

Validity and Reliability of the Clinical Competency Evaluation Instrument for Use among Physiotherapy Students

Pilot study

*Zailani Muhamad,¹ Ayiesah Ramli,¹ Salleh Amat²

صحة وموثوقية أداة تقييم الكفاءات السريرية لاستخدامها في تقييم طلاب العلاج الطبيعي دراسة تجريبية

زيلاني محمد، ايسه راملي، صالح أمات

ABSTRACT: Objectives: The aim of this study was to determine the content validity, internal consistency, test-retest reliability and inter-rater reliability of the Clinical Competency Evaluation Instrument (CCEVI) in assessing the clinical performance of physiotherapy students. **Methods:** This study was carried out between June and September 2013 at University Kebangsaan Malaysia (UKM), Kuala Lumpur, Malaysia. A panel of 10 experts were identified to establish content validity by evaluating and rating each of the items used in the CCEVI with regards to their relevance in measuring students' clinical competency. A total of 50 UKM undergraduate physiotherapy students were assessed throughout their clinical placement to determine the construct validity of these items. The instrument's reliability was determined through a cross-sectional study involving a clinical performance assessment of 14 final-year undergraduate physiotherapy students. **Results:** The content validity index of the entire CCEVI was 0.91, while the proportion of agreement on the content validity indices ranged from 0.83–1.00. The CCEVI construct validity was established with factor loading of ≥ 0.6 , while internal consistency (Cronbach's alpha) overall was 0.97. Test-retest reliability of the CCEVI was confirmed with a Pearson's correlation range of 0.91–0.97 and an intraclass coefficient correlation range of 0.95–0.98. Inter-rater reliability of the CCEVI domains ranged from 0.59 to 0.97 on initial and subsequent assessments. **Conclusion:** This pilot study confirmed the content validity of the CCEVI. It showed high internal consistency, thereby providing evidence that the CCEVI has moderate to excellent inter-rater reliability. However, additional refinement in the wording of the CCEVI items, particularly in the domains of safety and documentation, is recommended to further improve the validity and reliability of the instrument.

Keywords: Clinical Competence; Physiotherapy Speciality; Validity and Reliability; Malaysia.

الملخص: الهدف: من هذه الدراسة هو تحديد صلاحية المحتوى، والاتساق الداخلي، وموثوقية الاختبار وإعادة الاختبار وموثوقية ما بين التصنيفات لأداة تقييم الكفاءة السريرية (CCEVI) في تقييم الأداء السريري لطلاب العلاج الطبيعي. **الطريقة:** أجريت هذه الدراسة في الفترة ما بين شهري يوليو وسبتمبر 2013 في جامعة كيانغسان ماليزيا (UKM) تم تحديد عشرة خبراء لتحديد صلاحية المحتوى من خلال تقييم وتصنيف كل من البنود المستخدمة في CCEVI فيما يتعلق بأهميتها في قياس الكفاءة السريرية لدى الطلاب. تم تقييم 50 طالب من طلاب العلاج الطبيعي في المرحلة الجامعية خلال تسببهم السريري لتحديد صلاحية البنود المستخدمة. تم تحديد موثوقية الأداة من خلال دراسة مستعرضة تقييم الأداء السريري ل 14 من طلاب العلاج الطبيعي في السنة الجامعية النهائية. **النتائج:** مؤشر صحة المحتوى لأداة ال CCEVI كانت 0.91، بينما تراوحت النسبة المئوية للاتفاق على محتوى مؤشرات صحة (0.83–1.00). تم تثبيت الصلاحية البنائية ال CCEVI بعامل تحميل ≥ 0.6 ، في حين كان الاتساق الداخلي (كرونباخ ألفا) عموماً 0.97. وتم تأكيد موثوقية الاختبار وإعادة الاختبار لل CCEVI بمدى ارتباط بيرسون بين (0.91–0.97) وكان مدى معامل ارتباط التصنيف المتداخل ل CCEVI يتراوح بين (0.95–0.98) أما موثوقية ما بين المقيمين ل CCEVI فكانت بين (0.59–0.97) في التقديرات الأولية واللاحقة. **الخلاصة:** أكدت هذه الدراسة التجريبية صحة محتوى ال CCEVI وبينت أن الاتساق الداخلي لهذه الأداة مرتفع، وبالتالي تتوفر الأدلة على أن أداة ال CCEVI لديها موثوقية ما بين المقيمين متوسطة إلى ممتازة. ومع ذلك، فمن المستحسن أن يتم بعض الصقل الإضافي في صياغة البنود لأداة ال CCEVI ولا سيما في مجالات السلامة والوثائق لتحسين صحة وموثوقية هذه الأداة.

مفتاح الكلمات: الكفاءة السريرية؛ تخصص العلاج الطبيعي؛ الصحة والموثوقية؛ ماليزيا.

ADVANCES IN KNOWLEDGE

- The results of this study suggest that the Clinical Competency Evaluation Instrument (CCEVI) is a valuable preliminary instrument with psychometric properties for assessing the clinical competency of physiotherapy students in physiotherapy education programmes.
- This study demonstrated that the CCEVI has content validity and moderate to excellent inter-rater reliability.

APPLICATION TO PATIENT CARE

- Confirming the validity and reliability of the CCEVI ensures that it can effectively assess physiotherapy graduates' clinical competency, thereby verifying that graduates are providing quality health services in patient care and upholding patient safety standards.

THE COMPETENCY OF PHYSIOTHERAPY graduates is becoming a central issue of discussion among physiotherapy clinical educators and academic faculty experts in the healthcare profession.¹ The main concern is the instrument used to evaluate the clinical performance of students as a measure of competency.^{2,3} Such instruments should demonstrate psychometric properties that are valid and reliable.⁴⁻⁷

The increasing number of higher educational institutions that offer physiotherapy programmes has led to a vast variation in curriculum design and assessment approaches. In terms of the assessment of clinical competence, many academic programmes have developed their own assessment instrument that fulfils the needs of their curriculum. In most cases, the instruments used for evaluation are not standardised and differ between institutions.^{8,9} As a consequence, the quality of physiotherapy graduates qualifying for entry level positions in professional practice varies between institutions, potentially compromising the overall standard of physiotherapy care provided to patients. According to Wass *et al.*, there is a need to develop an assessment instrument for healthcare students that is accurate and able to measure clinical competence objectively.⁷ Therefore, the validity and reliability of an instrument is crucial in ensuring that it accurately measures the concepts/attributes that need to be measured according to a curriculum's requirements.¹⁰

Various assessment instruments have been developed and used by physiotherapy programmes around the world, such as the Physiotherapy Clinical Performance Instrument (PTCPI) which is used in the United States and Canada,¹¹ and the Assessment of Physiotherapy Practice used by physiotherapy programmes in Australia and New Zealand.¹² These two instruments are used to evaluate students' clinical competency at the entry level of practice. Similarly, tools such as the Clinical Internship Evaluation Tool, are used to evaluate students' clinical competency with regards to patient management skills.¹³

As a pioneer institution offering the first baccalaureate programme in physiotherapy in Malaysia, the academic staff members of the Physiotherapy Programme in the Faculty of Health

Sciences at University Kebangsaan Malaysia (UKM) in Kuala Lumpur, Malaysia, developed the Clinical Competency Evaluation Instrument (CCEVI). This instrument was developed to suit the local sociocultural context and the UKM physiotherapy curriculum with the aim of evaluating the clinical competency of UKM's physiotherapy students. To the best of the author's knowledge, no investigations of the psychometric properties of this instrument had previously been carried out. Therefore, the objective of this study was to determine the content and construct validity, test-retest reliability, internal consistency and the inter-rater reliability of the CCEVI among physiotherapy student at UKM.

Methods

This pilot study was carried out between June and September 2013. There were two phases to the methodology. Phase one aimed to determine the content validity of the CCEVI questionnaire, while phase two involved a test run of the questionnaire in order to determine the construct validity and reliability of the instrument. The CCEVI was administered in English.

The original version of the CCEVI consisted of 42 items in eight domains: (1) subjective; (2) objective; (3) analysis; (4) treatment; (5) plan and education; (6) safety; (7) documentation, and (8) *viva*. Subsequently, the Subjective, Objective, Treatment and Plan and Education domains were further subdivided into subscales of knowledge, skills and professional traits.

In June 2013, content validation in phase one was performed to improve the original version of the CCEVI questionnaire that was initially developed by the UKM Physiotherapy Task Force. A panel of 10 experts were identified with each expert possessing more than 10 years of experience in clinical teaching and evaluation of students' performance; their experience ranged from 10–24 years (mean: 18.9 years). Six of these experts were academicians (from UKM, the Mara University of Technology in Shah Alam, Malaysia, or the Training Division of the Malaysian Ministry of Health) and the remaining four

were clinical educators from four different teaching hospitals within Klang Valley in Kuala Lumpur. A copy of the CCEVI questionnaire was attached together with the item evaluation, which was then sent to the expert panel for review.¹⁴

A total of three indices (relevance, clarity and representativeness) were used to determine the content validity of each item in the instrument. A Likert-type rating scale of 1–4 was used to rate each item of the indices (1 = not relevant, 2 = somewhat relevant, 3 = relevant and 4 = very relevant). The completed rating scores from the experts were then collected to calculate the item content validity index (I-CVI) and the overall content validity of the instrument. The panel of experts were encouraged to give written feedback and recommendations on the overall structure of the CCEVI. The I-CVI of the entire instrument was calculated based on the proportion of items in the instrument that achieved a relevant rating by the content experts. It has been shown that an acceptable content validity index (CVI) score from a panel of 3–5 experts is 1.00, while a minimum CVI score of 0.78 is required for a panel of 6–10 experts.^{14,15}

Following the analysis of the content validity of the instrument, amendments were made to the initial version of the CCEVI. Although the original version of the CCEVI had 42 items, the revised version had been reduced to 40 items. To score the students' performance, a grading of a 5-point Likert scale ranging from 0–4 (0 = not competent, 1 = poor, 2 = fair, 3 = good and 4 = excellent) was used to reflect clinical competency. The revised version of the CCEVI was then sent back to the panel of experts to re-evaluate the clarity, appropriate use of language and overall presentation of the instrument. The feedback and comments were revised until no further changes were brought up by the experts. The CVI and inter-rater agreement of the revised instrument were then calculated again in order to compare it with the initial version of the CCEVI.

The final revised version of the CCEVI was then pilot tested to determine its reliability and validity. Over the period of July to September 2013, a cross-sectional pilot study was carried out using convenience sampling. A new set of five experts were invited to participate in the study. These experts were clinical physiotherapy educators working at a teaching hospital in Kuala Lumpur, with clinical experience ranging from 9–20 years (mean: 13.6 years). In addition, UKM undergraduate physiotherapy students in their final year of study, who had completed a minimum of six weeks of clinical placement, were also asked to join the study; a total of 50 students volunteered to participate (mean age: 23.3 years). All of the participants educators

and students were briefed on the conduct of study. The clinical educators were requested to assess the clinical competency of the students during their clinical placement using the revised CCEVI questionnaire. After assessment, a total of 50 completed CCEVI questionnaires were collected from the educators to determine the construct validity of the instrument.

In phase two, a test-retest method was used to determine the reliability of the instrument in. Two of the five aforementioned experts were randomly selected and were requested to conduct a screening at the physiotherapy outpatient department of Sungai Buloh Hospital in Kuala Lumpur. The purpose of the screening was to select patients with similar musculoskeletal problems who could be assessed and treated by physiotherapy students in a clinical competency assessment.

The selected patients were randomly assigned to 14 final-year UKM undergraduate physiotherapy students. These students were then assessed by the two aforementioned clinical educators as they carried out their assessment and treatment of the selected patients. The educators were requested to independently score the students' performances using the revised CCEVI and were not allowed to discuss the marks they had allocated to the students. The evaluation process was repeated again after a one-week interval. The 14 students were evaluated for a second time by the same clinical educators using the CCEVI and while assessing and treating the same patients in the same setting. The outcomes of the two assessments were then statistically analysed to determine stability, internal consistency and inter-rater reliability.

Data were analysed using the Statistical Package for the Social Sciences (SPSS) Version 20.0 (IBM Corp., Chicago, Illinois, USA). To calculate the I-CVI, scores were divided into two groups; relevant (with a score of 3 or 4) versus not relevant (with a score of 1 or 2). The I-CVI for each item on the CCEVI was calculated as the number of experts giving a rating of 3 (relevant) or 4 (very relevant) divided by the total number of experts. The CVI for the entire CCEVI was recalculated based on the percentage of total items rated by the experts as either 3 or 4. A CVI score of ≥ 0.80 was considered acceptable.¹⁴ The inter-rater agreement was calculated as the percentage of the CCEVI questionnaire that was considered relevant or very relevant by all experts.

To establish the construct validity of the instrument, each item in each of the CCEVI domains was evaluated using the principal components factor analysis with varimax rotation and Kaiser normalisation. Bartlett's test of sphericity was performed to determine the significance ($P \leq 0.005$) of correlation among the

Table 1: Content validity index of the revised Clinical Competency Evaluation Instrument among physiotherapy students in Malaysia

Domain	Construct	Item no.	CVI		
			Relevance	Clarity	Representativeness
Assessment	Knowledge	A 1	1.00	1.00	1.00
		A 2	1.00	1.00	1.00
		A 3	1.00	1.00	1.00
		A 4	1.00	1.00	1.00
		A 5	1.00	1.00	1.00
	Skills	A 6	1.00	1.00	1.00
		A 7	1.00	0.60	0.60
		A 8	1.00	1.00	1.00
		A 9	1.00	1.00	1.00
		A 10	1.00	0.70	1.00
		A 11	1.00	1.00	1.00
	Professional traits	A 12	1.00	0.70	0.80
		A 13	1.00	0.70	1.00
		A 14	0.90	0.80	0.80
Analysis	Knowledge	B 1	1.00	1.00	0.90
		B 2	1.00	1.00	1.00
		B 3	1.00	0.80	0.90
		B 4	1.00	1.00	1.00
		B 5	1.00	1.00	1.00
Treatment	Knowledge	C 1	0.80	1.00	1.00
		C 2	1.00	0.90	0.90
		C 3	1.00	1.00	1.00
	Skills	C 4	1.00	1.00	1.00
		C 5	1.00	1.00	1.00
		C 6	0.90	0.70	0.90
		C 7	0.90	1.00	1.00
		C 8	1.00	1.00	1.00
	Professional traits	C 9	0.80	0.80	1.00
		C 10	1.00	0.70	1.00
Patient and caregiver education	Knowledge	D 1	1.00	1.00	1.00
		D 2	1.00	1.00	1.00
	Skills	D 3	1.00	1.00	1.00
		D 4	1.00	1.00	0.60
	Professional traits	D 5	1.00	0.70	1.00
Safety	Skills	E 1	1.00	1.00	1.00
		E 2	1.00	1.00	1.00
		E 3	1.00	1.00	1.00

Documentation	Skills	F 1	1.00	1.00	1.00
		F 2	1.00	1.00	1.00
		F 3	1.00	1.00	1.00
CVI of content indices			1.00	0.83	0.95
Inter-rater agreement			0.88	0.73	0.80
CVI of entire instrument*			109/120 = 0.91		

No. = number; CVI = content validity index.

*Number of items with CVI of >0.80 divided by total number of items.

items. The factor analysis needed a bigger sample size; however, the Kaiser-Meyer-Olkin (KMO) measure (cut-off value: 0.60) was used to determine the sampling adequacy. Due to the small sample size of this pilot study, factor analysis was run for each domain instead of the entire instrument. The internal consistency of each domain was established using Cronbach's alpha reliability coefficient following the completion of the exploratory factor analysis.

For the inter-rater reliability, the two-way random effect model intraclass coefficient correlation (ICC 2,1) at a 95% confidence interval was used. Data were computed based on the percentage of the total score in each domain for the initial and repeated evaluations. The scores between the raters were compared to ascertain agreement. The following ICC values were set: ≤ 0.40 indicated poor reliability, $0.40-0.75$ signified fair to good reliability and ≥ 0.75 indicated excellent reliability.¹⁶ The stability of the instrument was examined through Pearson's correlation coefficient and ICC from two evaluations within a one week interval.

Approval from the Ethical Committee Board of UKM was granted (NN-090-2013) prior to the study and written consent was obtained from all of the clinical educators/experts and physiotherapy students included in the study.

Results

In the original version of the CCEVI, the initial 42 items were reviewed by experts for content validity. Qualitative and quantitative data were analysed and recommendations from the written feedback were reviewed. Quantitative analysis of the items demonstrated that the CVI for relevance, clarity and representativeness was 0.95 (40/42), 0.30 (13/42) and 0.67 (28/42), respectively. To further establish the content validity, the items with a CVI of < 0.80 were rephrased. The overall CVI of the entire instrument was found to be 0.64 (81/126). The inter-rater agreement for relevancy, clarity and representativeness was 0.79, 0.19 and 0.24, respectively. As a result, most of the initial items needed rephrasing/rewording to improve their

clarity and brevity. For example, the experts suggested merging the subjective and objective domains into one domain, assessment, in order to avoid redundant items in the assessment of the professional traits in both domains. There was also a suggestion that the five items (items 38–42) in the *viva* domain be relocated to the subscale of knowledge as this could be evaluated in the individual respective domains. The experts also commented on the inadequacy of some items in the documentation domain. This was addressed and one item was added to the subscale of professional traits in the treatment domain: "to comply with professional and ethical standards of practice".

Consequently, the revised CCEVI contained 40 items for measuring clinical competency in six domains; 14 assessment items; five analysis items; 10 treatment items; five patient and caregiver education items; three safety items, and three documentation items. After the revised version was sent back to the same panel of experts for their evaluation, and was subsequently further revised until no issues were highlighted by the experts, the CVI and inter-rater agreement were recalculated. The final revised version of the CCEVI showed improvement in the content validity in all three indices and the entire instrument [Table 1]. The items' CVI for relevance, clarity and representativeness was 1.00 (40/40), 0.83 (33/40) and 0.95 (38/40), respectively. The CVI for the entire instrument improved from 0.64 (81/126) to 0.91 (109/120). The inter-rater agreement for relevancy, clarity and representativeness were 0.88, 0.73 and 0.80 respectively [Table 1].

An exploratory factor analysis was employed to confirm the construct validity of each item in the instrument. When the KMO test for sampling adequacy ($KMO \geq 0.6$) and Bartlett's test of sphericity for the significance ($P \leq 0.005$) of correlation among the items were carried out, all items within the revised CCEVI met the criteria for both tests. A factor analysis on the items within each domain was run to ascertain the dimension among the items and whether the patterns fit well into each construct. A cut-off value of communalities of 0.5 was set before running the factor

Table 2: Factor loading and internal consistency of the revised Clinical Competency Evaluation Instrument among physiotherapy students in Malaysia

Domain and item no.	Factor loading	Internal consistency*
A: Assessment		0.94
A1	0.68	
A2	0.70	
A3	0.84	
A4	0.85	
A5	0.83	
A6	0.76	
A7	0.75	
A9	0.75	
A10	0.81	
A11	0.87	
A13	0.90	
B: Analysis		0.94
B1	0.88	
B2	0.94	
B4	0.96	
B5	0.91	
C: Treatment		0.95
C1	0.65	
C2	0.79	
C3	0.73	
C5	0.72	
C6	0.71	
C7	0.75	
C8	0.87	
C9	0.68	
D: Patient and caregiver education		0.95
D1	0.94	
D2	0.94	
D3	0.93	
D4	0.87	
D5	0.92	
E: Safety		0.79
E1	0.79	
E2	0.83	
E3	0.93	

F: Documentation		0.88
F1	0.88	
F2	0.93	
F3	0.91	
Overall		0.97

No. = number.

*Using Cronbach's alpha.

extraction. Items with a factor loading of ≥ 0.60 with an eigenvalue greater than 1.00 were accepted. Items A12 and A14 in the assessment domain, item B3 in the analysis and items C4 and C10 in the treatment domain were identified as problematic based on insignificant values in the correlation matrix table, indicating that the value on the communalities was either too low (≤ 0.40) or too high (≥ 0.9) [Table 2]. As a result, these items were eliminated from the study. The internal consistency using Cronbach's alpha was recalculated for each domain after these items were deleted, resulting in 35 items.

The factor loading of each item in their respective domains (assessment, analysis, treatment, patient and caregiver education, safety and documentation) was acceptable (≥ 0.6). The internal consistency of each domain was good to high, with the highest internal consistency observed in the patient and caregiver education domain (Cronbach's alpha: 0.95) and the lowest internal consistency in the safety domain (Cronbach's alpha: 0.79). The internal consistency overall for the CCEVI was 0.97 [Table 2].

The test-retest reliability further confirmed the stability of the CCEVI indicating a strong consistency between Pearson's correlation (r) (range: 0.91–0.97) and the ICC (range: 0.95–0.98) [Table 3].

The inter-rater reliability (ICC 2,1) was determined by comparing the total score of each domain between the two raters on the initial and subsequent evaluation separately. As observed in Table 4, the inter-rater correlation coefficient of the initial evaluation showed that the assessment, analysis, treatment, patient and caregiver education and documentation domains had excellent reliability (ICC range: 0.81–0.99). Only the safety domain showed moderate inter-rater reliability (ICC: 0.59). The inter-rater correlation coefficient on the subsequent evaluation indicated four domains with excellent inter-rater reliability, with ICCs of 0.76, 0.83, 0.87 and 0.89 for the safety, assessment, analysis and treatment domains, respectively. The patient and caregiver education and documentation domains showed moderate inter-rater reliability [Table 4].

Table 3: Correlation coefficient for test-retest reliability of the revised Clinical Competency Evaluation Instrument among physiotherapy students in Malaysia

Domain	Correlation, r	Intraclass correlation coefficient (95% CI)	P value
Assessment	0.96	0.98 (0.96–0.98)	<0.01
Analysis	0.94	0.97 (0.96–0.97)	<0.01
Treatment	0.97	0.98 (0.98–0.98)	<0.01
Patient and caregiver education	0.91	0.95 (0.95–0.96)	<0.01
Safety	0.93	0.96 (0.96–0.97)	<0.01
Documentation	0.96	0.97 (0.96–0.97)	<0.01

CI = confidence interval.

Discussion

The results of this pilot study showed that the CCEVI was accurate and reproducible when an assessment of competency among physiotherapy students was carried out, suggesting that it is a valid and reliable evaluation instrument. As seen in this study, clinical competency could not be measured directly; therefore, each item in an assessment instrument's questionnaire should be constructed to represent the domains of competencies intended to be measured. Such items should demonstrate a construct's unidimensionality.^{17,18} The content validity of an assessment instrument is usually based on the subjective judgment of the researcher, supported by a panel of experts.¹⁹ An objective measure to estimate the content validity of an instrument is therefore necessary. By using measures such as the CVI, the experts' responses can be evaluated and the questionnaire items can be rated according to their relevance.¹⁴ In addition, the content validity of an instrument is further established if its items indicate adequacy in representing a range of the attributes intended to be measured.²⁰

As observed in the findings of this study, there was adequate content validity of the overall CCEVI construct (CVI: 0.91). Through factor analysis, the relationship of the items in the instrument, in terms of which items belonged together, were determined and measured.²¹ In total, 35 items with factor loading of ≥ 0.60 were retained in the instrument.

Predetermined performance categories were clearly identified and each of the CCEVI items demonstrated high correlations to clinical competence. The internal

Table 4: Correlation coefficient for inter-rater reliability of the revised Clinical Competency Evaluation Instrument among physiotherapy students in Malaysia

Domain	First assessment ICC (95% CI)	P value	Second assessment ICC (95% CI)	P value
Assessment	0.81 (0.51–0.94)	<0.01	0.84 (0.57–0.95)	<0.01
Analysis	0.91 (0.73–0.97)	<0.01	0.88 (0.66–0.96)	<0.01
Treatment	0.81 (0.50–0.93)	<0.01	0.89 (0.70–0.96)	<0.01
Patient and caregiver education	0.78 (0.33–0.93)	<0.01	0.60 (0.10–0.85)	0.01
Safety	0.59 (0.14–0.84)	<0.01	0.76 (0.40–0.92)	0.01
Documentation	0.97 (0.73–0.97)	<0.01	0.69 (0.29–0.89)	<0.01

ICC = intraclass correlation coefficient; CI = confidence interval.

consistency of the items was evaluated through Cronbach's alpha coefficient. Cronbach's alpha is a reliability index that determines the inter-correlation of items in the instrument measuring the same construct.²² According to general guidelines, for reliability analysis, items with a Cronbach's alpha of >0.70 are considered to have good internal consistency.²³ The findings in the current study demonstrated high internal consistency in all six of the CCEVI domains, which is consistent with the findings of Fitzgerald *et al.*¹³ They reported high internal consistency (Cronbach's alpha: 0.98) for patient management items in the Clinical Internship Evaluation Tool.¹³ A study by Roach *et al.* evaluated the PTCPI and also found that its items showed high internal consistency, as the Cronbach's alpha was 0.99 for the total item scores.¹¹

The reliability of an assessment instrument is related to its consistency in reproducing accurate measurements and its ability to assess an individual's performance with minimum sources of error.^{12,24} One factor that may affect an assessment instrument's reliability is the raters' judgment of the students' performance.^{25,26} In the current study, the focus was on the repeatability and consistency of the scores between assessors when the assessment was conducted by multiple assessors or with the same assessor during repeated assessments. An intraclass correlation of 0.6 to 0.8 was utilised to represent substantial agreement between raters.²⁷ This study demonstrated a high level of agreement between the raters in five domains (ICC: 0.78–0.96) and moderate levels of agreement in the safety domain (ICC: 0.59) in the initial evaluation. However, with subsequent evaluation, the inter-rater

reliability coefficient indicated excellent agreement for four of the domains with a moderate level of agreement in the domains of patient and caregiver education and documentation (ICC: 0.59 and 0.68, respectively).

An earlier study by the American Physiotherapy Association found that the overall ICCs of the Clinical Performance Instruments for inter-rater reliability ranged from 0.50–0.75, which was considered a moderate level of agreement between raters.²⁸ Three other studies reported high levels of agreement between raters (clinical educators and academic faculty tutors) on the assessment of clinical performance.^{12,25,29} Of the three studies, Coote *et al.* and Meldrum *et al.* reported a similar ICC for the overall score (0.84) of their assessment instruments, while Dalton *et al.* reported an overall ICC of 0.92.^{12,25,29} The findings in these studies demonstrated almost perfect agreement between raters.

In contrast, a wide variance of scores between raters might be due to either overly generous or lenient marks given to students, which could lead to a measurement error.² Reubenson *et al.* suggested that performance scores should be awarded immediately after the observation of a student's clinical performance in order to avoid measurement errors and improve reliability.²⁶ Even so, raters' understanding of the performance criteria rating scale, the level of training they received regarding the assessment process and their interpretation of each performance item is likely to differ between individual raters.^{2,24,25} Meldrum *et al.* commented that the assessment of different domains in an assessment instrument may require different assessment skills; thus the competency of raters must be taken into consideration.²⁵

The small sample size in this study may have compromised the reliability of the findings.³⁰ Therefore, future studies on the CCEVI should be conducted with larger sample sizes in order to confirm the results of this study.³¹ It would be beneficial for future research to also incorporate extensive training and detailed guidelines for raters with regards to competency performance criteria and to use a single standard scoring scale to improve agreement and consistency between raters.^{12,25,26,29}

Conclusion

The CCEVI demonstrated high content validity and good to excellent internal consistency across all domains. The stability of the instrument was confirmed through the significant consistency of the scores across the two evaluations. The inter-rater reliability indicated a moderate to excellent correlation coefficient. The results of this study suggested that

the items in the safety and documentation domains required refinement in order to improve the CCEVI's reliability. Further evaluation of the instrument is necessary to strengthen its validity and reliability, as is the replication of this study with a larger sample size. This study suggests that instruments such as the CCEVI can provide an effective tool for physiotherapy academic programmes when assessing the clinical competency of students during their clinical education placement.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

References

1. Panzarella KJ, Manyon AT. Using the integrated standardized patient examination to assess clinical competence in physical therapist students. *J Phys Ther Educ* 2008; 22:24–32.
2. Baig LA, Violato C, Crutcher R. A construct validity study of clinical competence: A multitrait multimethod approach. *J Contin Educ Health Prof* 2010; 30:19–25. doi: 10.1002/chp.20052.
3. Watson R, Stimpson A, Topping A, Porock D. Clinical competence assessment in nursing: A systematic review of the literature. *J Adv Nurs* 2002; 39:421–31. doi: 10.1046/j.1365-2648.2002.02307.x.
4. Butler MP, Cassidy I, Quillinan B, Fahy A, Bradshaw C, Tuohy D, et al. Competency assessment methods - tools and process: A survey of nurse preceptors in Ireland. *Nurse Educ Pract* 2011; 11:298–303. doi: 10.1016/j.nepr.2011.01.006.
5. Chappell K, Koithan M. Validating clinical competence. *J Contin Educ Nurs* 2012; 43:293–4. doi: 10.3928/00220124-20120621-02.
6. Mehrnoosh P, Tahereh A, Sharareh K, Hamid AM. A valid and reliable tool to assess nursing students' clinical performance. *Int J Adv Nurs Studies* 2013; 2:36–9. doi: 10.14419/ijans.v2i1.346.
7. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. *Lancet* 2001; 357:945–9. doi: 10.1016/S0140-6736(00)04221-5.
8. Adams CL, Glavin K, Hutchins K. An evaluation of the internal reliability, construct validity, and predictive validity of the physical therapist clinical performance instrument (PT CPI). *J Phys Ther Educ* 2008; 22:42–50.
9. Ayiesah R, Leonard J, Nor Azlin M, Asfarina M. Reliability and Validity of Physiotherapy Clinical Evaluation Form. Second Teaching and Learning Conference, Penang, Malaysia, 2010.
10. Sangoseni O, Hellman M, Hill C. Development and validation of a questionnaire to assess the effect of online learning on behaviors, attitudes and clinical practices of physical therapists in the United States regarding evidence-based clinical practice. *Int J Allied Health Sci Prac* 2013; 11:1–12.
11. Roach KE, Frost JS, Francis NJ, Giles S, Nordrum JT, Delitto A. Validation of the revised physical therapist clinical performance instrument (PT CPI): Version 2006. *Phys Ther* 2012; 92:416–28. doi: 10.2522/ptj.20110129.
12. Dalton M, Davidson M, Keating JL. The assessment of physiotherapy practice (APP) is a reliable measure of professional competence of physiotherapy students: A reliability study. *J Physiother* 2012; 58:49–56. doi: 10.1016/S1836-9553(12)70072-3.
13. Fitzgerald LM, Delitto A, Irrgang JJ. Validation of the clinical internship evaluation tool. *Phys Ther* 2007; 87:844–60. doi: 10.2522/ptj.20060054.

14. Lynn MR. Determination and quantification of content validity. *Nurs Res* 1986; 35:382–5. doi: 10.1097/0006199-198611000-00017.
15. Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006; 29:489–97. doi: 10.1002/nur.20147.
16. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York: John Wiley & Sons Ltd., 1981, Pp. 38–46.
17. Lewis LK, Stiller K, Hardy F. A clinical assessment tool used for physiotherapy students: Is it reliable? *Physiother Theory Pract* 2008; 24:121–34. doi: 10.1080/09593980701508894.
18. Jette DU, Portney LG. Construct validation of a model for professional behavior in physical therapist students. *Phys Ther* 2003; 83:432–43.
19. Wynd CA, Schmidt B, Schaefer MA. Two quantitative approaches for estimating content validity. *West J Nurs Res* 2003; 25:508–18. doi: 10.1177/0193945903252998.
20. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, et al. A psychometric toolbox for testing validity and reliability. *J Nurs Scholarsh* 2007; 39:155–64. doi: 10.1111/j.1547-5069.2007.00161.x.
21. Munro BH. *Statistical Methods for Health Care Research*. 4th ed. Philadelphia, USA: Lippincott Williams & Wilkins, 2002.
22. DeVellis RF. *Scale Development: Theory and application*. 2nd ed. Thousand Oaks, California, USA: Sage Publications Inc., 2003.
23. Field A. *Discovering Statistics using SPSS (Introducing Statistical Methods Series)*. 3rd ed. Thousand Oaks, California, USA: Sage Publications Inc., 2009. Pp. 675–679.
24. Lagumen NG, Butterwick DJ, Paskevich DM, Fung TS, Donnon TL. The intra-rater reliability of nine content-validated technical skill assessment instruments (TSAI) for athletic taping skills. *Athlet Train Educ J* 2008; 3:91–101.
25. Meldrum D, Lydon AM, Loughnane M, Geary F, Shanley L, Sayers K, et al. Assessment of undergraduate physiotherapist clinical performance: Investigation of educator inter-rater reliability. *Physiother* 2008; 94:212–19. doi: 10.1016/j.physio.2008.03.003.
26. Reubenson A, Schnepf T, Waller R, Edmondson S. Inter-examiner agreement in clinical evaluation. *Clin Teach* 2012; 9:119–22. doi: 10.1111/j.1743-498X.2011.00509.x.
27. Shrout PE, Fleiss JL. Intraclass correlation: Uses in assessing rater reliability. *Psychol Bull* 1979; 86:420–8.
28. Task Force for the Development of Student Clinical Performance Instruments. The development and testing of APTA clinical performance instruments: American Physical Therapy Association. *Phys Ther* 2002; 82:329–53.
29. Coote S, Alpine L, Cassidy C, Loughnane M, McMahon S, Meldrum D, et al. The development and evaluation of a common assessment form for physiotherapy practice education in Ireland. *Physiother Ireland* 2007; 28:6–10.
30. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength and Cond Res* 2005; 19:231–40. doi: 10.1519/15184.1.
31. Perkins DO, Wyatt RJ, Bartko JJ. Penny-wise and pound-foolish: The impact of measurement error on sample size requirements in clinical trials. *Biol Psychiatry* 2000; 47:762–6. doi: 10.1016/S0006-3223(00)00837-4.