

Review began 09/06/2026

Review ended 10/12/2026

Published 10/18/2026

© Copyright 2026

Shevalkar et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118140

# The Development and Content Validation of a Student Satisfaction Questionnaire for Physiotherapy Teaching-Learning Methods: An Expert Panel Study

Rupali Shevalkar<sup>1</sup>, Bharati Asgaonkar<sup>2</sup>, Manali N. Yadav<sup>1</sup>

1. Musculoskeletal Physiotherapy, K J Somaiya College of Physiotherapy, Mumbai, IND 2. Physical Therapy, Topiwala National Medical College and BYL Nair Charitable Hospital, Mumbai, IND

**Corresponding author:** Rupali Shevalkar, rupali.salvi@somaiya.edu

## Abstract

### Background and objective

Problem-based learning (PBL) is increasingly used alongside traditional didactic teaching (TDT) in physiotherapy education, and student satisfaction is a key outcome for evaluating the acceptability of different teaching-learning methods. However, no content-validated, physiotherapy-specific instrument is available to measure student satisfaction across different teaching-learning methods. Hence, we aimed to develop a student satisfaction questionnaire to evaluate physiotherapy teaching-learning methods, establish its content validity through a structured expert panel review, and assess its preliminary internal consistency in a pilot sample.

### Methods

A 12-item questionnaire was developed to evaluate aspects such as session organisation, content relevance, clarity of facilitation, achievement of learning objectives, comprehension, confidence in application, perceived long-term retention, and overall participant satisfaction. The questionnaire was developed using a structured item-generation process. A purposively selected panel of six subject-matter experts assessed the draft questionnaire for relevance, clarity, and simplicity using a 4-point scale. For each dimension, the item-level content validity index (I-CVI) and scale-level content validity index (S-CVI) were computed. Items with an I-CVI below 0.78 were flagged for revision and re-evaluation. The modified kappa statistic ( $\kappa^*$ ) was used to adjust the I-CVI for potential chance agreement. Following revisions, the finalised questionnaire was pilot-tested in two separate groups of 30 final-year physiotherapy students. One group was evaluated following a PBL session, while the other was evaluated following a TDT session on the same topic. Internal consistency under each condition was assessed using Cronbach's alpha, corrected item-total correlations, and alpha-if-item-deleted values.

### Results

Ten out of the 12 items achieved an I-CVI of at least 0.78 across all three evaluated dimensions. However, two items did not meet the threshold for clarity and simplicity, with an I-CVI of 0.667, although they were deemed sufficiently relevant, with an I-CVI of 0.833, leading to revisions in their wording. The S-CVI/Ave scores were 0.944 for relevance, 0.889 for clarity, and 0.875 for simplicity, resulting in an overall average S-CVI/Ave of 0.903 across all dimensions. The modified kappa values varied between 0.565, indicating fair agreement, and 1.00, indicating excellent agreement, with the lowest values linked to the two revised items. In the pilot testing, the finalised questionnaire showed strong internal consistency under both conditions, with Cronbach's alpha scores of 0.929 for PBL and 0.972 for TDT. Corrected item-total correlations ranged from 0.559 to 0.800 for PBL and from 0.736 to 0.950 for TDT, and removing any item did not enhance the overall scale alpha in either condition.

### Conclusions

The 12-item questionnaire demonstrated strong content validity, with 10 items rated highly and two revised following expert review, as well as excellent internal consistency in both PBL and TDT pilot groups, with expert agreement exceeding chance levels. These findings provide preliminary support for using the questionnaire to assess student satisfaction with PBL and TDT in physiotherapy education. Its reliability across other teaching-learning methods, such as flipped classrooms and simulation, has not been evaluated and warrants further investigation.

**Categories:** Medical Education

**Keywords:** content validity, content validity index, expert panel, physiotherapy education, problem-based learning, questionnaire development, student satisfaction

#### How to cite this article

Shevalkar R, Asgaonkar B, Yadav M N (October 18, 2026) The Development and Content Validation of a Student Satisfaction Questionnaire for Physiotherapy Teaching-Learning Methods: An Expert Panel Study. Cureus 18(10): e118140. DOI 10.7759/cureus.118140

## Introduction

Problem-based learning (PBL) has been progressively adopted alongside, and sometimes in place of, traditional didactic teaching (TDT) in physiotherapy education. This approach is grounded in constructivist and self-directed learning principles, which position learners as active participants in knowledge construction rather than passive recipients of content [1]. A substantial body of evidence supports the positive effects of PBL on learner satisfaction and engagement across health professions education, with satisfaction consistently reported as one of its more reproducible benefits compared with conventional instruction [2]. Within physiotherapy education, PBL implementation has been associated with high levels of reported student satisfaction and improved self-directed learning behaviour [3]. However, the evidence base in this discipline remains comparatively underdeveloped compared with that in medicine and nursing, with much of the physiotherapy-specific literature relying on qualitative or observational research designs [4].

Numerous studies evaluating and comparing teaching-learning approaches, whether comparing TDT with a single innovative method or multiple innovative methods with one another, often rely on ad hoc, locally developed satisfaction surveys for data collection. These surveys typically lack formal evidence of content validity, raising concerns about whether the instruments adequately represent the intended construct rather than an incomplete or context-specific interpretation of it [5]. At the time this study was conducted, no content-validated, physiotherapy-specific student satisfaction questionnaire was identified in the literature that could be applied across a diverse range of innovative teaching-learning methods. This absence highlights a clear gap in measurement tools for researchers seeking to evaluate curricular innovations in a psychometrically sound and methodologically neutral manner.

Content validity, which refers to the extent to which an instrument's items adequately represent the construct being assessed, is considered a crucial, although not the only, step in developing a new measurement tool [6]. One of the most widely recognised methods for assessing content validity in health-related studies involves a structured review by an expert panel, followed by the computation of the item-level content validity index (I-CVI) and scale-level content validity index (S-CVI) [7]. Since raw I-CVI scores may not distinguish genuine expert consensus from agreement that could occur by chance, particularly with smaller panels, the modified kappa statistic ( $\kappa^*$ ) is recommended as an additional measure to account for chance agreement at the item level [8]. Recent methodological reviews have provided further guidance on appropriate procedures, thresholds, and reporting standards for content validity research [9], which informed the methodology adopted in this study. These recommendations include the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology, which identifies content validity as the initial and most important phase in a comprehensive, staged process for developing patient- and learner-reported outcome measures [10].

To address this measurement gap, we developed a novel 12-item student satisfaction questionnaire designed to be applicable across a wide range of teaching-learning methods in physiotherapy education. A structured expert panel review was conducted to establish the questionnaire's content validity. Subsequently, the final instrument was piloted to assess its internal consistency under two conditions for which comparative data were available: PBL and TDT. This report details the questionnaire development process, findings on item- and scale-level content validity, resulting revisions, and pilot reliability results for both conditions. The aim was to provide an instrument with preliminary psychometric support, alongside a transparent and replicable methodological template for future instrument-validation studies, including the extension of reliability testing to other innovative teaching-learning methods in physiotherapy and allied health education.

## Materials And Methods

### Study design and setting

This methodological instrument-development study was conducted in three sequential stages: (1) questionnaire development, (2) content validation by an expert panel, and (3) pilot reliability testing. The stages followed recommended practice for scale development and content validation [9,11,12]. They were also guided by the COSMIN framework, which treats content validity as the first and most important measurement property, to be established before reliability is examined [10]. Each stage had a distinct objective, sample, and analytic method, summarised in Table 1. The study was conducted in the Department of Physiotherapy, K. J. Somaiya College of Physiotherapy, Mumbai, India, between January 2023 and August 2023.

| Stage                        | Objective                                                                     | Participants                                                                                   | Method and framework                                                                                                                      | Output                                                         |
|------------------------------|-------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|
| 1. Questionnaire development | Generate and consolidate candidate items                                      | Study investigators                                                                            | Literature review and structured item generation; internal consolidation of overlapping items                                             | 12-item draft questionnaire                                    |
| 2. Content validation        | Establish content validity of the draft                                       | Six-member expert panel (purposive)                                                            | Independent ratings of relevance, clarity and simplicity on a 4-point scale; I-CVI, S-CVI/Ave, S-CVI/UA and modified kappa ( $\kappa^*$ ) | Item retention and revision decisions; finalised questionnaire |
| 3. Pilot reliability testing | Preliminary assessment of internal consistency of the finalised questionnaire | Two independent samples of 30 final-year physiotherapy students (PBL condition, TDT condition) | Cronbach's alpha, corrected item-total correlations and alpha-if-item-deleted, calculated separately per condition                        | Preliminary reliability estimates                              |

**TABLE 1: Stages of the instrument-development study**

I-CVI: item-level content validity index; S-CVI: scale-level content validity index; S-CVI/Ave: mean I-CVI across all 12 items; S-CVI/UA: proportion of items with I-CVI = 1.00 (universal agreement); PBL: problem-based learning; TDT: traditional didactic teaching

## Ethical considerations

The Institutional Ethics Committee of Topiwala National Medical College and BYL Nair Charitable Hospital, Mumbai, reviewed and approved the study protocol (approval number: ECARP/2022/74, approved on September 22, 2022). This study adhered to the ethical guidelines of the Declaration of Helsinki. Before participating, all expert and pilot student participants provided written informed consent after being informed of the study's purpose and the voluntary nature of their participation. This instrument validation study did not include any patient data.

## Questionnaire development

Items were generated by the study investigators, guided by published literature on student satisfaction with teaching-learning methods in health professions education [2,13] and by their experience in physiotherapy education. Four domains were specified a priori: content and organisation of the topic, teaching effectiveness, learning outcomes, and overall evaluation. A 12-item draft was developed directly from these domains, and no larger initial item pool was generated or subsequently reduced. Item wording and selection rested on investigator judgement informed by the literature, and no formal conceptual model of satisfaction was applied. The adequacy of content coverage was evaluated independently by the expert panel (Stage 2). Each item was rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree).

## Expert panel

Six experts were purposively selected by the investigators for Stage 2 based on their relevant experience in teaching, curriculum development, and educational research. The panel comprised two physiotherapy subject-matter experts (both MPT-qualified, with 20 and 10 years of experience), two curriculum committee members (both PhD-qualified, with 25 and 19 years of experience), and two health professions education researchers, one with 35 years of experience and the other with 11 years, both of whom were FAIMER fellows. All experts had at least 10 years of experience (median: 19.5 years; Table 2). They were not involved in item generation and rated the items independently. A panel of six experts meets the commonly cited minimum requirement for calculating a stable I-CVI [7,14].

| Expert | Role                                    | Highest qualification/credential | Years of experience |
|--------|-----------------------------------------|----------------------------------|---------------------|
| 1      | Subject-matter expert (physiotherapy)   | MPT                              | 20                  |
| 2      | Subject-matter expert (physiotherapy)   | MPT                              | 10                  |
| 3      | Curriculum committee member             | PhD                              | 25                  |
| 4      | Curriculum committee member             | PhD                              | 19                  |
| 5      | Health professions education researcher | FAIMER fellow                    | 35                  |
| 6      | Health professions education researcher | FAIMER fellow                    | 11                  |

**TABLE 2: Expert panel**

## Content validity assessment procedure

The questionnaire draft was provided to each expert, along with the definition of the study construct, a criteria checklist, and a structured content validity rating form. Each expert independently evaluated every item using three distinct 4-point ordinal scales, which correspond to three specific content-validity dimensions as recommended in the methodological literature [15]: relevance (the degree to which the item is pertinent to the construct of student satisfaction with a teaching-learning method), clarity (the degree to which the item is unambiguously worded), and simplicity (the degree to which the item is easily comprehensible to the target respondent population). Each dimension was rated as follows: 1 = not relevant/clear/simple, 2 = somewhat relevant/clear/simple, 3 = relevant/clear/simple, and 4 = highly relevant/clear/simple. In addition to numerical ratings, experts could record free-text comments on each item.

## Pilot testing and internal consistency

Pilot testing was conducted using the finalised post-revision questionnaire. The finalised 12-item questionnaire was piloted on two independent samples of 30 final-year physiotherapy students each, drawn from the initial phase of the broader research programme in which this instrument was used. One sample completed the final questionnaire immediately after a PBL session, and the other after a TDT session, with both sessions covering the same musculoskeletal topic, allowing internal consistency to be examined under two distinct teaching-learning conditions rather than a single condition. The internal consistency of the 12-item scale was assessed separately for each condition using Cronbach's alpha, together with corrected item-total correlations and alpha-if-item-deleted values for each item [16]. This step addresses reliability, which is a distinct psychometric property from the content validity established in the preceding stage. Because the questionnaire item content was designed to be method-agnostic rather than specific to any one teaching-learning approach, PBL and TDT were selected as the two conditions for which pilot data were available at the time of this study; reliability testing under other innovative teaching-learning methods was outside the scope of the present pilot and is identified as a direction for future research.

## Statistical analysis

I-CVI was determined for each item and dimension by dividing the number of experts rating the item as 3 or 4 by the total number of experts ( $N = 6$ ) [12]. In accordance with established guidelines for panels comprising six or more experts, items were retained if I-CVI was  $\geq 0.78$ ; items below this threshold were earmarked for revision of wording rather than automatic exclusion, provided their relevance to the construct remained acceptable [11]. The scale-level content validity index was calculated in two forms: S-CVI/Ave, which is the mean of the I-CVI values across all 12 items, with a threshold of  $\geq 0.90$  for scale-level acceptability, and S-CVI/UA, representing the proportion of items achieving universal agreement (I-CVI = 1.00) among all six experts [12]. Given that I-CVI does not account for the possibility of chance agreement, the modified kappa statistic ( $\kappa^*$ ) was also computed for each item. Kappa values were interpreted using standard benchmarks for normed assessment instruments: excellent ( $\kappa^* > 0.74$ ), good (0.60-0.74), fair (0.40-0.59), and poor ( $< 0.40$ ) [17]. For the pilot reliability analysis, Cronbach's alpha was calculated. Values were interpreted as excellent ( $\geq 0.90$ ), good (0.80-0.89), acceptable (0.70-0.79), questionable (0.60-0.69), poor (0.50-0.59), or unacceptable ( $< 0.50$ ) [18]. All analyses were conducted using Microsoft Excel 2013 (Microsoft Corporation, Redmond, WA) and SPSS Statistics v.21.0 (IBM Corp., Armonk, NY).

## Results

### Item-level content validity

Among the 12 items in the questionnaire, 10 achieved an I-CVI of at least 0.78 across all three assessed dimensions, as depicted in Table 3. In contrast, two items (item 7, originally stated as 'the session improved

my comprehensive understanding of the topic’ and later modified to ‘the session improved my overall understanding of the topic’, and item 8, initially phrased as ‘the session improved my confidence and ability of application of gained knowledge’ and subsequently revised to ‘the session increased my confidence in applying the knowledge gained’) did not reach the 0.78 threshold for clarity and simplicity, each scoring an I-CVI of 0.667 in these areas. However, both items surpassed the relevance threshold, with an I-CVI of 0.833. The modified kappa values for relevance were 1.00 for the eight items that received unanimous expert agreement (6/6) and 0.816 for the four items with one dissenting expert, indicating excellent agreement. For clarity and simplicity, kappa values were similarly 1.00 or 0.816 for all retained items, but dropped to 0.565 (indicating fair agreement) for items 7 and 8 on both dimensions. This finding independently supports the decision to revise the two items based on the I-CVI results.

| Item | Statement (pre-revision wording)                          | I-CVI (R) | $\kappa^*$ (R) | I-CVI (C) | $\kappa^*$ (C) | I-CVI (S) | $\kappa^*$ (S) | Decision |
|------|-----------------------------------------------------------|-----------|----------------|-----------|----------------|-----------|----------------|----------|
| 1    | Session organization was clear and structured             | 1         | 1              | 0.833     | 0.816          | 0.833     | 0.816          | Retained |
| 2    | Subject content was relevant and up to date               | 1         | 1              | 1         | 1              | 1         | 1              | Retained |
| 3    | Teacher/tutor explained/facilitated concepts clearly      | 1         | 1              | 1         | 1              | 0.833     | 0.816          | Retained |
| 4    | Session stimulated understanding of learning objectives   | 0.833     | 0.816          | 0.833     | 0.816          | 0.833     | 0.816          | Retained |
| 5    | Session improved understanding of theoretical concepts    | 1         | 1              | 1         | 1              | 1         | 1              | Retained |
| 6    | Session improved understanding of practical application   | 1         | 1              | 0.833     | 0.816          | 0.833     | 0.816          | Retained |
| 7    | Session improved comprehensive understanding of the topic | 0.833     | 0.816          | 0.667     | 0.565          | 0.667     | 0.565          | Revised  |
| 8    | Session improved confidence/ability of application        | 0.833     | 0.816          | 0.667     | 0.565          | 0.667     | 0.565          | Revised  |
| 9    | Method likely to retain knowledge long-term               | 0.833     | 0.816          | 0.833     | 0.816          | 0.833     | 0.816          | Retained |
| 10   | Overall quality of session was satisfactory               | 1         | 1              | 1         | 1              | 1         | 1              | Retained |
| 11   | Found the teaching-learning method effective              | 1         | 1              | 1         | 1              | 1         | 1              | Retained |
| 12   | Recommend the method for future topics                    | 1         | 1              | 1         | 1              | 1         | 1              | Retained |

**TABLE 3: I-CVI for relevance, clarity, and simplicity, and modified kappa ( $\kappa^*$ ) for each dimension, by item (N = 6 experts)**

I-CVI  $\geq 0.78$  = acceptable.  $\kappa^*$  was interpreted as excellent ( $> 0.74$ ), good (0.60–0.74), fair (0.40–0.59), and poor ( $< 0.40$ ). Items 7 and 8 indicate the items revised for clarity and simplicity

I-CVI (R/C/S): item-level content validity index for relevance/clarity/simplicity.  $\kappa^*$ : modified kappa

### Scale-level content validity

S-CVI/Ave surpassed the recommended threshold of 0.90 for relevance, achieving a value of 0.944, indicating strong relevance. However, it slightly underperformed in terms of clarity and simplicity, with values of 0.889 and 0.875, respectively (Table 4). This suggests that the two items require revision in these areas. The overall mean S-CVI/Ave across all three dimensions was 0.903. The proportion of items that attained universal expert agreement (6/6) was 0.667 for relevance, 0.500 for clarity, and 0.417 for simplicity. The I-CVI and  $\kappa^*$  values in Table 3 refer to the pre-revision wording. Content validity was established for the 10 items that met the threshold in the first round; the revised wording of items 7 and 8 reflects expert feedback but has not been formally re-rated.

| Dimension  | S-CVI/Ave | S-CVI/UA | Threshold | Verdict          |
|------------|-----------|----------|-----------|------------------|
| Relevance  | 0.944     | 0.667    | ≥ 0.90    | Acceptable       |
| Clarity    | 0.889     | 0.5      | ≥ 0.90    | Marginally below |
| Simplicity | 0.875     | 0.417    | ≥ 0.90    | Marginally below |

**TABLE 4: S-CVI by dimension (N = 6 experts, 12 items)**

Scale-level acceptability threshold: S-CVI/Ave ≥ 0.90

S-CVI: scale-level content validity index; I-CVI: item-level content validity index; S-CVI/Ave: mean I-CVI across all 12 items; S-CVI/UA: proportion of items with I-CVI = 1.00 (universal agreement)

Pilot reliability testing

The finalised questionnaire was piloted with two independent samples, each comprising 30 final-year physiotherapy students. One sample was assessed following a PBL session, and the other was evaluated after a TDT session on the same topic. The 12-item scale exhibited excellent internal consistency in both scenarios. Specifically, Cronbach’s alpha was 0.929 post-PBL (mean total score 48.27, SD 5.91, out of 60) and 0.972 post-TDT (mean total score 46.70, SD 7.68) (Table 5). In the PBL sample, the corrected item-total correlations ranged from 0.559 to 0.800, with alpha-if-item-deleted values spanning 0.919 to 0.929. In the TDT sample, the corrected item-total correlations varied from 0.736 to 0.950, and the alpha-if-item-deleted values ranged from 0.968 to 0.973. The removal of any item did not surpass the overall scale alpha in either condition, thereby supporting the retention of all 12 items under both conditions for which pilot data were available.

| Item | PBL mean (SD) | PBL item-total r | PBL α if deleted | TDT mean (SD) | TDT item-total r | TDT α if deleted |
|------|---------------|------------------|------------------|---------------|------------------|------------------|
| 1    | 3.93 (0.58)   | 0.755            | 0.921            | 3.93 (0.69)   | 0.841            | 0.97             |
| 2    | 4.30 (0.47)   | 0.722            | 0.924            | 3.87 (0.63)   | 0.906            | 0.969            |
| 3    | 3.97 (0.56)   | 0.763            | 0.921            | 3.83 (0.75)   | 0.885            | 0.969            |
| 4    | 4.23 (0.50)   | 0.688            | 0.924            | 4.03 (0.81)   | 0.804            | 0.971            |
| 5    | 3.90 (0.80)   | 0.722            | 0.923            | 4.03 (0.61)   | 0.866            | 0.97             |
| 6    | 3.90 (0.84)   | 0.661            | 0.926            | 3.80 (0.85)   | 0.801            | 0.971            |
| 7    | 4.13 (0.57)   | 0.733            | 0.922            | 3.93 (0.64)   | 0.95             | 0.968            |
| 8    | 4.03 (0.61)   | 0.565            | 0.928            | 3.83 (0.79)   | 0.897            | 0.969            |
| 9    | 4.03 (0.72)   | 0.559            | 0.929            | 3.90 (0.71)   | 0.905            | 0.968            |
| 10   | 3.93 (0.69)   | 0.749            | 0.921            | 3.73 (0.78)   | 0.736            | 0.973            |
| 11   | 3.97 (0.72)   | 0.787            | 0.919            | 3.93 (0.74)   | 0.887            | 0.969            |
| 12   | 3.93 (0.69)   | 0.8              | 0.919            | 3.87 (0.73)   | 0.81             | 0.971            |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>TABLE 5: Pilot reliability statistics by condition: item means, corrected item-total correlations, and alpha-if-item-deleted (N = 30 per condition)</b></p> <p>Overall Cronbach’s alpha: PBL = 0.929, TDT = 0.972 (each N = 30, k = 12 items), both interpreted as excellent per standard benchmarks. No item’s alpha-if-deleted value exceeded the overall scale alpha in either condition. These two conditions represent the only teaching-learning methods for which pilot reliability data are currently available</p> <p>PBL: problem-based learning; SD: standard deviation; TDT: traditional didactic teaching</p> |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Discussion

This study developed a 12-item questionnaire and established its content validity for assessing and

comparing innovative teaching-learning approaches, such as PBL, with traditional didactic methods in physiotherapy education. The questionnaire demonstrated strong overall content validity: 10 of the 12 items were retained unchanged, while the two items requiring revision underwent only minor wording changes to improve clarity and simplicity, as their relevance to the core construct was not in doubt. The overall mean S-CVI/Ave of 0.903 is favourable compared with the content validity indices reported for other newly developed instruments in health professions education, which typically report S-CVI/Ave values ranging from 0.80 to 0.95 after one or more rounds of expert review and revision [6].

A significant methodological strength of this study is the consistency between the revision decisions based on the I-CVI and the independently calculated modified kappa statistic. Items 7 and 8, which were the only items falling below the I-CVI acceptability threshold for clarity and simplicity, also had the lowest kappa values (0.565, indicating fair agreement) across all items and dimensions, whereas all retained items had kappa values of 0.816 or 1.00, indicating excellent agreement. This concordance between the raw agreement-based index and the chance-corrected statistic supports the interpretation that the observed I-CVI pattern reflects genuine variation in expert judgement rather than an artefact of the specific index used. This finding addresses a criticism of content validity research that relies solely on the I-CVI.

In addition to content validity, the pilot assessment of internal consistency provides an important complementary source of psychometric evidence. The final questionnaire demonstrated excellent internal consistency across two independent samples of 30 students each: one assessed following PBL, with a Cronbach's alpha of 0.929, and the other following TDT, with a Cronbach's alpha of 0.972. In both conditions, every item demonstrated a corrected item-total correlation exceeding 0.55, and removing any item did not improve the overall scale's alpha. Content validity and internal consistency are distinct psychometric properties. The former establishes whether expert judges consider the items appropriate representations of the construct, while the latter indicates the extent to which responses to the items are internally consistent. A validation study reporting only content validity, as is common in the literature, leaves the instrument's practical measurement performance unexamined [19]. The favourable findings for both properties in this questionnaire, derived from independent sources of evidence (expert judgement and student responses) and assessed using distinct statistical methods, together with the maintenance of internal consistency across two different teaching-learning conditions, strengthen confidence in the questionnaire's suitability for its intended use in these specific contexts.

The item-level findings are readily interpretable. Items 7 and 8 used compound and relatively abstract phrasing, such as 'comprehensive understanding' and 'confidence and ability of application'. These phrases may have been interpreted differently by different experts, even though the underlying constructs, namely overall understanding and confidence in applying knowledge, were considered relevant to student satisfaction. Based on expert feedback, both items were revised using simpler, more direct wording focused on a single construct. This approach aligns with established best practices for iterative item refinement in questionnaire development, whereby items are revised and re-evaluated rather than discarded when clarity or simplicity, rather than relevance, is the primary concern [20].

The questionnaire developed and evaluated in this study addresses a significant gap in instrumentation in physiotherapy education research, where ad hoc, non-validated satisfaction surveys have been prevalent [21]. Unlike most comparable instrument-development reports in this field, which typically focus solely on content validity, this study also provides preliminary evidence of internal consistency across two distinct teaching-learning conditions. This dual-source evidence, comprising expert judgement and student response data, offers prospective users greater confidence in the instrument before its adoption. Although preliminary, the practical implications of this study are clear: physiotherapy educators and researchers evaluating PBL or TDT can use this questionnaire as a measure with preliminary psychometric support for these specific conditions, reducing the need to develop and validate an ad hoc instrument.

Researchers seeking to evaluate other innovative methods, such as flipped classrooms or simulation-based learning, may adopt the same item content but should consider its reliability in those contexts an untested hypothesis that requires confirmation through a comparable pilot study rather than an established property of the questionnaire. The transparent, formula-based validation procedure reported here, including the less commonly reported chance-corrected kappa statistic alongside the standard item-level and scale-level content validity indices (I-CVI and S-CVI) and the two-condition pilot reliability analysis, may also serve as a practical, replicable template for physiotherapy and allied health educators seeking to develop and validate similar instruments. This approach could be extended to assess reliability across additional teaching-learning methods not examined in the present study.

## Limitations

Item generation relied on a literature review and investigator judgement within prespecified domains, without a formal conceptual framework or student input. Consequently, content coverage was assessed solely through expert panel review. Both the expert panel and the pilot samples were small. The revised versions of items 7 and 8 were not returned to the expert panel; therefore, their content validity following revision has not been formally established, and the reported content validity indices reflect the original wording of these items. A second round of expert review should confirm their content validity before the

instrument is used more widely. The expert panel and item pool were also developed within a single institutional and disciplinary context. Although the items address generic dimensions of teaching-learning satisfaction that may generalise across physiotherapy and allied health settings, validation across multiple institutions would provide stronger evidence of the questionnaire's broader applicability.

## Conclusions

A 12-item student satisfaction questionnaire for evaluating teaching-learning methods in physiotherapy education was developed and assessed for content validity and preliminary internal consistency. Following review by a six-member expert panel, 10 items met the I-CVI thresholds for relevance, clarity, and simplicity, while two items (7 and 8) fell below the thresholds for clarity and simplicity and were reworded based on expert feedback but were not formally re-evaluated. Pilot testing of the revised questionnaire in two samples of 30 students each, following PBL and TDT sessions, demonstrated high preliminary internal consistency, although the very high values may partly reflect item redundancy. These findings provide initial evidence supporting the use of the questionnaire in PBL and TDT settings. Further steps include re-evaluating the revised items, examining the factor structure, assessing test-retest reliability, and conducting external validation in larger, independent, multi-institutional samples.

## Additional Information

### Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

**Concept and design:** Rupali Shevalkar, Bharati Asgaonkar, Manali N. Yadav

**Acquisition, analysis, or interpretation of data:** Rupali Shevalkar, Bharati Asgaonkar, Manali N. Yadav

**Drafting of the manuscript:** Rupali Shevalkar, Manali N. Yadav

**Critical review of the manuscript for important intellectual content:** Rupali Shevalkar, Bharati Asgaonkar

**Supervision:** Bharati Asgaonkar

### Disclosures

**Human subjects:** Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Ethics Committee for Academic Research and Projects issued approval ECARP/2022/74. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

## References

1. Solomon P: Problem-based learning: a review of current issues relevant to physiotherapy education . *Physiother Theory Pract.* 2005, 21:37-49. [10.1080/09593980590911499](https://doi.org/10.1080/09593980590911499)
2. Telang A: Problem-based learning in health professions education: an overview . *Arch Med Health Sci.* 2014, 2:243. [10.4103/2321-4848.144363](https://doi.org/10.4103/2321-4848.144363)
3. Eksteen CA, Slabbert JA: Problem based curricula and problem based learning in physiotherapy: a critical review. *S Afr J Physiother.* 2001, 57:23-8. [10.4102/sajp.v57i4.520](https://doi.org/10.4102/sajp.v57i4.520)
4. Kumar S, Sisodia V, Kumar A: Evidence base for physiotherapy/physical therapy: a specialty-based quantitative trend analysis of articles. *Saudi J Health Sci.* 2013, 2:23. [10.4103/2278-0521.112627](https://doi.org/10.4103/2278-0521.112627)
5. Spector PE: *Survey Design and Measure Development*. Oxford University Press, Oxford, UK; 2013. [10.1093/oxfordhb/9780199934874.013.0009](https://doi.org/10.1093/oxfordhb/9780199934874.013.0009)
6. Zapata-Ospina JP, García-Valencia J: Validity based on content: a challenge in health measurement scales . *J Health Psychol.* 2022, 27:481-93. [10.1177/1359105320953477](https://doi.org/10.1177/1359105320953477)
7. Shi J, Mo X, Sun Z: Content validity index in scale development (Article in Chinese) . *Zhong Nan Da Xue Xue Bao Yi Xue Ban.* 2012, 37:152-5. [10.3969/j.issn.1672-7347.2012.02.007](https://doi.org/10.3969/j.issn.1672-7347.2012.02.007)
8. Yimprayoon P: Sample size determination and power analysis for modified Cohen's kappa statistic . *Appl Math Sci.* 2013, 7:6153-66. [10.12988/ams.2013.39496](https://doi.org/10.12988/ams.2013.39496)
9. Yusoff MSB: ABC of content validation and content validity index calculation . *Educ Med J.* 2019, 11:49-54. [10.21315/eimj2019.11.2.6](https://doi.org/10.21315/eimj2019.11.2.6)
10. Mokkink LB, Terwee CB, Knol DL, et al.: Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. *BMC Med Res Methodol.* 2006, 6:2. [10.1186/1471-2288-6-2](https://doi.org/10.1186/1471-2288-6-2)

11. Wynd CA, Schmidt B, Schaefer MA: Two quantitative approaches for estimating content validity . West J Nurs Res. 2003, 25:508-18. [10.1177/0193945903252998](#)
12. Lukmanul Hakim NAM, Rasidi Pairan M, Ikram Zakaria M: Step-by-step guide to calculating content validity index (CVI) for single constructs using Excel. IJRIS. 2025, 9:1717-26. [10.47772/ijriss.2025.90300135](#)
13. Vélez Silva VN, Palma Zambrano GA, Borroto Cruz ER: Problem-based learning vs. the traditional method in the development of competencies in students in the health area. A systematic review. Salud Cienc Tecnol. 2024, 4:16-8. [10.56294/saludcyt2024.601](#)
14. Abu Bakar N, Bulgiba A, Isahak M: Validity and reliability of the Malay version of the PSC-12 scale among healthcare workers in Malaysia. PLoS One. 2025, 20:e0317744. [10.1371/journal.pone.0317744](#)
15. Patra A, Guha A: Content validity index: a simple but significant approach in social science research . Academia. 2018, 8:3-4. [10.5958/2249-7137.2018.00020.4](#)
16. Fong DY, Ho SY, Lam TH: Evaluation of internal reliability in the presence of inconsistent responses . Health Qual Life Outcomes. 2010, 8:27. [10.1186/1477-7525-8-27](#)
17. Cao H, Sen PK, Peery AF, Dellon ES: Assessing agreement with multiple raters on correlated kappa statistics . Biom J. 2016, 58:935-43. [10.1002/bimj.201500029](#)
18. Peterson RA, Kim Y: On the relationship between coefficient alpha and composite reliability . J Appl Psychol. 2013, 98:194-8. [10.1037/a0030767](#)
19. Souza AC, Alexandre NM, Guirardello EB: Psychometric properties in instruments evaluation of reliability and validity. Epidemiol Serv Saude. 2017, 26:649-59. [10.5123/S1679-49742017000300022](#)
20. Wolf MG, Ihm E, Maul A, Taves A: The response-process-evaluation method: a new approach to survey-item validation. Adv Methods Pract Psychol Sci. 2025, 8:6-8. [10.1177/25152459251353135](#)
21. Temesgen MH, Girma Y, Dugo T, Azeze G, Dejen M, Deres M, Janakiraman B: Factors influencing student's satisfaction in the physiotherapy education program. Adv Med Educ Pract. 2021, 12:133-40. [10.2147/AMEP.S289134](#)