

Development and Psychometric Validation of the Guideline-Concordant Low Back Pain Vignette (GCLBP-V) in Undergraduate Physiotherapy Students

Ahmad Muhsen ¹, Ahmad N Abudoush ², Bodor Bin Sheeha ³, Mohammad Z Darabseh ⁴, Amjad Shallan ¹, Adnan Wshah ¹, Owis Jafar Eilayyan ¹

¹Department of Physical Therapy, Faculty of Applied Medical Sciences, The Hashemite University, Zarqa, Jordan; ²Department of Psychology, Faculty of Arts, The University of Jordan, Amman, Jordan; ³Department of Rehabilitation Sciences, College of Health and Rehabilitation Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia; ⁴Department of Physiotherapy, School of Rehabilitation Sciences, The University of Jordan, Amman, Jordan

Correspondence: Ahmad Muhsen, Department of Physical Therapy, Faculty of Applied Medical Sciences, The Hashemite University, P.O Box 330127, Zarqa, 13133, Jordan, Tel +962 5 3903333, Ext: 5550, Fax +962 5 3903368, Email muhsen_a@hu.edu.jo

Background: Low back pain (LBP) is a leading cause of global disability. Clinical guidelines recommend maintaining activity, early return to work, exercise, and avoiding bed rest. Although vignette-based assessments are common in pain education, they are rarely psychometrically evaluated. This study aimed to develop and validate the Guideline-Concordant Low Back Pain Vignette (GCLBP-V) in undergraduate physiotherapy students.

Methods: A cross-sectional study included 126 undergraduate physiotherapy students in Jordan. The four-item GCLBP-V assesses guideline-concordant recommendations across activity, work, exercise, and bed rest. Content validity was evaluated via expert review ($n = 8$) and cognitive pre-testing. Construct validity was assessed using known-groups comparisons (junior vs senior) and convergent validity with pain knowledge and pain-related beliefs. Test-retest reliability (intraclass correlation coefficient, ICC (2,1)), measurement error (standard error of measurement, SEM; minimal detectable change at 95% confidence, MDC₉₅), internal consistency (Kuder-Richardson Formula 20, KR-20), and score distribution were evaluated in accordance with COSMIN recommendations.

Results: Content validity was excellent (scale-level average content validity index, S-CVI/Ave = 0.969). Known-groups validity was supported by higher scores among senior versus junior students (Cliff's $\delta = 0.510$, $p < 0.001$). The GCLBP-V showed moderate associations with pain knowledge ($\rho = 0.603$) and maladaptive pain-related beliefs ($\rho = -0.466$) ($p < 0.001$). Test-retest reliability was good to excellent (ICC (2,1) = 0.858, 95% CI 0.769–0.921), with SEM = 0.531 and MDC₉₅ = 1.471. Internal consistency was modest (KR-20 = 0.549). Ceiling and floor effects were 18.3% and 12.7%, respectively.

Conclusion: The GCLBP-V demonstrates acceptable psychometric properties as a brief measure of guideline-concordant clinical recommendations. It addresses a key methodological gap by providing a standardized vignette-based behavioral outcome measure suitable for group-level comparisons in pain education. Further research should evaluate responsiveness and generalizability across different populations, disciplines, and settings.

Plain Language Summary: Low back pain is one of the most common causes of disability worldwide. Clinical guidelines recommend that people with low back pain stay active, continue working when possible, exercise, and avoid long periods of bed rest. Researchers often use short written patient stories, called vignettes, to check whether healthcare students and professionals give recommendations that match these guidelines. However, most of these vignette-based tools have not been formally tested to determine whether they accurately assess guideline-based recommendations.

The team developed a short questionnaire called the Guideline-Concordant Low Back Pain Vignette (GCLBP-V). They tested it with undergraduate physiotherapy students in Jordan. The study examined whether the tool was easy to understand, produced stable results over time, and could distinguish between students with different levels of training.



The findings showed that the GCLBP-V performed well across these areas. Senior students scored higher than junior students. Students who knew more about pain and had healthier beliefs about pain tended to score higher on the tool.

This tool may help pain education researchers evaluate whether educational programs are preparing future healthcare professionals to provide recommendations supported by research for people with low back pain.

Keywords: low back pain, pain education, case vignette, clinical decision-making, psychometric validation, physiotherapy students, guideline-concordant recommendations

Introduction

Low back pain remains the leading cause of years lived with disability (YLDs) worldwide and a primary driver of healthcare use and work absence across all regions and income levels.^{1,2} Recent international clinical practice guidelines consistently advocate a biopsychosocial management model for non-specific LBP, emphasizing maintenance of usual activities, early or supported return to work, graded exercise, and avoidance of bed rest.³⁻⁶ These recommendations underpin evidence-based musculoskeletal care.⁴⁻⁷

Despite established guidelines, gaps persist between evidence-based recommendations and clinical decision-making among healthcare professionals and students.^{8,9} This discrepancy has been linked in previous research to residual biomedical attitudes and beliefs about pain and disability, which are associated with guideline-discordant practices, including activity avoidance and delayed return to work.¹⁰ In response, pedagogical frameworks grounded in contemporary pain science, particularly pain neuroscience education (PNE), have emerged to improve pain knowledge and beliefs, and may support more guideline-concordant clinical behaviors.¹¹⁻¹³

Case vignettes have been widely employed in pain education research to approximate clinical behavior through standardized assessment of clinical recommendations, particularly regarding activity, work participation, exercise, and rest following educational interventions.¹¹⁻¹⁴ Vignette-based outcomes have also demonstrated meaningful associations with real-world clinical decision-making in controlled research contexts, supporting their practical utility as proxies for assessing clinical practice.¹⁵ However, vignettes have often been used primarily as convenient assessment formats with limited psychometric evaluation.¹⁰⁻¹³ The absence of such evaluation limits the interpretability of vignette scores, compromises cross-study comparisons, and constrains their suitability for use as primary outcomes in educational trials.

Alongside this methodological gap, a geographical imbalance persists, with pain education research increasingly extending beyond high-income Western countries over recent decades, including emerging work in regions such as the Middle East. This issue is particularly relevant in Jordan, where physiotherapy education has become increasingly established over recent decades.¹⁶ Recent findings also indicate that pain-related knowledge and beliefs among healthcare students may not always align with contemporary pain science, underscoring the need for validated educational outcome measures.¹⁶ Yet, tools to measure students' evidence-based clinical recommendations for low back pain remain limited. This lack of validated tools restricts both local curriculum development and contributions to the international pain education literature.

To address these gaps, we developed the Guideline-Concordant Low Back Pain Vignette (GCLBP-V). This brief, standardized case vignette was independently developed and anchored to international guideline principles to assess clinical recommendations for non-specific LBP across four core domains (daily activities, work, exercise, and bed rest).³⁻⁶ Although conceptually aligned with guideline-based domains used in earlier vignette-based research assessing clinical recommendations for LBP,¹¹⁻¹³ the GCLBP-V incorporates a standardized scoring framework and was specifically designed for formal psychometric evaluation rather than solely as an educational outcome measure. Therefore, this study aimed to develop and psychometrically validate the GCLBP-V in a sample of undergraduate physiotherapy students in Jordan, assessing the instrument's content validity, construct validity, test-retest reliability, measurement error, and score distribution, following COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) recommendations for the evaluation of measurement properties.¹⁷ Establishing the psychometric properties of the GCLBP-V will enhance the methodological rigor of pain research and improve the quality of evidence generated from future trials.

Materials and Methods

Study Design and Setting

This psychometric cross-sectional study was conducted to develop and validate the Guideline-Concordant Low Back Pain Vignette (GCLBP-V). Data were collected between February and April 2025 at the Department of Physical Therapy, Faculty of Applied Medical Sciences, The Hashemite University (HU), Jordan. The study followed the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) recommendations for the development, evaluation, and reporting of measurement properties.¹⁷

Ethics

Ethical approval was granted by the Institutional Review Board of The Hashemite University (HU-IRB No. 33/1/2024/2025). All participants provided written informed consent prior to participation. Students were informed that participation was voluntary, responses were collected anonymously at all time points, and participation would not affect their academic standing or grades. All procedures complied with the Declaration of Helsinki.

Participants

All undergraduate Bachelor of Science (BSc) pre-registration physical therapy (PT) students from the Hashemite University were invited to participate. Inclusion criteria were (1) active enrollment in Years 1–4 of the BSc physical therapy (PT) program and (2) the ability to provide informed consent. Exclusion criteria were current internship or graduation status, or academic leave at the time of data collection.

Recruitment was conducted via email announcements, campus advertisements, and brief in-class presentations. For known-groups validity testing, students were grouped a priori as junior students (Years 1–2) or senior students (Years 3–4). This classification was based on the hypothesis that senior students would demonstrate more guideline-concordant recommendations due to greater exposure to musculoskeletal and pain-related content.

Instrument Development

Conceptual Framework and Item Generation

The Guideline-Concordant Low Back Pain Vignette (GCLBP-V) was independently developed to assess the alignment of clinical recommendations for non-specific low back pain (LBP) with contemporary international practice guidelines. Item generation was informed by core themes consistently emphasized across major clinical practice guidelines, including those issued by the National Institute for Health and Care Excellence (NICE), the American College of Physicians (ACP), the American Physical Therapy Association (APTA), and the US Department of Veterans Affairs/Department of Defense (VA/DoD).^{3–6} These four main domains, maintaining usual daily activities, supporting return to work, engaging in graded exercise, and avoiding bed rest, formed the conceptual foundation of the vignette.

Vignette Construction

The case vignette depicts a working-age adult presenting with non-specific LBP, in the absence of red flags or features suggesting serious spinal pathology. Participants are then instructed to select one clinical recommendation for each of four management domains: daily activities, work, exercise, and bed rest. For each domain, four response options are provided, and only one option is prespecified as guideline-concordant practice, as defined by contemporary guidelines.^{3–6} The GCLBP-V was administered in English, which is the language of instruction within the physiotherapy program. The complete GCLBP-V instrument, including the vignette, response options, scoring instructions, and answer key, is provided in [Supplementary File 1](#).

Scoring

Responses are scored dichotomously: 1 for the guideline-concordant response and 0 for other options. The total score is the sum of the four domain scores (range 0–4), with higher scores reflecting greater guideline concordance. If any domain

response is absent, the total score is recorded as missing (no prorating). Detailed scoring instructions are provided in [Supplementary File 1](#).

Sample Size

Sample size was calculated using Stata/SE version 17 (StataCorp, College Station, TX). For convergent validity, a minimum of 113 participants was required to detect a correlation of $|\rho| = 0.30$, which is considered indicative of a small-to-moderate association in biomedical research,^{18,19} based on a two-sided $\alpha = 0.05$ and 90% statistical power. To account for potential incomplete responses or non-response, the target sample size was increased to 126 participants.

For known-groups validity, the final sample comprised 64 junior and 62 senior students, enabling estimation of effect size using Cliff's δ ,²⁰ with bias-corrected bootstrap confidence intervals. Test-retest reliability was assessed in a stratified subsample of 60 participants; ICC (2,1) estimates were calculated and reported with 95% confidence intervals in accordance with methodological guidance for reliability studies.²¹ The sample size was considered sufficient for reliability and known-groups analyses based on methodological recommendations for psychometric studies.

Score Distribution and Floor/Ceiling Effects

Descriptive statistics (mean, median, standard deviation (SD), interquartile range (IQR), and full range) were calculated for GCLBP-V total scores. Floor and ceiling effects were examined and defined as present if more than 15% of respondents achieved the minimum and maximum possible total score, consistent with COSMIN recommendations.¹⁷

Content Validity

Content validity was assessed through expert panel review followed by cognitive pre-testing with the target population.

Expert Panel

An expert panel (n=8) with backgrounds in physiotherapy education, musculoskeletal rehabilitation, pain science, and clinical research assessed the content validity of the GCLBP-V. Panel members independently rated the relevance of each item and its response options using a 4-point ordinal scale (1: not relevant, 4: highly relevant). Content validity was quantified using: the Item-Level Content Validity Index (I-CVI), the Scale-Level Content Validity Index/Average (S-CVI /Ave), and a modified kappa (k^*) to adjust for chance agreement. Acceptable thresholds for validity were set *a priori* as I-CVI ≥ 0.78 , S-CVI/Ave ≥ 0.90 , and $k^* \geq 0.74$, in accordance with established methods.^{22,23} The panel's qualitative feedback was synthesized and used to guide minor wording refinements, enhancing item clarity and precision without changing the underlying constructs, response structure, or scoring criteria.

Cognitive Pre-Testing

Following expert review, cognitive pre-testing was conducted using a think-aloud approach with a convenience sample of eight final-year PT students. Students verbalized their interpretation and reasoning while completing the vignette, focusing on clarity, comprehension, and interpretation of response options. Feedback prompted final minor refinements (eg, simplifying timeframes), while preserving construct intent and scoring structure.

Construct Validity

To evaluate construct validity, two approaches were used: known-groups and convergent validity analyses, based on a priori hypotheses.

Known-groups validity was tested by comparing GCLBP-V scores between junior (Years 1–2) and senior (Years 3–4) PT students. It was hypothesized that senior students would score higher on the GCLBP-V than their junior counterparts. Effect size was measured using Cliff's delta (δ). Convergent validity was examined by correlating GCLBP-V scores with two theoretically related constructs. Pain neurophysiology knowledge, measured using the Revised Neurophysiology of Pain Questionnaire (RNPQ),²⁴ and pain-related disability beliefs, assessed using the modified Health Care Providers' Pain and Impairment Relationship Scale (HC-PAIRS).²⁵ We hypothesized that GCLBP-V scores would be positively correlated with RNPQ scores and negatively correlated with HC-PAIRS scores. Associations were expected to be of small-to-moderate magnitude (approximately $|\rho| = 0.30$ – 0.50), based on commonly used interpretative frameworks for correlation coefficients in biomedical research.^{18,19}

Reliability and Measurement Error

Test-retest reliability was evaluated in a convenience subsample of 60 students who completed the GCLBP-V on two occasions separated by a 7–10-day interval, during which no formal pain-related instruction was delivered. The retest sample included equal numbers of junior (Years 1–2; $n = 30$) and senior (Years 3–4; $n = 30$) students. Reliability was estimated using the intraclass correlation coefficient (ICC (2,1)), based on a two-way random-effects model with absolute agreement, and reported with 95% confidence intervals.²¹

Measurement error was expressed as the standard error of measurement (SEM) and the minimal detectable change at the 95% confidence level (MDC₉₅). Systematic differences between sessions were assessed using the Wilcoxon signed-rank test. Agreement and potential bias were examined using Bland–Altman analysis (mean difference and limits of agreement).²⁶

Internal Consistency

Internal consistency of the 4-item dichotomous total score was estimated using the Kuder–Richardson Formula 20 (KR-20), which was selected as the appropriate internal consistency measure for dichotomous items, supported by item-difficulty indices, and corrected point-biserial item–total correlations.

Statistical Analysis

All statistical analyses were conducted using Stata/SE version 17 (StataCorp, College Station, TX), with two-sided significance set at $\alpha = 0.05$. All analyses were conducted using complete-case data, with no imputation performed. Complete-case analysis was deemed appropriate given the absence of missing data.

Descriptive statistics (mean, median, standard deviation, interquartile range, and full range) were calculated for GCLBP-V total scores. Floor and ceiling effects were examined descriptively and interpreted relative to the predefined 15% threshold.¹⁷

Content validity indices (I-CVI, S-CVI/Ave, and modified kappa) were calculated according to established methodology.^{22,23} Known-groups validity was examined by comparing junior and senior students using the Mann–Whitney *U*-test, with effect size quantified using Cliff's δ and 95% bias-corrected bootstrap confidence intervals. Convergent validity was evaluated using Spearman's rank correlation coefficients (ρ) between GCLBP-V scores and RNPQ and HC-PAIRS scores. Confidence intervals were estimated using non-parametric bootstrap resampling, with 1000 iterations.

Test–retest reliability was estimated using intraclass correlation coefficients (ICC (2,1)), based on a two-way random-effects model with absolute agreement and reported with 95% confidence intervals.²¹ Measurement error was quantified using the standard error of measurement (SEM) and the minimal detectable change at the 95% confidence level (MDC₉₅). Agreement and potential systematic bias were further evaluated using Bland–Altman methods.²⁶

Internal consistency was estimated using the Kuder–Richardson Formula 20 (KR-20), supplemented by item difficulty indices and corrected point-biserial item–total correlations.

Results

Sample Description

A total of 126 undergraduate physiotherapy students met the eligibility criteria and participated in the validation phase. All 126 participants completed the four GCLBP-V domains; therefore, no item-level missing data or missing total scores were observed. This sample consisted of 64 junior students (Years 1–2) and 62 senior students (Years 3–4), almost evenly distributed across academic years as Year 1 ($n = 32$), Year 2 ($n = 32$), Year 3 ($n = 31$), and Year 4 ($n = 31$). The mean age of the sample was 20.10 years ($SD = 1.37$), with a slight female majority accounting for 59.5% ($n = 75$), and males for 40.5% ($n = 51$). All students provided complete data across the vignette items and the Revised Neurophysiology of Pain Questionnaire (RNPQ) and the modified Health Care Providers' Pain and Impairment Relationship Scale (HC-PAIRS), enabling complete-case analyses ($N = 126$). Descriptive statistics for all variables are presented in [Table 1](#).

Table 1 Descriptive Characteristics of the Study Sample

Variable	Total Sample N = 126	Junior (Y1–2) n = 64	Senior (Y3–4) n = 62
Age (years)	20.10 (1.37)	18.98 (0.70)	21.24 (0.84)
Sex, n (%)			
Female	75 (59.5%)	37 (57.8%)	38 (61.3%)
Male	51 (40.5%)	27 (42.2%)	24 (38.7%)
GCLBP-V	2.02 (1.30)	1.44 (1.08)	2.63 (1.23)
GCLBP-V, median (IQR)	2 (1–3)	1 (1–2)	3 (2–4)
RNPQ	5.99 (1.78)	4.91 (1.42)	7.11 (1.38)
HC-PAIRS	51.28 (8.61)	55.34 (5.00)	47.08 (9.54)

Note: Data are presented as mean (SD) unless otherwise specified.

Abbreviations: GCLBP-V, Guideline-Concordant Low Back Pain Vignette; RNPQ, Revised Neurophysiology of Pain Questionnaire; HC-PAIRS, Health Care Providers’ Pain and Impairment Relationship Scale.

Content Validity

All four vignette domains satisfied the prespecified content validity criteria. Item-level content validity indices (I-CVI) ranged from 0.875 to 1.00, with corresponding modified kappa (k^*) values between 0.871 and 1.00. The scale-level average content validity index (S-CVI/Ave) was 0.969 (Table 2).

Score Distribution, Item Performance, and Internal Consistency

Across the full sample, the full range of possible GCLBP-V scores (0 to 4) was observed, with a mean of 2.02 (SD = 1.30) and a median of 2 (IQR: 1–3) (Table 1). Floor effects were present in 12.7% of participants and ceiling effects in 18.3%. Among senior students (Years 3–4), 29% (18/62) achieved the maximum total score of 4, compared with 1.6% (1/64) of junior students (Years 1–2). Item difficulty indices ranged from 0.444 to 0.548, indicating moderate endorsement of guideline-concordant responses across all domains, with no item demonstrating extremely low or high endorsement.

Table 2 Content Validity, Item Performance, and Internal Consistency of the GCLBP-V

Domain/Metric	I-CVI	Modified Kappa (k^*)	Item Difficulty	Item-Total Correlation
Item-level indices				
Daily activities	0.875	0.871	0.548	0.375
Work	1.000	1.000	0.444	0.296
Exercise	0.875	0.871	0.492	0.332
Bed rest	0.875	0.871	0.468	0.310
Scale-level indices				
S-CVI/Ave	0.969	–	–	–
KR-20	–	–	–	0.549

Notes: Modified kappa (k^*) represents chance-corrected agreement for item-level content validity. Item difficulty represents the proportion of participants selecting the guideline-concordant response.

Abbreviations: GCLBP-V, Guideline-Concordant Low Back Pain Vignette; I-CVI, Item-level Content Validity Index; S-CVI/Ave, Scale-level Content Validity Index/Average; KR-20, Kuder–Richardson Formula 20.

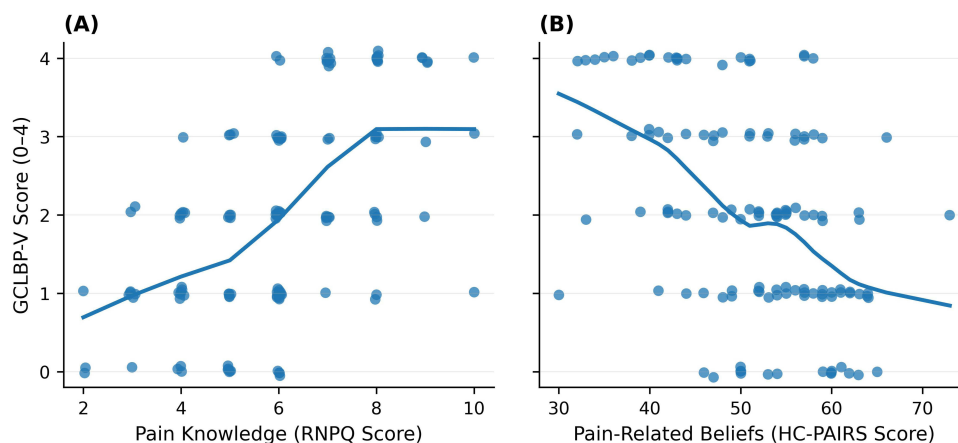


Figure 1 Association between GCLBP-V scores and (A) pain knowledge measured using the Revised Neurophysiology of Pain Questionnaire (RNPQ) and (B) pain-related beliefs measured using the Health Care Providers' Pain and Impairment Relationship Scale (HC-PAIRS). Scatterplots display individual participant data points with locally weighted scatterplot smoothing (LOWESS) curves illustrating the observed relationships. Panel A demonstrates a positive association between GCLBP-V and RNPQ scores, whereas Panel B demonstrates a negative association between GCLBP-V and HC-PAIRS scores.

Abbreviations: GCLBP-V, Guideline-Concordant Low Back Pain Vignette; RNPQ, Revised Neurophysiology of Pain Questionnaire; HC-PAIRS, Health Care Providers' Pain and Impairment Relationship Scale.

Corrected point-biserial item–total correlations ranged from 0.296 to 0.375, and internal consistency (KR-20) was 0.549 (Table 2).

Construct Validity

Known-Groups Validity

GCLBP-V scores were significantly higher among senior students (mean = 2.63, SD = 1.23) compared with junior students (mean = 1.44, SD = 1.08), Mann–Whitney $U = 2995.5$, $p < 0.001$. The effect size was moderate (Cliff's $\delta = 0.510$, 95% CI 0.336 to 0.668).

Convergent Validity

GCLBP-V scores demonstrated a moderate positive correlation with pain neurophysiology knowledge (RNPQ, $\rho = 0.603$, 95% CI 0.488–0.702, $p < 0.001$) and a moderate negative correlation with pain-related disability beliefs (HC-PAIRS, $\rho = -0.466$, 95% CI -0.597 to -0.311 , $p < 0.001$). The LOWESS (Locally Weighted Scatterplot Smoothing) lines demonstrated an increasing trend in GCLBP-V scores with higher RNPQ scores (Figure 1A) and a decreasing trend with increasing HC-PAIRS scores (Figure 1B), consistent with the observed associations.

Reliability and Measurement Error

The intraclass correlation coefficient (ICC (2,1)) was 0.858 (95% CI 0.769 to 0.921), indicating good to excellent test-retest reliability.²¹ There was no systematic difference observed between the two administrations (Wilcoxon signed-rank $p = 0.498$). Measurement error estimates included a standard error of measurement (SEM) of 0.531 and a minimal detectable change at the 95% confidence level (MDC_{95}) of 1.471. Bland–Altman analysis demonstrated a mean difference of 0.067, with limits of agreement from -1.371 to 1.504, indicating acceptable agreement between test and retest scores. The Bland–Altman plot (Figure 2) demonstrated that most observations were distributed within the limits of agreement, with no clear pattern of systematic bias across the range of scores. Detailed reliability and validity metrics are summarized in Table 3.

Discussion

This study developed and psychometrically validated the Guideline-Concordant Low Back Pain Vignette (GCLBP-V), a brief instrument designed to evaluate alignment between clinical recommendations and current evidence-based guidelines for non-specific low back pain. The findings provide converging evidence supporting the instrument's validity and

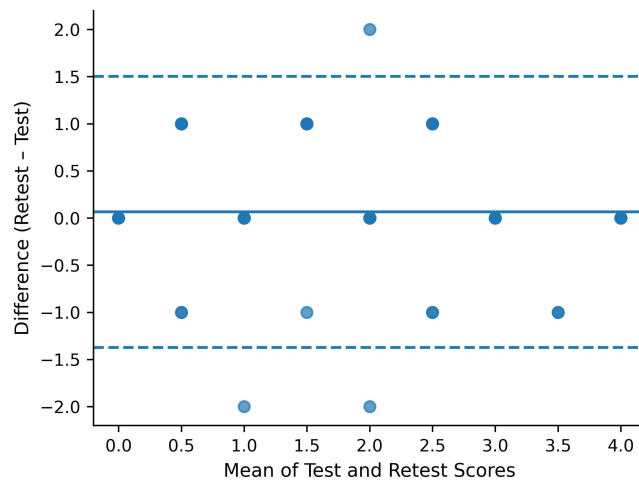


Figure 2 Bland–Altman plot of test–retest agreement for GCLBP-V scores. The solid line represents the mean difference (bias) between test and retest scores, and the dashed lines indicate the 95% limits of agreement. Data points represent individual participants plotted against the mean of the test and retest scores.

reliability. Content validity indices exceeded established thresholds for expert agreement.^{22,23} Construct validity was supported through known-groups discrimination and theoretically consistent correlations with pain-related knowledge and beliefs. The association with pain knowledge ($\rho = 0.603$) exceeded the hypothesized lower bound, suggesting that the GCLBP-V captures clinically meaningful aspects of guideline-concordant decision-making beyond simple knowledge

Table 3 Construct Validity, Reliability, and Measurement Error of the GCLBP-V

Metric	Result	95% CI	p-value
Construct validity			
Known-groups validity			
Mann–Whitney U	2995.5	–	<0.001
Cliff’s δ	0.510	0.336, 0.668	–
Convergent validity			
ρ (GCLBP-V with RNPQ)	0.603	0.488, 0.702	<0.001
ρ (GCLBP-V with HC-PAIRS)	–0.466	–0.597, –0.311	<0.001
Reliability and measurement error			
Test–retest reliability (n = 60)			
ICC (2,1)	0.858	0.769, 0.921	–
Wilcoxon signed-rank p	–	–	0.498
SEM	0.531	–	–
MDC ₉₅	1.471	–	–
Agreement (Bland–Altman)			
Mean difference (bias)	0.067	–	–
Limits of agreement	–1.371, 1.504	–	–

Note: Data are presented as estimates (95% CI) unless otherwise specified.
Abbreviations: GCLBP-V, Guideline-Concordant Low Back Pain Vignette; CI, confidence interval; ICC, intraclass correlation coefficient; ρ , Spearman’s rank correlation coefficient; δ , Cliff’s delta; SEM, standard error of measurement; MDC₉₅, minimal detectable change at the 95% confidence level.

recall. The observed test–retest reliability was good to excellent ($ICC(2,1) = 0.858$, 95% CI 0.769–0.921), satisfying the criteria for group-level comparisons,²¹ with acceptable measurement error and no systematic bias noted.

Case vignettes are commonly employed as standardized measures in healthcare research to approximate clinical behavior and assess adherence to evidence-based recommendations. Their methodological utility has been widely reported; when rigorously constructed and standardized, vignette responses can reasonably approximate aspects of clinical practice under controlled conditions.¹⁵ In pain education research, vignette-based outcomes have been consistently employed to assess clinical recommendations for key low back pain management domains, including activity, work participation, exercise, and bed rest, specifically following pain neuroscience education (PNE) interventions.^{11–13} These findings suggest that vignette-based tools are sensitive to educational interventions and can detect shifts towards more guideline-concordant recommendations.

However, despite their widespread use, vignette-based measures in pain education research have generally been adopted as practical survey or outcome tools for assessing behavioral outcomes rather than as formally developed instruments. Psychometric evaluation of these measures is often underreported, which limits interpretability and comparability of findings across studies. This methodological limitation was highlighted in a recent systematic review and meta-analysis, which reported that, due to the complexity of the construct, behavioral outcomes in pain education research are frequently measured using heterogeneous or non-validated measures.¹⁴ To our knowledge, previous vignette-based studies in pain education, including those by Colleary et al,¹¹ Maguire et al,¹² and Mankelow et al,¹³ have not reported formal psychometric evaluation of their vignette measures, such as content validity indices, test–retest reliability, or measurement error parameters. Accordingly, the present study extends this literature by introducing a structured scoring framework and comprehensive psychometric evaluation, thereby establishing the GCLBP-V as a standardized measurement instrument suitable for educational and research applications. The GCLBP-V may therefore help improve the standardization of behavioral outcome assessment in pain education research.

The observed associations between GCLBP-V scores and measures of pain knowledge and beliefs reinforce the instrument's construct validity. In particular, GCLBP-V scores showed a moderate positive correlation with pain neurophysiology knowledge ($\rho = 0.603$), and a moderate negative correlation with pain-related disability beliefs ($\rho = -0.466$). These findings suggest that better pain knowledge and more adaptive beliefs are linked to more guideline-concordant clinical recommendations. This aligns with the theoretical framework underpinning contemporary pain education, where knowledge and beliefs are considered central determinants of clinical decision-making.^{11–14}

Previous research in pain education has demonstrated that improvements in pain knowledge and attitudes following interventions such as pain neuroscience education (PNE) are often accompanied by shifts toward more guideline-aligned decisions when measured using vignette-based tools.^{11,12} However, the evidence for meaningful behavioral change remains inconclusive, with limited or inconsistent effects observed on clinical recommendations.¹³ These findings are consistent with a recent systematic review and meta-analysis, which reported reliable gains in knowledge and beliefs, but weaker and more variable behavioral effects.¹⁴ While previous studies have focused on effects following educational interventions, the present study provides direct evidence of the relationship between knowledge, beliefs, and clinical recommendations. This finding further supports the construct validity of the GCLBP-V as a measure of guideline-concordant clinical recommendations, suggesting it captures a construct that approximates real-world clinical decision-making more closely than knowledge or belief measures alone.

The GCLBP-V also showed strong known-groups validity, with senior students outperforming their junior counterparts and demonstrating a moderate effect size (Cliff's $\delta = 0.510$). This suggests that the instrument can differentiate between groups with different levels of educational exposure and clinical decision-making ability. Moreover, the effect size implies that the GCLBP-V captures meaningful, rather than insignificant, differences in guideline-concordant recommendations.

These findings align with evidence that pain knowledge and attitudes generally improve with exposure to more pain education, though not consistently observed across healthcare disciplines. In their large-scale international cross-sectional study, Mankelow et al²⁷ reported clinically meaningful gains in pain knowledge and attitudes among physiotherapy students across years of training, compared to minimal change seen in other disciplines, highlighting variability in outcomes. This variability is also evident within pre-registration healthcare curricula, where pain education is often

omitted and inconsistent.²⁸ Consequently, clearer definitions and more standardized pain content are needed within physiotherapy education.²⁹ The GCLBP-V's ability to distinguish between junior and senior physiotherapy students further supports its educational sensitivity and potential utility for curriculum evaluation and benchmarking.

The internal consistency of the GCLBP-V was modest ($KR-20 = 0.549$), which aligns with expectations for brief, multidomain tools.³⁰ The GCLBP-V, unlike unidimensional scales, intentionally includes distinct yet conceptually related domains of evidence-based recommendations: activity, work participation, exercise, and bed rest. In such contexts, high internal consistency is neither expected nor particularly desirable, as it may reflect item redundancy rather than adequate construct coverage.

From a measurement standpoint, reliability coefficients such as $KR-20$ (the dichotomous equivalent of Cronbach's alpha) are determined by both the number of items and the dimensionality of the scale. Short instruments and multidimensional measures typically exhibit lower internal consistency estimates.^{30,31} Therefore, the modest $KR-20$ value should be understood in the light of the instrument's design rather than interpreted as poor measurement quality. This interpretation is further supported by generally acceptable item difficulty indices and corrected item-total correlations. Although one item demonstrated a corrected item-total correlation slightly below the commonly cited 0.30 threshold (0.296), all items exceeded 0.20, a value that has been considered acceptable for exploratory and early-stage instrument development, whereas values of 0.30 or higher are generally regarded as desirable.³² These findings suggest that each item contributes to the overall construct while preserving the conceptual breadth of the instrument.

The GCLBP-V demonstrated good to excellent test-retest reliability ($ICC(2,1) = 0.858$, 95% CI 0.769–0.921), reflecting stable scores over time without intervention. This figure exceeds commonly accepted reliability thresholds for clinical and educational measurement, underscoring the instrument's utility for group-level comparisons.²¹ The instrument's temporal stability is further supported by the absence of systematic bias across repeated administrations. Such reliability is considered appropriate for a brief multidomain instrument of this nature.

Measurement error analysis further clarifies the interpretability of the GCLBP-V. The minimal detectable change at the 95% confidence level ($MDC_{95} = 1.47$) represents the smallest change required to reliably distinguish a true change in an individual's score from measurement error. Although MDC_{95} is primarily intended for interpreting change at the individual level, measurement error is reduced when scores are averaged across groups. Consequently, the GCLBP-V may be capable of detecting smaller between-group or pre-post differences in educational research; however, responsiveness and meaningful group-level change thresholds have not yet been established and require further investigation. Given the limited scale range (0–4), the MDC constitutes a sizeable proportion of the scale, indicating limited sensitivity to detect small individual-level changes. This is a common feature of brief instruments with narrow score ranges, where measurement error represents a considerable proportion of the total scale.¹⁷ Therefore, these findings should be interpreted in light of the instrument's intended use, with the GCLBP-V being most appropriate for group-level comparisons, particularly in educational or intervention contexts.

A modest ceiling effect was observed, marginally exceeding the widely cited 15% threshold.¹⁷ This cut-off should be viewed as a pragmatic reference point rather than an absolute indicator of measurement flaw. Ceiling effects above this level have been reported in widely recognized instruments and typically reflect reduced discriminatory power at the upper end of the scale, rather than a core limitation in measurement precision.³³ Notably, the ceiling effect was concentrated among senior students, with 29% achieving the maximum score. This finding likely reflects greater exposure to evidence-based pain education and clinical training; however, it may also indicate reduced sensitivity for detecting further improvements among more advanced learners.

In this study, the observed value of the ceiling effect likely derives from the combination of a brief four-item format, a binary scoring system, and a trained student cohort, all of which restrict score variability. Yet the instrument preserves its capacity to discriminate between known groups and to display anticipated theoretically convergent patterns with related constructs. Therefore, the ceiling effect reflects the instrument's design and the target population rather than a threat to its validity.

The present findings have significant implications for pain education and research. The GCLBP-V addresses an important gap in the measurement of behavioral outcomes in pain education, where changes in pain knowledge and beliefs are more consistently observed than shifts in clinical recommendations.^{13,14} Educationally, the instrument

provides a practical means to assess guideline-concordant clinical decision-making and adherence in pre-registration physiotherapy programs, facilitating curriculum evaluation and benchmarking across training years. For research, the GCLBP-V's established psychometric properties justify its application as a standardized proxy for clinical behavior in pain education intervention studies, particularly for between-group comparisons. The GCLBP-V assesses guideline-concordant clinical recommendations using a standardized vignette format, thereby contributing to a more ecologically relevant and methodologically robust measure of educational impact.

From an educational perspective, the GCLBP-V was intentionally developed as a brief and standardized assessment of guideline-concordant recommendations for a commonly encountered presentation of non-specific low back pain. Although real-world clinical practice often involves more complex and overlapping presentations, standardized vignettes offer a practical approach for assessing key clinical recommendations while minimizing sources of extraneous variability. Future research may explore more complex vignette scenarios incorporating diagnostic uncertainty, psychosocial factors, or coexisting conditions to further evaluate clinical reasoning and decision-making in pain management.

Limitations

Several limitations should be acknowledged when interpreting the present findings. First, the study sample was limited to physiotherapy students from a single institution in Jordan, which may limit generalizability to other healthcare disciplines, educational contexts, and practicing clinicians. Validation across diverse cultural and educational settings and professional groups is therefore warranted. Second, the GCLBP-V is based on a single vignette scenario with a limited set of domains; although it serves as a proxy for clinical behavior, it may not fully capture real-world practice or the full scope of decision-making in low back pain management. Third, the brief four-item dichotomous format, while feasible, constrains score variability and may reduce sensitivity to small individual-level changes due to the magnitude of measurement error relative to the scale range. Fourth, because the instrument was administered in English, findings may not generalize to populations with different levels of English proficiency. Finally, longitudinal and intervention-based studies are needed to evaluate responsiveness to change, which cannot be assessed within the present cross-sectional design.

Conclusion

This study presents the development and psychometric validation of the GCLBP-V, a brief standardized instrument designed to assess guideline-concordant recommendations in non-specific low back pain in a cohort of undergraduate physiotherapy students in Jordan. The findings demonstrate acceptable and consistent validity and reliability, supporting the instrument's utility as a measure of guideline-concordant clinical recommendations. The GCLBP-V addresses a key gap in psychometrically evaluated vignette-based instruments and may strengthen methodological rigor in behavioral outcome assessment within pain education research. Furthermore, its development within a Middle Eastern educational setting helps address a geographical gap in the existing literature. The instrument's brevity and standardized format facilitate its application in educational and research contexts, particularly for group-level comparisons. Future research should evaluate responsiveness and extend validation across diverse populations, disciplines, and settings.

Abbreviations

ACP, American College of Physicians; APTA, American Physical Therapy Association; BSc, Bachelor of Science; COSMIN, Consensus-based Standards for the selection of health Measurement Instruments; GCLBP-V, Guideline-Concordant Low Back Pain Vignette; HC-PAIRS, Health Care Providers' Pain and Impairment Relationship Scale; HU, Hashemite University; I-CVI, Item-Level Content Validity Index; ICC, Intraclass Correlation Coefficient; IQR, Interquartile Range; k^* Modified kappa coefficient; KR-20, Kuder–Richardson Formula 20; LBP, Low Back Pain; MDC₉₅, Minimal Detectable Change at the 95% Confidence Level; NICE, National Institute for Health and Care Excellence; PNE, Pain Neuroscience Education; PT, Physiotherapy; RNPQ, Revised Neurophysiology of Pain Questionnaire; S-CVI /Ave, Scale-Level Content Validity Index/Average; SD, Standard Deviation; SEM, Standard Error of Measurement; VA/DoD, Veterans Affairs/Department of Defense; YLDs, Years Lived with Disability.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Ethics Approval and Informed Consent

Ethical approval for this study was obtained from the Institutional Review Board of The Hashemite University (HU-IRB No. 33/1/2024/2025) prior to study commencement. The study was conducted in accordance with the principles of the Declaration of Helsinki. All participants were provided with information about the study and gave informed consent prior to participation. Participation was voluntary, had no bearing on academic standing or grades, and responses were collected anonymously at all time points.

Acknowledgments

The authors express their gratitude for the support received from Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2026R422), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors also acknowledge The Hashemite University, Jordan, the host institution for this study, for providing the necessary facilities and support. The authors would also like to gratefully acknowledge all physiotherapy students who contributed their time to support the study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This project was funded by the Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2026R422), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Disclosure

The authors report no conflicts of interest in this work.

References

1. GBD 2021 Low Back Pain Collaborators. Global, regional, and national burden of low back pain, 1990–2020, its attributable risk factors, and projections to 2050: a systematic analysis of the Global Burden of Disease Study 2021. *Lancet Rheumatol.* 2023;5(6):e316–e329. doi:10.1016/S2665-9913(23)00098-X
2. Hartvigsen J, Hancock MJ, Kongsted A, et al. What low back pain is and why we need to pay attention. *Lancet.* 2018;391(10137):2356–2367. doi:10.1016/S0140-6736(18)30480-X
3. National Institute for Health and Care Excellence. Low back pain and sciatica in over 16s: assessment and management (NG59). London: National Institute for Health and Care Excellence; 2016 [updated 2020]. Available from: <https://www.nice.org.uk/guidance/ng59>. Accessed February 22, 2025.
4. Qaseem A, Wilt TJ, McLean RM, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med.* 2017;166(7):514–530. doi:10.7326/M16-2367
5. George SZ, Fritz JM, Silfies SP, et al. Interventions for the management of acute and chronic low back pain: revision 2021. *J Orthop Sports Phys Ther.* 2021;51:CPG1–CPG60. doi:10.2519/jospt.2021.0304
6. U.S. Department of Veterans Affairs, U.S. Department of Defense. VA/DoD clinical practice guideline for the diagnosis and treatment of low back pain. Version 3.0. Washington, DC: U.S. Department of Veterans Affairs and U.S. Department of Defense; 2022. Available from: <https://www.healthquality.va.gov/guidelines/Pain/lbp/>. Accessed February 22, 2025.
7. Dahm KT, Brurberg KG, Jamtvedt G, Hagen KB. Advice to rest in bed versus advice to stay active for acute low-back pain and sciatica. *Cochrane Database Syst Rev.* 2010;2010(6):CD007612. doi:10.1002/14651858.CD007612.pub2
8. Ladeira CE, Cheng MS, da Silva RA. Clinical specialization and adherence to evidence-based practice guidelines for low back pain management: a survey of US physical therapists. *J Orthop Sports Phys Ther.* 2017;47(5):347–358. doi:10.2519/jospt.2017.6561
9. Briggs AM, Slater H, Bunzli S, et al. Consumers' experiences of back pain in rural Western Australia: access to information and services, and self-management behaviours. *BMC Health Serv Res.* 2012;12:357. doi:10.1186/1472-6963-12-357

10. Bishop A, Foster NE, Thomas E, Hay EM. How does the self-reported clinical management of patients with low back pain relate to the attitudes and beliefs of health care practitioners? A survey of UK general practitioners and physiotherapists. *Pain*. 2008;135(1–2):187–195. doi:10.1016/j.pain.2007.11.010
11. Colleary G, O’Sullivan K, Griffin D, Ryan CG, Martin DJ. Effect of pain neurophysiology education on physiotherapy students’ understanding of chronic pain, clinical recommendations and attitudes towards people with chronic pain: a randomised controlled trial. *Physiotherapy*. 2017;103(4):423–429. doi:10.1016/j.physio.2017.01.006
12. Maguire N, Chesterton P, Ryan C. The effect of pain neuroscience education on sports therapy and rehabilitation students knowledge, attitudes, and clinical recommendations toward athletes with chronic pain. *J Sport Rehabil*. 2019;28(5):438–443. doi:10.1123/jsr.2017-0212
13. Mankelov J, Ryan C, Taylor P, Martin D. The effect of pain neurophysiology education on healthcare students’ knowledge, attitudes and behaviours: a mixed-methods randomized controlled trial. *Musculoskelet Sci Pract*. 2020;50:102–249. doi:10.1016/j.msksp.2020.102249
14. Mankelov J, Ryan C, Taylor P, Atkinson G, Martin D. A systematic review and meta-analysis of the effects of biopsychosocial pain education upon health care professional pain attitudes, knowledge, behavior and patient outcomes. *J Pain*. 2022;23(1):1–24. doi:10.1016/j.jpain.2021.06.010
15. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA*. 2000;283(13):1715–1722. doi:10.1001/jama.283.13.1715
16. Muhsen A, Madi M, Wshah A, et al. Pain knowledge, attitudes, and beliefs among final-year physiotherapy and occupational therapy students in Jordan: a cross-sectional study. *J Pain Res*. 2025;18:6243–6257. doi:10.2147/JPR.S565885
17. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170. doi:10.1007/s11136-018-1829-0
18. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018;126(5):1763–1768. doi:10.1213/ANE.0000000000002864
19. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. 2012;24(3):69–71.
20. Vargha A, Delaney HD. A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *J Educ Behav Stat*. 2000;25(2):101–132. doi:10.3102/10769986025002101
21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
22. Lynn MR. Determination and quantification of content validity. *Nurs Res*. 1986;35(6):382–385. doi:10.1097/00006199-198611000-00017
23. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health*. 2007;30(4):459–467. doi:10.1002/nur.20199
24. Catley MJ, O’Connell NE, Moseley GL. How good is the neurophysiology of pain questionnaire? A Rasch analysis of psychometric properties. *J Pain*. 2013;14:818–827. doi:10.1016/j.jpain.2013.02.008
25. Houben RM, Vlaeyen JW, Peters M, et al. Health care providers’ attitudes and beliefs towards common low back pain: factor structure and psychometric properties of the HC-PAIRS. *Clin J Pain*. 2004;20:37–44. doi:10.1097/00002508-200401000-00008
26. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.
27. Mankelov J, Ryan CG, Taylor PC, et al. International, multi-disciplinary, cross-sectional study of pain knowledge and attitudes in nursing, midwifery and allied health professions students. *BMC Med Educ*. 2022;22(1):547. doi:10.1186/s12909-022-03488-3
28. Briggs EV, Carr EC, Whittaker MS. Survey of undergraduate pain curricula for healthcare professionals in the United Kingdom. *Eur J Pain*. 2011;15(8):789–795. doi:10.1016/j.ejpain.2011.01.006
29. Reezigt R, Beetsma A, Köke A, Hobbelen H, Reneman M. Toward consensus on pain-related content in the pre-registration, undergraduate physical therapy curriculum: a Delphi-study. *Physiother Theory Pract*. 2024;40(5):1040–1053. doi:10.1080/09593985.2022.2144562
30. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80(1):99–103. doi:10.1207/S15327752JPA8001_18
31. Tavakol M, Dennick R. Making sense of Cronbach’s alpha. *Int J Med Educ*. 2011;2:53–55. doi:10.5116/ijme.4dfb.8dfd
32. DeVellis RF, Thorpe CT. *Scale Development: Theory and Applications*. 5th ed. Thousand Oaks (CA): Sage Publications; 2021.
33. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307. doi:10.1007/BF01593882

Journal of Pain Research

Publish your work in this journal

The Journal of Pain Research is an international, peer reviewed, open access, online journal that welcomes laboratory and clinical findings in the fields of pain research and the prevention and management of pain. Original research, reviews, symposium reports, hypothesis formation and commentaries are all considered for publication. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-pain-research-journal>

Dovepress
Taylor & Francis Group