Development of the Canadian Physiotherapy Assessment of Clinical Performance: A New Tool to Assess Physiotherapy Students' Performance in Clinical Education

Brenda Mori, BScPT, MSc, PhD;^{*†} Dina Brooks, BScPT, PhD;^{*}
Kathleen E. Norman, BScPT, PhD;[‡] Jodi Herold, BHSc(PT), PhD;[§]
Dorcas E. Beaton, BScOT, PhD^{¶**††}

ABSTRACT

Purpose: To develop the first draft of a Canadian tool to assess physiotherapy (PT) students' performance in clinical education (CE). Phase 1: to gain consensus on the items within the new tool, the number and placement of the comment boxes, and the rating scale; Phase 2: to explore the face and content validity of the draft tool. **Methods:** Phase 1 used the Delphi method; Phase 2 used cognitive interviewing methods with recent graduates and clinical instructors (CIs) and detailed interviews with clinical education and measurement experts. **Results:** Consensus was reached on the first draft of the new tool by round 3 of the Delphi process, which was completed by 21 participants. Interviews were completed with 13 CIs, 6 recent graduates, and 7 experts. Recent graduates and CIs were able to interpret the tool accurately, felt they could apply it to a recent CE experience, and provided suggestions to improve the draft. Experts provided salient advice. **Conclusions:** The first draft of a new tool to assess PT students in CE, the Canadian Physiotherapy Assessment of Clinical Performance (ACP), was developed and will undergo further development and testing, including national consultation with stakeholders. Data from Phase 2 will contribute to developing an online education module for CIs and students.

Key Words: educational measurement; internship and residency; students.

RÉSUMÉ

Objectif : Créer le premier projet d'outil canadien d'évaluation de la formation clinique des étudiants en physiothérapie. Phase 1: dégager un consensus sur les éléments constitutifs du nouvel outil, le nombre et l'emplacement des encadrés servant aux commentaires et l'échelle d'évaluation; Phase 2: explorer l'apparence et la validité du contenu du projet d'outil. **Méthodes :** On a utilisé, au cours de la Phase 1, la méthode Delphi et, au cours de la Phase 2, des entrevues cognitives menées auprès de nouveaux diplômés et de formateurs cliniques, ainsi que des entrevues détaillées menées auprès de spécialistes de l'éducation clinique et de la mesure. **Résultats :** On a dégagé un consensus sur la première version du nouvel outil au cours du cycle 3 du processus Delphi, que 21 participants ont terminé. On a réalisé des entrevues avec 13 formateurs cliniques, 6 nouveaux diplômés et 7 experts. Les nouveaux diplômés et les formateurs cliniques ont pu interpréter l'outil avec précision, qu'ils croyaient pouvoir appliquer à une expérience d'éducation clinique récente, et ont suggéré des améliorations. Les spécialistes ont fourni des conseils marquants. **Conclusions :** La première version d'un nouvel outil d'évaluation de la formation clinique des étudiants en physiothérapie, l'évaluation du rendement clinique (ERC) en physiothérapie canadienne, a été mis au point et fera l'objet d'autres améliorations et essais, y compris d'une consultation nationale des intervenants. Les données tirées de la Phase 2 contribueront à l'élaboration du module d'éducation en ligne.

University programmes in the health professions must assess their students' ability to practice against the standards expected of an entry-level health care professional. In physiotherapy (PT), this includes assessing

students' performance in clinical internships. In Canada, although core competencies for entry-level practice have been developed, no national standardized tool exists to measure PT students' clinical performance behaviours

From the: §Postgraduate Medical Education and ¶Department of Occupational Science & Occupational Therapy, Faculty of Medicine, University of Toronto; *Department of Physical Therapy; †Centre for Faculty Development, Faculty of Medicine, **Musculoskeletal Health and Outcomes Research, University of Toronto at the Li Ka Shing Knowledge Institute of St. Michael's Hospital; ‡†Institute for Work and Health, Toronto; ‡Physical Therapy Program, School of Rehabilitation Therapy, Queen's University, Kingston, Ont.

Correspondence to: Brenda Mori, 160–500 University Ave., Toronto, ON M5G 1V7 brenda.mori@utoronto.ca

Contributors: All authors designed the study; collected, analyzed, and interpreted the data; drafted or critically revised the article; and approved the final draft.

Competing Interests: None declared.

Acknowledgements: Use of the rating scale anchors and descriptors adapted and revised from the PT CPI I Web (Alexandria, VA: American Physical Therapy Association; 2006) is by nonexclusive license from the American Physical Therapy Association.

Physiotherapy Canada 2015; 67(3);281–289; doi:10.3138/ptc.2014-29E

in relation to the level of competence expected for an entry-level physiotherapist. This article describes the development of a standardized measure to capture the Canadian entry-level physiotherapy competencies (response set, instructions, scaling, and scoring) and examines the face and content validity of this new tool, the Canadian Physiotherapy Assessment of Clinical Performance (ACP).

A clinical education (CE) assessment tool is used by students and clinical instructors (CIs) at the midpoint and the end of each internship. The assessment process is instrumental in students' development toward entry-level PT practice.¹ Most Canadian PT programmes currently use the US-based Clinical Performance Instrument, version 1997² (PT-CPI); however, CIs have expressed concerns that the PT-CPI is time consuming, does not always apply to their practice setting, and may have a US bias.^{3–5} For these reasons, a new assessment tool is required for use by Canadian PT students and their CIs to describe how closely the students' behaviours, as observed in the CE setting, approach the level of competence expected for an entry-level physiotherapist.

The Canadian National Association for Clinical Education in Physiotherapy (NACEP)—whose members are the faculty CE leads for entry-to-practice PT degree programmes at Canadian universities—led the initiative to find or develop a new tool for assessing PT students in CE. NACEP developed a list of key attributes required in a new tool: it must be available in English and French, psychometrically sound, feasible, relevant to Canadian PT practice, and competency based; must have the capacity to be completed online; and must be user-friendly, cost-effective, and robust.⁶ In addition, PT academic programmes wanted electronic access to their own data and the ability to share anonymized assessment data across schools from consenting individuals,⁶ which would be possible with data from an online measure. National longitudinal data would support national research initiatives, enabling NACEP to make data-informed decisions about student performance and to assess the measurement properties of the assessment form.

A subcommittee of NACEP investigated two assessment tools—the PT-CPI: Version 2006⁷ and the Assessment of Physiotherapy Practice,^{8,9} developed in Australia—for their potential use with and applicability to Canadian PT students. For several reasons, the subcommittee concluded that these tools did not meet the key attributes developed by NACEP: they were not deemed entirely relevant to Canadian PT practice, were not user-friendly, and were not amenable to pooling assessment data. Therefore, the academic leaders (NACEP and the chairs/directors of the PT programmes) agreed to develop a new Canadian tool based on the Canadian Essential Competency Profile for Physiotherapists (ECP),^{10,11} which

describes the essential competencies (i.e., the knowledge, skills and attitudes) required by physiotherapists in Canada at the beginning of and throughout their careers. It also provides guidance for physiotherapists to build on their competencies over time. The ECP reflects the diversity of physiotherapy practice in Canada and helps support evolution of the profession in relation to the changing nature of practice environments and advances in evidence-informed practice.^{10(p.4)}

Opinions differ on using competency documents as the basis for developing measurement tools to assess competency. Lurie and colleagues,¹² after completing a systematic review to evaluate the extent to which the six general competencies of the Accreditation Council for Graduate Medical Education in the United States can be independently measured in a valid and reliable way, concluded that the literature does not support any one method to assess the general competencies as independent constructs; competencies should be used to guide and coordinate specific evaluation efforts rather than to measure the competencies directly. In contrast, Green and Holmboe¹³ have argued that the chief challenges in assessing competencies can be mitigated through adequate training of those who use such tools to assess learners in CE, citing evidence of some successes in using tools that measure general competencies.

Typical development stages for a measure of this type include item selection, item reduction, development, pretesting, and testing.¹⁴ In this case, because the items were generated from the ECP, we will report on the first two phases of the process: item reduction and development. The objective for Phase 1 was to use the Delphi approach to gain consensus on the rating scale and the ECP items to be used in the tool, as well as on the number and placement of comment boxes. The objective of Phase 2 was to explore the face and content validity of the draft tool in consultation with a variety of stakeholders, including recent graduates, clinicians, and education and measurement experts, through individual interviews, as well as to gain buy-in from key stakeholders. While the American Educational Research Association does not support face validity as a type of validity,^{15,16} other measurement experts do support the concept of face validity.^{17,18} In this study, gathering evidence on face validity also served as a means to engage stakeholders in the development of a national tool and to demonstrate that we value their input.

METHODS

Both study phases were approved by the Health Sciences Research Ethics Board at the University of Toronto and conformed to the Human and Animal Rights requirements of the February 2006 International Committee of Medical Journal Editors' Uniform Requirements for Manuscripts Submitted to Biomedical Journals. All participants provided informed consent.

Phase 1

Participants

A purposive sample¹⁹ was used for our three-stage online national Delphi consensus study, which took place between August and October 2012. Members of NACEP and of the Canadian Council of Physiotherapy University Programs (Academic Council)—which comprises the directors or chairs of all academic PT programmes in Canadian universities—were invited to complete a series of surveys. Each university might have more than one member represented on these committees; five universities had two representatives on the Academic Council, and eight universities had more than one representative in NACEP. At the time of our Delphi study, all 47 members of both groups (20 members of the Academic Council and 27 members of NACEP) were invited to participate as the Expert Consultant Panel. The surveys were available in English only; we did not anticipate that this would limit participation in the Delphi process, as all national meetings are conducted in English, but participants were encouraged to respond to open-ended questions in French if they preferred.

Process

Phase 1 used the Delphi process,²⁰ an accepted method for consensus building based on the rationale that pooled intelligence enhances individual judgment and provides the collective opinion of a group of experts.²¹ In our Delphi process, each round consisted of an electronic survey distributed to the Expert Consultant Panel; in each round, participants were given the option of printing their completed responses, and after round 1, aggregate results of the previous round were shared with participants during each new round. Participants thus had opportunities to reflect on their own or the group's previous responses before responding in the next round.

The questionnaires

Since the items included in the tool were defined by the ECP, our study focused on using the Delphi process to refine the number of items to be assessed, the number and placement of comment boxes, and the rating scale. In rounds 1 and 2, the Delphi survey had four components: Part A—Consideration of items that need a rating scale; Part B—Consideration of items that need a comment box; Part C—Consideration of the rating scale descriptors; and Part D—Demographic information. In round 1, participants were asked to indicate on a 5-point Likert-type scale (1 = strongly disagree, 5 = strongly agree) their agreement with the statement “I believe the [Role or Key Competency] can be assessed with one rating scale.” Those whose response was ≤ 3 (neutral, disagree, or strongly disagree) were asked to answer the same question for each Key Competency or Enabling Competency within that role. In round 2, participants were asked their opinion on those roles and competencies for

which we did not achieve consensus in round 1, using a dichotomous scale (agree/disagree). Round 3 proposed a draft assessment tool, followed by the statement, “The proposed assessment tool is acceptable to move on to the next stage of development”; again, response options were “I agree” or “I disagree.”

Analysis

For the purpose of our analysis, consensus was considered to be achieved if $\geq 80\%$ of participants expressed agreement,²² identified by ratings of 4 or 5 (agree or strongly agree) on the Likert scale in round 1 or “I agree” in rounds 2 and 3. To maintain confidentiality, we analyzed data from the item reduction competency component and the rating scale component of the Delphi separately from the demographic data.

Phase 2

Participants

Phase 2 used one-on-one interviewing with two groups of participants. The first group consisted of experts in measurement and/or CE identified by the authors either as having recently completed similar national projects in physiotherapy or other health professions, or as having expertise in measurement or competence assessment. Experts were interviewed about their area of expertise, their experiences, and their advice on developing a new measure.

Recent graduates (RGs) and CIs made up the second group of participants. We used cognitive interviewing methods, including think-aloud and verbal probing;²³ interview questions for this group focused on comprehension of the tool and on the process respondents used to formulate their answers to a question (what the literature terms *judgment*).²⁴ For this group, we sought a representative sample of CIs across the continuum of care and a variety of practice areas; RGs were invited to participate to provide the student-user perspective.

Process and analysis

Based on the work of Willis (1999),²⁴ we anticipated that approximately 12–15 interviews each from the CI and RG cohorts would meet the goals of the cognitive interview and provide the desired mix of participants to ensure representation of CIs' opinions from a variety of practice areas and across the continuum of care. Potential participants were contacted by email by the principal investigator (BM) and invited to participate; those who agreed received the draft tool by email so that they could review it before the interview.

Box 1 shows the interview guide. The interviewer took notes during the interview; these were then returned to the participant, who could review, modify, or edit them to ensure that his or her opinions were accurately recorded. The approved interview notes were reviewed by the principal investigator for common topics and issues.

Box 1 Interview Guide**Clinical Education and/or Measurement Experts: (questions to be modified based on the individual's expertise)**

- The rating scale is a modification from other rating scales used in assessment. Do you think it would allow a clinician to accurately assess a student?
- Based on your experience, do you anticipate any challenges with using this scale?
- Do you think this tool would be able to capture a student who was performing poorly/could fail an internship?
- Do you think this tool would be able to capture a student who was performing really well?
- Do you anticipate any challenges with this measure?
- Do you foresee any unintended consequences/potential pitfalls with the proposed tool?
- Do you have any allergic reactions to this draft tool?
- Do you have any advice regarding continuing to develop this tool?

Clinical Instructors and Recent Graduates (potential questions):

- Area of practice (CIs only)

Questions about the rating scale:

1. A Beginner student is defined in the tool. Can you rephrase your understanding of a beginner student in your own words?
2. An Advanced Beginner student is defined in the tool. Can you rephrase your understanding of an advanced beginner student in your own words?
3. An intermediate student is defined in the tool. Can you rephrase your understanding of an intermediate student in your own words?
4. An advanced intermediate student is defined in the tool. Can you rephrase your understanding of an advanced intermediate student in your own words?
5. An entry level student is defined in the tool. Can you rephrase your understanding of an advanced intermediate student in your own words?
6. What do you think about the 75% caseload description for the entry level student?
7. A student who is performing with distinction is defined in the tool. Can you rephrase your understanding of a student performing with distinction in your own words?
8. Do you think “performance with distinction” is attainable?
9. Would you prefer a visual analogue scale or an ordinal scale or something in between (Examples provided)
10. Where would you have rated your last student on the tool for Expert—Assessment? How did you come to that rating?
11. Can you describe a simple patient on your service?
12. Can you describe a complex patient on your service?

Questions about the items that comprise the tool:

13. Can you explain the Role of Expert—Focus on Assessment in your own words
14. Can you explain the Role of Expert—Focus on Analysis in your own words
15. Can you explain the Role of Expert—Focus on Intervention in your own words
16. What would an entry level student look like on Communicator 2.1 to you?
17. How would you describe a beginner level student on Communicator 2.2?
18. How would you describe an intermediate student on Communicator 2.3?
19. Do you think we would need 3 rating scales for communicator role?
20. What does 3.1 mean to you?
21. What does 3.2 mean to you?
22. Should 3.1 and 3.2 be separate items?
23. What does 4.1 means to you?
24. What does 4.3 mean to you?
25. Can you describe an entry level student for 5.0—advocate?
26. Can you describe a beginner student for the Scholarly Practitioner role?
27. What does 7.1 mean to you?
28. What does 7.2 mean to you?
29. What does 7.3 mean to you?

General impressions of the ACP:

30. Would this tool allow you to capture a student doing poorly?
31. If the student was failing—would it provide adequate feedback to them that they were failing?
32. Would this tool allow you to capture a student doing very well?
33. How would you suggest improving the tool?
34. How long do you think it would take to complete it?
35. Would this be a realistic ask?
36. Would you like to complete it on paper or electronically?

Our analysis was guided by the approach in Willis' guide:²⁴ CI and RG interview notes were considered together for consistency in interpretation of the wording of the key competencies and rating scale anchors, as well as relevant examples that could be used for the ACP education module, while interview notes from the experts were reviewed for suggestions and advice that would improve the tool and/or future studies examining the ACP. The findings were used to modify the draft tool as well as to provide information for the education module for the study piloting the new tool.

RESULTS

Phase 1

Table 1 describes the participants in each round and summarizes the proportion of agreement on whether each key competency should have its own rating scale or comment box. In Round 3, 81% agreed with the proposed draft, representing consensus on the first draft of the tool.

Phase 2

The 26 Phase 2 participants are described in Table 2. Results from the CI and RG interviews are summarized below, followed by the summary of the Expert interviews.

Rating scale responses

In general, CI and RG participants were able to accurately describe each level of the rating scale in their own words (questions 1–5). One component of the “Entry-level performance” description was, “the student is capable of maintaining at minimum 75% of a full-time PT's caseload in a cost-effective manner”; participants generally considered maintaining 75% of a caseload achievable, but CIs working in private practice or in high-acuity areas felt that it might be difficult:

My patients have chronic conditions and I'm always there... The way I manage the caseload is that I get them [the student(s)] to act with me as a buffer—they have their independence and they do some things, and I do things they don't know how to do. (CI#10)

While many participants felt that they would not often be able to use the “with distinction” descriptor, some participants felt that “with distinction” would be achievable for some students who were able to manage their caseload efficiently and effectively.

Participants generally appreciated that the draft ACP included a rating scale with defined anchors, but they had varying ideas of what the scale should look like (a line with anchors, a checkbox for each anchor, or a checkbox for each anchor plus an additional checkbox between anchors).

Item responses

Participants seemed able to accurately interpret the items within the tool. However, several participants did

question why key competency 1.5 (“develops and recommends an intervention strategy”) was categorized in the section “Expert—Focus on Intervention.”

It [1.5] fits more with the analysis section. I see 1.5 as developing the plan of care—I see—intervention as doing it—I'm going to walk the patient then do the stairs. 1.5 requires reasoning—so they might not be able to reason out that the patient has stairs at home and we need to do it—but the student could be great at guarding the patient on the stairs. These are two very different things and having 1.5 in analysis allows me to break it up like that. (CI#7)

There was some ambiguity regarding the Advocate role, especially when participants were asked about applying that role to their last internship (RGs) or student supervised (CIs):

I think it's really hard to be an advocate in a 5-week [internship]—I mean you are always an advocate to an extent and definitely an advocate for your patient but to be an advocate for the profession is more difficult so it would be difficult to rate myself based on the client or the profession. (RG#2)

Participants demonstrated an ability to apply the rating scale in the measure to the last student they supervised (CIs) or to their own performance in their last clinical internship (RGs). When a recent graduate was asked to rate her performance on her most recent internship (final internship) on the Expert—Focus on Intervention item, she responded,

Entry-level. I felt I had all the tools—that I learned throughout the internship—by the end of the [internship] I was able to successfully deliver the treatments to the patients. (RG#5)

General impressions of the tool

Participants commented that the draft ACP was clearer, less redundant, and more applicable to their practice than the PT-CPI, and that the categories contributed to a more obvious focus:

I wish this one was used when I went through—much clearer and ... [it] seems like a better assessment of a student as a PT than the current tool. And it's nice that it is significantly shorter. I also like that there are not necessarily comment boxes for every scale because that was something ... I ended up reaching for comments and [they were] not as meaningful and it makes more sense to have 1 or 2 comment boxes for the role. (RG#1)

They felt that the ACP would help them identify students who are performing very well and also to identify those who are performing poorly. Two participants suggested adding a box to indicate the need to contact the Director/Academic Coordinator of Clinical Education at the University programme if there were significant concerns about a student's performance; this method of

Table 1 Phase 1 Delphi Results

		Round 1 Results		Round 2 Results		Round 3 Results & Proposed		
Participants	No. of responses	26 complete; 3 incomplete*		23 complete		21		
	NACEP	18		17		16		
	Academic Chairs	8		6		5		
	Average years on faculty	9.2		8.4		9.1		
Role		KC	% (no.) SA + A Need its own RS? Need its own CB?		% (no.) I agree Need its own RS? Need its own CB?		Need its own RS?	Need its own CB?
Essential Competency Profile Section (RS and CB)	1.0 Expert	1.1	—	—	✓1 RS Expert-Assessment, 83% (23)	✓1 CB, 82% (22)	✓Round 2	✓Round 2
		1.2	—	—				
		1.3	—	—	✓1 RS Expert-Analysis, 83% (23)	✓1 CB, 82% (22)	✓Round 2	✓Round 2
		1.4	—	—				
		1.5	—	—	✓1 RS Expert-Intervention, 83% (23)	✓1 CB, 82% (22)	✓Round 2	✓Round 2
		1.6	—	—				
		1.7	—	—				
		1.8	—	—				
	2.0 Communicator	2.1	—	✓1 CB, 85% (26)	—	✓Round 1	RS proposed	✓Round 1
		2.2	—		—		RS proposed	
		2.3	—		—		RS proposed	
	3.0 Collaborator	3.1	—	✓1 CB, 85% (26)	—	✓Round 1	RS proposed	✓Round 1
		3.2	—		—		RS proposed	
	4.0 Manager	4.1	—	✓1 CB, 85% (26)	—	✓Round 1	RS proposed	✓Round 1
		4.2	—		—		RS proposed	
		4.3	—		—		RS proposed	
	5.0 Advocate	5.1	✓1 RS, 89% (28)	✓1 CB, 96% (26)	✓Round 1	✓Round 1	✓Round 1	✓Round 1
	6.0 Scholarly Practitioner	6.1	—	✓1 CB, 85% (26)	✓Consensus on 1 RS, 83% (23)	✓Round 1	✓Round 2	✓Round 1
		6.2	—					
		6.3	—					
	7.0 Professional	7.1	—	✓1 CB, 85% (26)	—	✓Round 1	RS proposed	✓Round 1
		7.2	—		—		RS proposed	
		7.3	—		—		RS proposed	
							Total: 16 RS	Total: 9 CB
Rating Scale Section	Beginner	65% (26) with suggestions to reword			✓Consensus on descriptors 96% (22)		✓Round 2	
	Advanced Beginner	65% (26) with suggestions to reword			✓Consensus on descriptors 87% (22)		✓Round 2	
	Intermediate	✓Consensus on descriptors 92% (25)			✓Round 1		✓Round 1	
	Advanced Intermediate	✓Consensus on descriptors 91% (23)			✓Round 1		✓Round 1	
	Entry-Level	✓Consensus on descriptors 81% (26)			✓Consensus on slight modification 91% (22)		✓Round 2	
	Beyond Entry Level	65% (26) with suggestions to reword			✓Consensus on descriptor as a “with distinction box” 87%		✓Round 2	
	Scale as a whole	62%; would prefer to have “With Distinction” box instead of “Beyond Entry-Level Performance”			Agreement on rating scale: 76% (21)		RESULT: 81% (17) agreed to proposed tool	

SA = Strongly Agree; A = Agree; RS = Rating Scale; CB = Comment Box; KC = Key Competency; — = no consensus; ✓ = consensus achieved.

*Since the demographic questions constituted the final component of the questionnaire, the demographics from the 3 incomplete responses in Round 1 were not captured.

Table 2 Phase 2 Participants

Participants	Personal Features	Current Practice		
		Practice Settings	Area of Practice	Interview Mode
CIs (<i>n</i> = 13)	10 F, 3 M	Acute care (5) Inpatient rehab (2)	Oncology (3) Neurology (3)	In person (8); by phone (5)
	Mean 10.5 (SD 7.6) y in practice	Outpatient hospital rehabilitation (2) Private practice (3) Community care (1)	Cardiorespiratory (3) Musculoskeletal (4) Paediatrics (2) General PT (1)	
RGs (<i>n</i> = 6)	6 F	N/A	N/A	By phone (6)
Experts (<i>n</i> = 7)	6 F, 1 M	Physiotherapist (5)		In person (4)
		Occupational therapist (1) Social worker (1)		By phone (2) Internet call (1)

CI = clinical instructor; F = female; M = male; RG = recent graduate; PT = physiotherapy; N/A = not applicable.

identifying cause for concern, they felt, would be a clearer signal to students who might not be performing well. Two participants did not like use of the word *Expert*, as they felt that a student should not be assessed in relation to an expert.

Expert interviews

The measurement experts we interviewed consistently advised us to consider using a rating scale that would make the tool informative to users and robust enough to allow for statistical analyses, as well as providing the same data whether the tool was completed online or on paper. They also suggested collecting data longitudinally over several cohorts to allow for norm referencing, which would help users interpret the scores on the rating scale and describe typical performance of students at specific stages. Finally, they suggested that we consider the number of items assessed with a rating scale and how this might affect the overall weighting and scoring of the tool.

DISCUSSION

This article describes the process used to develop a new Canadian tool to assess PT students' performance in CE, the ACP, and explores the face and content validity of the ACP through consultation with a variety of stakeholders, including recent graduates, clinicians, and experts in education and measurement.

As noted above, opinions differ as to whether competency documents should be used to develop measurement tools aimed at assessing competence.^{12,13} Our goal was to develop a measure that would produce a comprehensive picture of the PT student's performance in the CE setting, rather than assessing each competency independently of the rest. Observing students in practice is an essential component of accurately assessing students' competence,²⁵ and Canadian PT education has a strong culture of observing students in CE. Therefore, basing the new tool on the ECP has the potential to provide accurate information on students' performance

relative to the expectations of a Canadian physiotherapist. Moreover, using the ECP as the framework for the ACP would make explicit and transparent to PT students and educators that the key competencies to be developed by PT students are the same as those expected of all physiotherapists practising in Canadian contexts.

During round 3 of the Delphi process in Phase 1, we proposed a draft tool with three rating scales for the ECP role of Expert, one rating scale for the Scholarly Practitioner role, and one rating scale for the remaining key competencies (3 for Communicator, 2 for Collaborator, 4 for Manager, 3 for Professional), for a total of 16 rating scales. Consultation with stakeholders indicated that item 1.5 did not seem to fit in the section "Expert—Focus on Intervention." Reading the enabling competencies for item 1.5 made it clearer to us that while the topic is intervention, the enabling competencies describe the clinical reasoning necessary to write goals and develop a treatment plan (an analytical process), which seemed to us to fit better in the "Expert—Focus on Analysis" section. This prompted further consultation with NACEP and a review of the literature. Given that Higgs and Jones describe clinical reasoning as a process of knowledge and cognition bridged with metacognition or reflective self-awareness,²⁶ and in light of the enabling competencies for item 1.5, we moved this item, with NACEP's support, to the "Expert—Focus on Analysis" section.

Rating scales are an important component in measurement tools that assess learners' clinical performance, and they can have a significant impact on the tool's psychometric qualities.¹⁸ Rating scales can take several forms. The *visual analogue scale* (VAS) is a continuous line with defined end points. *Adjectival scales* are unipolar scales and use descriptors along a continuum (e.g., reported health: poor/fair/good/very good/excellent; student performance: unsatisfactory/satisfactory/excellent/superior).¹⁸ *Likert-type scales* are similar to adjectival scales, with the exception that they are bipolar (e.g.,

strongly disagree, neutral, strongly agree). VASs intuitively seem to provide greater measurement accuracy, since the point on the line can be measured in millimetres, but a study of pain rating scales showed a strong correlation between the VAS and ordinal scales and found that an 11-point *numeric rating scale* can perform better than either a 4-point descriptive scale or a VAS.²⁷ Previous studies have suggested that the most appropriate number of categories on a rating scale is seven, plus or minus two categories,^{28,29} based on typical short-term memory capacity and ability to process information.

In Phases 1 and 2 of our development process, the proposed scale was an adjectival scale with five defined anchors (Beginner, Advanced Beginner, Intermediate, Advanced Intermediate, Entry-Level) on a line and a checkbox for “with distinction” to the right of Entry-Level. While we researched several anchor descriptors and rating scale formats, a Canadian survey of physiotherapists completed in 2012 indicated a preference for this graphic representation of the rating scale.⁴ The preferred rating scale was also accepted by the Expert Consultant Panel. We modified the anchor descriptions from the PT CPI: Version 2006,⁷ with permission from the American Physical Therapy Association (see Acknowledgements). During interviews with CIs and RGs, we showed participants three different models of the same scale: a VAS, a categorical scale with 6 anchors and 6 boxes, and a categorical scale with 6 anchors and 10 boxes (i.e., a box between each pair of anchors except the last). While some participants had a strong preference for one format over another, responses were varied. Thus, these two developmental phases were successful in defining the anchors for the ACP, and the participants demonstrated an ability to interpret and apply the anchors to rate their most recent student (or themselves) in a seemingly accurate manner, but the variety of opinions on the appearance of the scale indicates a need for further investigation. This investigation will be a core component of the next phase of development: Phase 3 will specifically ask physiotherapists their opinions on the appearance of the rating scale, the grouping of the proposed items within the roles, and their overall impression of the measure.

Our development process produced the first draft the ACP, a new Canadian assessment tool based on the ECP that consisted of 16 items assessed by a defined rating scale, 9 comment boxes, and an adjectival rating scale represented on a VAS with the option to select “with distinction” as a defined anchor. Item 1.5, “Develops and recommends an intervention strategy,” was moved to the Expert—Focus on Analysis section. The ACP also allows the assessor to indicate significant concerns about a student’s performance in each role.

Data from this study will also contribute to the development of an online education module for the new

tool, which will include clarification regarding expected caseload and descriptive examples collected from the interviews for the items and scale anchors. The online module will clarify that “Beginner” performance is acceptable performance for a junior student and will guide assessors to consider the enabling competencies as a description of the key competency rather than an exhaustive list that must be achieved.

This is the first body of research focused on developing a Canadian measure to assess PT students in CE since the work of Loomis in the mid-1980s.^{30,31} Further research to develop and refine the ACP will help move Canadian PT CE practice toward our goal of accurately and meaningfully assessing PT students in practice.

LIMITATIONS

Our study has several limitations. First, we received input in all rounds of Phase 1 from only 21 of 47 academic leaders; 16 represented NACEP (maximum participation would have been 27), the core group within the Expert Consultant Panel with the most relevant experience in assessing students in CE. In Phase 2, we achieved representation across a variety of practice areas from the 13 participating CIs, but all participants were working in an urban setting in the same two cities in southern Ontario. Additional areas of practice, such as rural practice or general medicine, could have added additional perspectives on the proposed draft. Second, our target was 12–15 interviews with recent graduates, but only 6 potential participants were interested and consented to participate in the study, and all were from the same academic institution. The next phase of testing will ensure broader representation.

CONCLUSIONS

The first draft of the ACP, a Canadian tool to assess PT students’ performance in CE, was developed in consultation with experts, clinicians, and students, and there is evidence to support the face and content validity of the ACP. The next phase development will involve broad consultation with physiotherapists across Canada to obtain feedback on the proposed draft; in particular, we will seek input on the grouping of items, the appearance of the rating scale, the name of the tool, and overall impressions. Following broad consultation, the resulting draft will be pilot-tested across Canada.

KEY MESSAGES

What is already known on this topic

Assessing students in clinical education is an important component of their development toward becoming entry-level physiotherapy professionals. A new tool that is relevant to physiotherapy practice in Canada would be well received by clinical instructors.

What this study adds

This article reports on the first two stages in developing the Canadian Physiotherapy Assessment of Clinical Performance (ACP), a new Canadian tool to assess physiotherapy students in clinical education, based on the Essential Competency Profile for Physiotherapists in Canada. Consensus was achieved on the items of the ACP, the number of comment boxes, and the rating scale to be used for the items. Interviews with stakeholders indicate that both clinical instructors and students were able to interpret the tool accurately and to apply it to their last clinical education experience.

REFERENCES

- Keating J, Dalton M, Davidson M. Assessment in clinical education. In: Delany C, Molloy E, editors. *Clinical education in the health professions*. Sydney: Churchill Livingstone; 2009. p. 147–72.
- American Physical Therapy Association (APTA). *Physical Therapist Clinical Performance Instrument*. Alexandria (VA): The Association; 1997.
- Anderson C, Cosgrove M, Lees D, et al. What clinical instructors want: perspectives on a new assessment tool for students in the clinical environment. *Physiother Can*. 2014;66(3):322–8. <http://dx.doi.org/10.3138/ptc.2013-27>. Medline:25125788
- Hall MD. Identifying the factors contributing to Canadian physiotherapists' decisions to supervise physiotherapy students: results from a national survey [dissertation]. Edmonton: University of Alberta; 2013.
- Creaser GA. An exploration of clinical instructors' experiences and perceptions of the physical therapy clinical performance instrument [dissertation]. Halifax (NS): Mount Saint Vincent University; 2006.
- National Association for Clinical Education in Physiotherapy. Annual general meeting minutes. Ottawa; 2008.
- Roach KE, Frost JS, Francis NJ, et al. Validation of the Revised Physical Therapist Clinical Performance Instrument (PT CPI): version 2006. *Phys Ther*. 2012;92(3):416–28. <http://dx.doi.org/10.2522/ptj.20110129>. Medline:22135710
- Dalton M, Davidson M, Keating J. The Assessment of Physiotherapy Practice (APP) is a valid measure of professional competence of physiotherapy students: a cross-sectional study with Rasch analysis. *J Physiother*. 2011;57(4):239–46. [http://dx.doi.org/10.1016/S1836-9553\(11\)70054-6](http://dx.doi.org/10.1016/S1836-9553(11)70054-6). Medline:22093122
- 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(1):49–56. [http://dx.doi.org/10.1016/S1836-9553\(12\)70072-3](http://dx.doi.org/10.1016/S1836-9553(12)70072-3). Medline:22341382
- National Physiotherapy Advisory Group. Essential competency profile for physiotherapists in Canada. Toronto: Canadian Physiotherapy Association; 2009.
- National Association for Clinical Education in Physiotherapy. Annual general meeting minutes. Calgary; 2009.
- Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med*. 2009;84(3):301–9. <http://dx.doi.org/10.1097/ACM.0b013e3181971f08>. Medline:19240434
- Green ML, Holmboe E. Perspective: the ACGME toolbox: half empty or half full? *Acad Med*. 2010;85(5):787–90. <http://dx.doi.org/10.1097/ACM.0b013e3181d737a6>. Medline:20520026
- Guyatt GH, Bombardier C, Tugwell PX. Measuring disease-specific quality of life in clinical trials. *CMAJ*. 1986;134(8):889–95. Medline:3955482
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington (DC): American Educational Research Association; 1999.
- Martella RC, Nelson JR, Morgan RL, et al. *Understanding and interpreting educational research*. New York: Guilford; 2013.
- Feinstein AR. *Clinimetrics*. New Haven (CT): Yale University Press; 1987.
- Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008. <http://dx.doi.org/10.1093/acprof:oso/9780199231881.001.0001>.
- Domholdt E. *Physical therapy research: principles and applications*. Philadelphia: W.B. Saunders; 1993.
- Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. *Practical Assessment, Research & Evaluation*. 2012;12(10). Available from: <http://pareonline.net/pdf/v12n10.pdf>.
- Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. *J Inf Technol Educ*. 2007;6:1–21.
- Ulshak FL. *Human resource development: the theory and practice of need assessment*. Reston (VA): Reston Publishing; 1983.
- Jobe JB. Cognitive psychology and self-reports: models and methods. *Qual Life Res*. 2003;12(3):219–27. <http://dx.doi.org/10.1023/A:1023279029852>. Medline:12769134
- Willis GB. Cognitive interviewing: a “how to” guide. Presented at the 1999 Meeting of the American Statistical Association. Available from: <http://www.hkr.se/pagefiles/35002/gordonwillis.pdf>.
- van der Vleuten CP, Schuwirth LW, Scheele F, et al. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol*. 2010;24(6):703–19. <http://dx.doi.org/10.1016/j.bpobgyn.2010.04.001>. Medline:20510653
- Higgs J, Jones M. Clinical reasoning in the health professions. In: Higgs JM, editor. *Clinical reasoning in the health professions*. 2nd ed. London: Butterworth-Heinemann; 2000. p. 3–14.
- Downie WW, Leatham PA, Rhind VM, et al. Studies with pain rating scales. *Ann Rheum Dis*. 1978;37(4):378–81. <http://dx.doi.org/10.1136/ard.37.4.378>. Medline:686873
- Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health*. 1990;13(4):227–36. <http://dx.doi.org/10.1002/nur.4770130405>. Medline:2197679
- Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev*. 1956;63(2):81–97. <http://dx.doi.org/10.1037/h0043158>. Medline:13310704
- Loomis J. Evaluating clinical competence of physical therapy students. Part 1: the development of an instrument. *Physiother Can*. 1985;37(2):83–9. Medline:10270852
- Loomis J. Evaluating clinical competence of physical therapy students. Part 2: assessing the reliability, validity and usability of a new instrument. *Physiother Can*. 1985;37(2):91–8. Medline:10270853