quality rating.[35] Its merits include a 'very good' SRQ design, 'adequate' content validity, and 'very good' internal consistency. Using the Motivated Strategies for Learning Questionnaire enhances its construct validity by measuring students' self-regulated learning in a blended environment. As blended learning evolves, it influences student learning. Ballouk et al. offers a tool capturing students' proactive learning, factoring in environmental and motivational elements, underlining the influence of blended learning on student behaviors.[35]

Ballouk et al. achieved a sufficient level of content validity in terms of methodological quality for the BLQ. However, they did not perform a CFA. The CFA is an integral part of the scale development process; it tests the data's fit with the presumed factor structure.[39] Conducting a CFA can enhance the validity of a self-report questionnaire, a step that was not taken by either Akkoyunlu and Yilmaz-Soylu or Ballouk et al..[34,35] The absence of a CFA is a common psychometric limitation in the scale development process.[40]

The study acknowledges several limitations that may affect the validity and generalizability of its findings. First, heterogeneity in study populations, settings, and questionnaire versions could introduce inconsistencies that challenge the review's generalizability. Second, the subjective nature of comparing SRQ content may impact the recommendations provided. Lastly, the review emphasizes the evolving nature of SRQ for blended learning and the need for continuous assessment, suggesting that future evidence could alter the current conclusions. These limitations highlight areas for future research to address and improve the evidence base regarding the psychometric properties of SRQs for evaluating blended learning in health science programs.

This review has several strengths that enhance its credibility and utility. Utilizing an extensive search strategy across multiple databases, the study ensures a comprehensive inclusion of pertinent literature. By adhering to PRISMA and COSMIN guidelines, the review upholds high methodological quality. It offers a thorough examination of various measurement properties, such as reliability, validity, and responsiveness, contributing to a detailed analysis. The process of achieving consensus through discussion at each review stage promotes collaborative and transparent decision-making. The initial pilot search demonstrated the feasibility of the systematic review, evidenced by several potentially relevant articles. Additionally, the review not only identifies gaps in the current application of SRQs to health science programs but also recommends ongoing assessment and improvement, advocating for a rigorous scale development process. These strengths collectively provide valuable insights into the psychometric properties of SRQs, informing future research and practice in evaluating blended learning within health science education.

This systematic review highlights critical implications for HEIs, focusing on the optimization of assessment practices in BTL. It is imperative for HEIs to meticulously evaluate the psychometric properties of SRQs, selecting SRQs that have proven reliability, validity, and responsiveness for educational evaluations. Continuous evaluation of SRQs is recommended to ensure their alignment with the dynamic needs of both students and teachers. Customizing SRQs to reflect the unique characteristics and objectives of the HEI context can significantly improve their utility and relevance. Furthermore, professional development in assessment methodologies is crucial for teachers and administrators, enhancing the rigor of evaluation practices. Implementing a systematic quality assurance process for evaluating BTL, including regular reviews of SRQs' psychometric properties, is essential for maintaining educational quality. HEIs are also encouraged to engage in an iterative process of SRQ refinement, incorporating feedback from all stakeholders to improve the assessment process. Given the evolving nature of digital technologies in clinical education, continuous update and validation of these SRQs within BTL contexts are necessary. Despite moderate evidence supporting the effectiveness of certain SRQs for BTL evaluation, the review underscores the necessity for ongoing validation efforts across diverse populations to ensure their reliability and applicability. Embracing these practice implications can significantly enhance assessment quality and effectiveness in BTL, contributing to better educational outcomes in health science programs.

We advocate for the initiation of longitudinal studies to evaluate the temporal consistency and sensitivity of SRQs, aiming to determine the evolution of students' perceptions and experiences with BTL. We call for cross-cultural validation studies to ensure these tools are culturally adaptable and universally applicable, considering the diverse cultural tapestry of health science students. The exploration of new constructs is also deemed essential to encompass emerging trends in BTL, thereby enabling the development of more comprehensive assessment tools. Moreover, the review encourages the adoption of innovative research methodologies, including mixed methods approaches and advanced statistical analyses, to deepen the understanding of student experiences in BTL. There is the need for detailed reporting on aspects like recall period, completion time, and overall usability. By adhering to and promoting COSMIN guidelines, this review not only elevates the standards for future research but also underscores the importance of context-specific teaching evaluations. Through addressing these research implications, researchers are poised to significantly contribute to the enhancement of BTL evaluations and the optimization of educational practices within health science education.

CONCLUSION

This review classified SRQs on BTL within health science education into categories based on their validation status and psychometric properties. Category A includes instruments with proven content validity and reliability, notably the Learner's View on Blended Learning and Its Implementation Instrument by Akkoyunlu and Yilmaz-Soylu, the BLQ by Ballouk et al., and the BLS by Lazar et al. These are validated for health science education use due to their high-quality measurement properties. Conversely, the Blended e-Learning Scale by Wu et al. is designated as Category B, indicating it requires further research for validation and improvement. This distinction underscores the necessity of employing psychometrically validated SRQs for effective assessment in blended learning contexts, aiming to advance educational evaluations and enhance student learning outcomes.

Statement of Authorship

All authors certified fulfillment of ICMJE authorship criteria.

Author Disclosure

All authors declared no conflicts of interest.

Funding Source

This review was funded by the National Research Council of the Philippines, Department of Science and Technology (Grant No. NRCP GIA FY 2022, Project No. A-109). The authors declared that the funding source had no role in the study's design, conduct, analysis, interpretation, or reporting. No additional funding or financial benefits were received from any parties with interests in the research outcomes. The authors affirmed that this review was conducted with an objective and unbiased approach.

REFERENCES

1. Graham CR. Blended learning systems: Definition, current trends, and future directions. In: Bonk CJ, Graham CR, eds. Handbook of blended learning: Global perspectives, local designs. San Francisco, CA: Pfeiffer; 2006. pp. 3–21.
2. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res. 2016 Jan 4;18(1):e4807. doi: 10.2196/jmir.4807. PMID: 26729058; PMCID: 4717286.
3. McCutcheon K, Lohan M, Traynor M. A systematic review protocol on the use of online learning versus blended learning for teaching clinical skills to undergraduate health professional students. High Educ Pedagog. 2016 Jan 1;1(1):82–8. doi.org/10.1080/23752696.2015.1134204.
4. Commission on Higher Education. CHED guidelines for the prevention, control, and mitigation of the spread of the 2019-NOVEL CORONA VIRUS ACUTE RESPIRATORY DISEASE (2019-nCOV ARD) in higher education institutions (HEIs) | Kalinga State University [Internet]. 2020 [cited 2022 Oct 29]. Available from: https://ched.gov.ph/wp-content/uploads/CMO-No.-4-s.-2020-Guidelines-on-the-Implementation-of-Flexible-Learning.pdf
5. Alsoufi A, Alsuyihili A, Msherghi A, Elhadi A, Atiyah H, Ashini A, et al. Impact of the COVID-19 pandemic on medical education: Medical students' knowledge, attitudes, and practices regarding electronic learning. doi: 10.1371/journal.pone.0242905. PMID:33237962; PMCID: PMC7688124.
6. Camargo CP, Tempski PZ, Busnardo FF, de Arruda Martins M, Gemperli R. Online learning and COVID-19: a meta-synthesis analysis. doi: 10.6061/clinics/2020/e2286. PMID: 33174948; PMCID: PMC7605278.
7. Al Zahrani EM, Al Naam YA, AlRabeeah SM, Aldossary DN, Al-Jamea LH, Woodman A, et al. E- Learning experience of the medical profession's college students during COVID-19 pandemic in Saudi Arabia. doi: 10.1186/s12909-021-02860-z. PMID: 34416889; PMCID: PMC8378523.
8. Butz N. Motivation in Synchronous Hybrid Graduate Business Programs: A Self-Determination Approach to Contrasting Online and On-Campus Students. J Online Learn Teach [Internet]. 2014 Jan 1 [cited 2022 Nov 4]. Available from: https://www.academia.edu/77969044/Motivation_in_Synchronous_Hybrid_Graduate_Business_Programs_A_Self_Determination_Approach_to_Contrasting_Online_and_On_Campus_Students
9. Gray K, Tobin J. Introducing an online community into a clinical education setting: a pilot study of student and staff engagement and outcomes using blended learning. BMC Med Educ. 2010 Jan 26;10(1):6. doi: 10.1186/1472-6920-10-6. PMID: 20100354; PMCID: PMC2828452.
10. García Rodicio H, Sánchez E, Acuña SR. Support for self-regulation in learning complex topics from multimedia explanations: do learners need extensive or minimal support? Instr Sci. 2013 May;41(3):539–53. https://doi.org/10.1007/s11251-012-9243-4
11. Shim TE, Lee SY. College students' experience of emergency remote teaching due to COVID-19. Child Youth Serv Rev. 2020 Dec 1;119:105578. doi: 10.1016/j.childyouth.2020.105578. PMID: 33071405; PMCID: PMC7550861.
12. Chesterton P, Richardson M, Tears C. Student physiotherapists perceptions of online curriculum delivery during the COVID-19 pandemic. BMC Med Educ. 2022 Jun 7;22(1):440. doi: 10.1186/s12909-022-03486-5. PMID: 35672778; PMCID: PMC9172598.
13. Coe R, Aloisi C, Higgins S, Major LE. What makes great teaching? Review of the underpinning research. Sutton Trust. 2014 Oct.
14. Lin WS, Wang CH. Antecedences to continued intentions of adopting e-learning system in blended learning instruction: A contingency framework based on models of information system success and task-technology fit. Comput Educ. 2012 Jan 1;58(1):88–99. doi: 10.1016/j.compedu.2011.07.008.
15. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. JBJS. 2009 May 1;91(Supplement_3):99–106. doi: 10.2106/JBJS.H.01624. PMID: 19411507.
16. Lyons C, Brown T, Bourke-Taylor H. The Classroom Environment Questionnaire (CEQ): Development and preliminary structural validity. Aust Occup Ther J. 2018 Oct;65(5):363–75. doi: 10.1111/1440-1630.12474. PMID: 29659028.
17. Ginns P, Barrie S. Developing and testing a student-focussed teaching evaluation survey for university instructors. Psychol Rep. 2009 Jun 1;104(3):1019–32. doi: 10.2466/PR0.104.3.1019-1032. PMID: 19708424.
18. Kossioni AE, Lyrakos G, Ntinalexi I, Varela R, Economu I. The development and validation of a questionnaire to measure the clinical learning environment for undergraduate dental students (DECLEI). doi: 10.1111/eje.12051. PMID: 24521086.
19. Krois NR, Kossioni AE, Barlow PB, Tabrizi M, Marchini L. Steps towards validation of the Dental Education Clinical Learning Instrument (DECLEI) in American dental schools (DECLEI-USA). J Dent Educ. 2020 Aug;84(8):895–901. doi: 10.1002/jdd.12145. PMID: 32166746.

20. Bangert AW. Building a validity argument for the community of inquiry survey instrument. Internet High Educ. 2009 Jun 1;12(2):104–11. doi: 10.1016/j.iheduc.2009.06.001.
21. Matosas-López L, Aguado-Franco JC, Gómez-Galán J. Constructing an instrument with behavioral scales to assess teaching quality in blended learning modalities. J New Approaches Educ Res. 2019 Jul 15;8(2):142. doi: 10.7821/naer.2019.7.410.
22. Berk RA. Face-to-face versus online course evaluations: A "Consumer's Guide" to seven strategies. 2013;9(1).
23. Dziuban C, Graham CR, Moskal PD, Norberg A, Sicilia N. Blended learning: the new normal and emerging technologies. Int J Educ Technol High Educ. 2018 Feb 15;15(1):3. doi: 10.1186/s41239-017-0087-5.
24. Grammatikopoulos V, Linardakis M, Gregoriadis A, Oikonomidis V. Assessing the Students' Evaluations of Educational Quality (SEEQ) questionnaire in Greek higher education. High Educ. 2015;70(3): 395–408. doi: 10.1007/s10734-014-9837-7.
25. Rothman T, Romeo L, Brennan M, Mitchell D. Criteria for assessing student satisfaction with online courses. Int J E-Learn Secur. 2011 Mar 1;1(1):27–32. doi: 10.20533/IJELS.2046.4568.2011.0004.
26. Thomas JE, Graham CR, Piña AA. Current practices of online instructor evaluation in higher education.
27. Cobo-Rendón R, Bruna Jofre C, Lobos K, Cisternas San Martin N, Guzman E. Return to University Classrooms With Blended Learning: A Possible Post-pandemic COVID-19 Scenario. Front Educ [Internet]. 2022 [cited 2023 Aug 10];7. Available from: https://www.frontiersin.org/articles/10.3389/feduc.2022.957175. doi: 10.3389/feduc.2022.957175
28. Dayagbil FT, Palompon DR, Garcia LL, Olvido MMJ. Teaching and learning continuity amid and beyond the pandemic. Front Educ 2021;6. doi: 10.3389/feduc.2021.678692.
29. Joaquin JJB, Biana HT, Dacela MA. The Philippine higher education sector in the time of COVID-19. Front Educ. 2020;5. doi: 10.3389/feduc.2020.576371.
30. Dones VC, Dalusong MTB, Manlapaz D, Rojas JA, Ho KC, Reyes JJ, et al. Evidence on psychometric properties of self-report questionnaires in evaluating blended learning in health sciences university students: research protocol for systematic review and meta-analysis. BMJ Open. 2024 Aug;14(8):e075266. doi: 10.1136/bmjopen-2023-075266. PMID: 39153782; PMCID: PMC11331856.
31. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res Int J Qual Life Asp Treat Care Rehabil. 2018 May;27(5):1147–57. doi: 10.1007/s11136-018-1798-3 PMID: 29435801; PMCID: PMC5891568.
32. Ket H. Search Blocks - BMI Group [Internet]. Creative Commons Naamsvermelding-NietCommercieel-GelijkDelen 4.0 Internationaal-licentie. [cited 2023 Jan 3]. Available from: https://www.bmi-online.nl/searchblocks/search-blocks-bmi/
33. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018 May;27(5):1159–70. doi: 10.1007/s11136-018-1829-0. PMID: 29550964; PMCID: PMC5891557.
34. Akkoyunlu B, Soylu MY. A study of student's perceptions in a blended learning environment based on different learning styles. J Educ Technol Soc. 2008;11(1):183–93.
35. Ballouk R, Mansour V, Dalziel B, Hegazi I. The development and validation of a questionnaire to explore medical students' learning in a blended learning environment. BMC Med Educ. 2022 Jan 3;22(1):4. doi: 10.1186/s12909-021-03045-4. PMID: 34980084; PMCID: PMC8722320.
36. Karadenyz. Development of student satisfaction with collected learning. 2012;11(40):161–72.
37. Lazar IM, Panisoara G, Panisoara IO. Digital technology adoption scale in the blended learning context in higher education: Development, validation and testing of a specific tool. PLoS One. 2020 Jul 10;15(7):e0235957. doi: 10.1371/journal.pone.0235957. PMID: 32649691; PMCID: PMC7351189.
38. Wu JH, Tennyson RD, Hsia TL. A study of student satisfaction in a blended e-learning system environment. Comput Educ. 2010 Aug 1;55(1):155–64. doi: 10.1016/j.compedu.2009.12.012.
39. Prudon P. Confirmatory factor analysis as a tool in research using questionnaires: a critique. Compr Psychol. 2015 Jan 1;4:03.CP.4.10. doi.org/10.2466/03.CP.4.10.
40. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. Psicol Reflexao E Crit Rev Semest Dep Psicol UFRGS. 2017 Jan 25;30(1):3. doi: 10.1186/s41155-016-0057-1. PMID: 32025957; PMCID: PMC6966966..