Canadian Physiotherapy Assessment of Clinical Performance: Face and Content Validity

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

ABSTRACT

Purpose: To investigate face and content validity of a draft measure to be used across Canada to assess physiotherapy students' performance in clinical education, through broad consultation with physiotherapy clinical instructors (CIs) across Canada. **Methods:** An online survey was used to collect input on the draft measure. In addition to demographics, the questionnaire included questions on the preferred rating scale, the items within the measure that should have their own rating scale, and general impressions. **Results:** Among the 259 CIs who completed the survey, a discrete rating scale with six anchors and 10 boxes or a continuous-line rating scale with six anchors was preferred. Respondents favoured using one rating scale for each key competency in the Expert role but considered a single rating scale sufficient for assessing the Scholarly Practitioner role. CIs agreed that the proposed measure would allow them to assess a student who was performing poorly or very well. The name *Canadian Physiotherapy Assessment of Clinical Performance* (ACP) received the most votes in the questionnaire. **Conclusions:** CIs' collective preferences on the design, organization, and naming of the measure they will use in evaluating students are reflected in the second draft of the ACP.

Key Words: clinical competence; educational assessment; internship and residency; students.

RÉSUMÉ

Objectif : Examiner l'apparence et la validité du contenu d'un projet de mesure qui servira dans tout le Canada pour évaluer le rendement en formation clinique des étudiants en physiothérapie, grâce à une vaste consultation auprès des enseignants cliniques à l'échelle du Canada. **Méthodes :** Un sondage en ligne a permis de recueillir des commentaires sur le projet de mesure. Outre les données démographiques, le questionnaire comprenait des questions sur l'échelle de cotation préférée, les éléments de la mesure qui devraient avoir leur propre échelle de cotation et les impressions générales. **Résultats :** Les 259 enseignants cliniques qui ont répondu au sondage préféraient généralement une échelle de cotation discrète comportant 6 ancres et 10 cases ou une échelle de cotation doublée de 6 ancres. Les répondants préféraient utiliser une échelle de cotation pour chaque compétence clé du rôle d'expert, mais considéraient qu'une seule échelle de cotation était suffisante pour évaluer le rôle d'érudit. Les enseignants cliniques étaient d'accord avec le fait que la mesure proposée leur permettrait d'évaluer les étudiants dont le rendement était mauvais ou très bon. L'appellation « évaluation du rendement clinique (ERC) de la physiothérapie au Canada » a obtenu le plus de votes dans le questionnaire. **Conclusions :** Les préférences collectives des enseignants cliniques en matière de conception, d'organisation, et d'appellation de la mesure qu'ils utiliseront pour évaluer les étudiants sont prises en compte dans la deuxième version de l'ERC.

Assessment during clinical education (CE) is a critical component of physiotherapy (PT) students' journey in their development as competent entry-level physiotherapists. Assessment forms are commonly used by both students and clinical instructors (CIs) twice during each internship, at middle and final points, to assess the student's performance behaviours. These forms provide for-

mative and summative feedback to students, provide information regarding the student's performance, allow students to develop their self-assessment skills, and allow faculty to decide whether a student should pass a CE course. A new measure that will be used across Canada to assess PT students' clinical performance behaviours on the basis of longitudinal observations in the CE setting

From the: ^{*}Department of Physical Therapy; Centre for Faculty Development; [§]Postgraduate Medical Education and [¶]Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of Toronto; ^{**}Musculoskeletal Health and Outcomes Research, Li Ka Shing Knowledge Institute, 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, Department of Physical Therapy, Faculty of Medicine, University of Toronto, 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.

Physiotherapy Canada 2016; 68(1);64–72; doi:10.3138/ptc.2015-35E

is currently being developed, based on the Essential Competency Profile for Physiotherapists in Canada (ECP),¹ a national document that

describes the essential competencies, (i.e., the knowledge, skills and attitudes) required by physiotherapists in Canada at the beginning of and throughout their career. It also provides guidance for physiotherapists to build on their competencies over time. The Profile reflects the diversity of physiotherapy practice and helps support evolution of the profession in relation to the changing nature of practice environments and advances in evidence-informed practice.^{1(p.4)}

In this article, we describe the third phase of development of the Canadian Physiotherapy Assessment of Clinical Performance (ACP). This process began with the development of a draft of the measure in Phases 1 and 2 and initial exploration of face and content validity with a focused sample. In Phase 1, an expert consultant panel participated in a study that used the Delphi² approach to gain consensus on the rating scale, the items that would compose the measure, and the number and placement of comment boxes. However, because the expert consultant panel was so intimately engaged with the concepts of the ECP and with clinical education and assessment, we also needed to establish face and content validity with future users of the measure.

In Phase 2, therefore, we interviewed a variety of stakeholders, including CIs, recent graduates, and experts in clinical education and measurement. The result was a first draft of the assessment measure that included a total of 16 rating scales assessing seven roles and nine comment boxes, as reported elsewhere.³ Table 1 summarizes the roles and items in the first draft.

The new measure, based on the ECP, is a descriptive measure used by PT students and their CIs to describe the students' behaviours as observed in the CE setting relative to what might be expected of an entry-level physiotherapist at middle and final points of the internship (e.g., during the 3rd week and the final week of a 5-wk internship). After Phases 1 and 2, the draft measure included an entry-level physiotherapist's seven roles and each role's key competencies: Expert—Focus on Assessment (one item), Expert—Focus on Analysis (one item), Expert—Focus on Intervention (one item), Communicator (three items), Collaborator (two items), Manager (three items), Advocate (one item), Scholarly Practitioner (1 item), and Professional (3 items). The draft measure also included a rating scale with six defined anchors.

There are two popular paradigms to approach validity: a health measurement (psychometric or clinimetric approach)⁴ and an educationalist (or modern) approach.⁵ The health measurement approach typically emphasizes amassing an adequate volume of data to validate an instrument for a specific application.⁶ In contrast, the educationalist approach, often linked to Messick's⁷ unified theory of validity, focuses on the use of data to support the validity of inferences about the measure. This

article principally takes a health measurement approach to validity and uses language and a format familiar to the health measurement field and, most important, to the end users of this information and our key stakeholders: clinicians measuring the competence of their students in clinical education settings.

The goal of this phase was to investigate the face and content validity of the draft measure through broad consultation with physiotherapists across Canada who had supervised and assessed at least one student in clinical practice in the past 12 months. *Face validity* is the extent to which a measure appears to be measuring what it is intended to measure,⁸ and *content validity* is defined as the extent to which a measure consists of a comprehensive sample of items that completely assess the domain of interest.⁸ Both are essential components of an assessment measure. Results of the survey will be used to inform the next draft, which will be pilot tested across Canada.

METHODS

We used LimeSurvey (<https://www.limesurvey.org/en/>), an online survey application, to conduct a cross-sectional survey. Ethics approval was granted by the Health Sciences Research Ethics Board at the University of Toronto.

Participants

We recruited PT CIs across Canada who were registered to practise in their province or territory and had supervised at least one student in the previous 12 months for whom they had completed a formal evaluation using the current evaluation measure (the Physical Therapy—Clinical Performance Instrument).^{9,10}

CIs were invited to participate in this study in January 2013 via the Canadian Physiotherapy Association's electronic biweekly newsletter, *National Rounds*, with a paragraph introducing the study and a link to the draft measure and the online survey. A reminder was included in the newsletter 2 weeks after the initial invitation. Physiotherapists were also invited to participate by email through each university PT programme's Director of Clinical Education or Academic Coordinator of Clinical Education. The survey was open for 3 weeks.

Survey content

To help contextualize the new assessment form, the survey asked respondents to consider the last student they supervised when answering the questions. The survey asked CIs to choose the rating scale option they thought would best allow them to assess a student on each of the items in the measure; the three proposed rating scales, which appeared in a randomized order in the survey, were (a) a discrete adjectival rating scale with six defined anchors and radio buttons, (b) a discrete adjectival rating scale with six defined anchors and radio buttons between each pair of anchors, and (c) a continuous-line adjectival rating scale with six defined

Table 1 Changes to the ECP on the Basis of Collected Data

ECP role and key competency	Draft 1 (developed in Phases 1 and 2)		Draft 2 (developed in Phase 3)	
Expert – Focus on Assessment				
1.1. Consults with the client to obtain information about his or her health, associated history, previous health interventions, and associated outcomes.	One RS for these two key competencies	One CB	RS for 1.1	One CB
1.2. Collects assessment data relevant to the client's needs and physiotherapy practice.			RS for 1.2	
Expert – Focus on Analysis				
1.3. Analyzes assessment findings.	One RS for these three key competencies	One CB	RS for 1.3	One CB
1.4. Establishes a physiotherapy diagnosis and prognosis.			RS for 1.4	
1.5. Develops and recommends an intervention strategy.			RS for 1.5	
Expert – Focus on Intervention				
1.6. Implements intervention.	One RS for these three key competencies	One CB	RS for 1.6	One CB
1.7. Evaluates the effectiveness of interventions.			RS for 1.7	
1.8. Completes physiotherapy services.			RS for 1.8	
Communicator				
2.1. Develops, builds, and maintains rapport, trust, and ethical professional relationships through effective communication.	RS for 2.1	One CB	RS for 2.1	One CB
2.2. Elicits, analyzes, records, applies, conveys, and shares information.	RS for 2.2		RS for 2.2	
2.3. Employs effective and appropriate verbal, non-verbal, written, and electronic communications.	RS for 2.3		RS for 2.3	
Collaborator				
3.1. Establishes and maintains inter-professional relationships, which foster effective client-centred collaboration.	RS for 3.1	One CB	RS for 3.1	One CB
3.2. Collaborates with others to prevent, manage, and resolve conflict.	RS for 3.2		RS for 3.2	
Manager				
4.1. Manages individual practice effectively.	RS for 4.1	One CB	RS for 4.1	One CB
4.2. Manages and supervises personnel involved in the delivery of physiotherapy services.	RS for 4.2		RS for 4.2	
4.3. Participates in activities that contribute to safe and effective physiotherapy practice.	RS for 4.3		RS for 4.3	
Advocate				
5.1. Works collaboratively to identify, respond to, and promote the health needs and concerns of individual clients, populations, and communities.	RS for 5.1	One CB	RS for 5.1	One CB
Scholarly Practitioner				
6.1. Uses a reflective approach to practice.	One RS for these three key competencies	One CB	One RS for these three key competencies	One CB
6.2. Incorporates lifelong learning and experiences into best practice.				
6.3. Engages in scholarly inquiry.				
Professional				
7.1. Conducts self within legal and ethical requirements.	RS for 7.1	One CB	RS for 7.1	One CB
7.2. Respects the individuality and autonomy of the client.	RS for 7.2		RS for 7.2	
7.3. Contributes to the development of the physiotherapy profession.	RS for 7.3		RS for 7.3	

ECP = Essential Competency Profile; RS = rating scale; CB = comment box.

a) Discrete adjectival rating scale (6 anchors, 6 radio buttons)

b) Discrete adjectival rating scale (6 anchors, 10 radio buttons)

c) Continuous lined adjectival rating scale (6 anchors, including 1 box)

Figure 1 Three rating scale options proposed in the survey.

anchor points, one of which is a box at the rightmost end of the scale (see Figure 1). The survey also asked CIs about the grouping of the proposed items within the roles and their alignment into logical subdomains.

Although the content of the measure was defined by the ECP, we sought some input as to the specificity with which the measure would address the ECP competencies and planned to make further refinements to the measure if CIs felt that some items were redundant or could easily be grouped together under a higher order attribute to create a more parsimonious measure. Finally, two questions with Likert-scale response options asked for CIs' overall impression of the measure, and respondents were asked in an open-ended question to share any additional comments about the new tool. We also asked CIs for input on the name of the new assessment measure, presenting six response options and "I have no preference." Demographic questions (gender, year of birth, year of graduation from PT, province or territory of residence) and questions about area of practice and number of students supervised in the previous 5 years were also included in the survey to allow us to describe the sample and comment on the extent to which it was representative of Canadian PT practice as compared with the Canadian Institute of Health Information's *Physiotherapists in Canada, 2010*¹¹ and a recent Ontario publication on CE placements.¹²

Analysis

We calculated parametric and non-parametric descriptive statistics, as appropriate to the data, to characterize the responses and describe the results. First, the sample was described, including gender, year of birth, year of graduation from PT, and province or territory of residence, for comparison with the Canadian PT context. Next, the quality of the data was described. We analyzed responses regarding the rating scale. Subsequently, we analyzed the survey data on item refinement and the name of the measure. Last, we analyzed data from the 5-point Likert-scale items (1 = *strongly disagree*, 5 = *strongly agree*) that asked for CIs' level of agreement with the statements "This measure would allow you to adequately identify a student who was performing poorly" and "This measure would allow you to adequately identify a student who was performing very well." In addition, we reviewed responses to the open-ended question about respondents' general opinions of the ACP.

Terminology

Throughout this article, the term *measure* refers to the clinical education assessment instrument currently under development, which is titled ACP for the purposes of this study phase. The term *role* refers to the domains of the measure; the roles are the same ones identified in the ECP (Expert, Communicator, Collaborator, Manager, Advocate, Scholarly Practitioner, Professional). The term

items refers to the key competencies that describe a role, which are assessed using the rating scale. The *rating scale* is the anchored response scale that the student or the CI uses to assess each item.

RESULTS

Demographics

Of the 537 CPA members who consented to participate, 390 self-identified as meeting the inclusion criteria and proceeded to the survey, and 259 completed it. Tables 2 and 3 provide descriptive information about the participants. The average age of the sample was 40.2 years; 84.4% were female, and participants graduated an average of 15.8 years ago. The majority (57.3%) had supervised between one and five students in the previous 5 years. Of those who answered the demographics questions, 77% were working in an urban centre, 40.9% worked in an acute-care centre, representing a variety of practice areas (musculoskeletal, 22.2%; neurological, 17.6%; or mixed caseload, 17.5%) and 33.8% were practising in Ontario.

Rating scale

When asked "Considering the last student you supervised, please choose the scale option that you think would best allow you to assess the student on each item with a rating scale," 40% preferred the ordinal scale with 10 boxes (option b), 42% preferred the continuous-line adjectival rating scale (option c), and 8% indicated no preference.

Table 2 Demographics of the Physiotherapy Clinical Instructors Sample

Characteristic	No. (%) of respondents*
Gender (<i>n</i> = 263)	
Female	222 (84.4)
Male	41 (15.6)
Age (<i>n</i> = 259), y, mean (SD), range	40.2 (10.4), 25–65
Time since graduation (<i>n</i> = 263), y, mean (SD); range	15.8 (10.8); <1–40
No. of students supervised in past 5 y (<i>n</i> = 267)	
1–5	153 (57.3)
6–10	76 (28.5)
11–15	22 (8.2)
16–20	9 (3.4)
>20	7 (2.6)
Current practice location (<i>n</i> = 266)	
British Columbia	36 (13.5)
Alberta	37 (13.9)
Saskatchewan	25 (9.4)
Manitoba	12 (4.5)
Ontario	90 (33.8)
Quebec	21 (7.9)
New Brunswick	9 (3.4)
Nova Scotia	4 (1.5)
Prince Edward Island	17 (6.4)
Newfoundland and Labrador	14 (5.3)
Yukon	1 (0.4)

*Unless otherwise indicated.

Table 3 Practice Setting

Survey item	No. (%) of responses
Location of where you work (<i>n</i> = 274)	
Urban	211 (77.0)
Semi-rural	28 (10.2)
Rural	29 (10.6)
Remote	6 (2.2)
Age group of patients or clients you primarily work with (<i>n</i> = 263)	
0–18 years	56 (21.3)
19–65 years	49 (18.6)
66 and older	47 (17.9)
Mixed ages	111 (42.2)
Your areas of practice (<i>n</i> = 270)*	
Musculoskeletal or orthopaedics	121 (22.2)
Neurology or neurosciences	96 (17.6)
Mixed caseload	95 (17.5)
Cardiopulmonary	62 (11.4)
General rehabilitation	46 (8.5)
Pain	20 (3.7)
Oncology	19 (3.5)
Rheumatology	15 (2.8)
Hand therapy or plastic surgery	13 (2.4)
Burns and wound care	7 (1.3)
Other	50 (9.2)
Your practice setting (<i>n</i> = 270)*	
Acute care hospital	131 (40.9)
Rehabilitation hospital or facility	74 (23.1)
Private practice	38 (11.9)
Community care or home care	17 (5.3)
Community health centre	15 (4.7)
Long-term care facility	13 (4.1)
Administration or research	6 (1.9)
Insurance sector	1 (0.3)
Other	25 (7.8)

*Multiple responses permitted; percentages for this section are based on 544 total responses.

Further item refinement

The first draft of the measure proposed in this survey had one rating scale to assess the student's performance under the heading Expert—Focus on Assessment, which included ECP items 1.1 and 1.2. The majority (57.8%) said they would prefer to assess each key competency on its own rating scale. Similarly, the draft version grouped key competencies 1.3, 1.4, and 1.5 together with one rating scale under the heading Expert—Focus on Analysis, but the majority of responses (59.0%) indicated a preference for one rating scale per key competency. Respondents were more evenly divided on key competencies 1.6, 1.7, and 1.8 (Expert—Focus on Intervention); 50.2% felt that these three could be assessed with one rating scale, and 49% preferred each key competency to have its own rating scale. In contrast, 80.2% felt that one rating scale would be adequate to assess the student's performance in the role of Scholarly Practitioner (composed of key competencies 6.1, 6.2, and 6.3).

Table 4 Number of Rating Scales for Each Role

Question	No. (%) of respondents		
	Prefer one rating scale	Prefer multiple rating scales	No response
Expert – Focus on Assessment: Should we keep 1.1 and 1.2 on one rating scale? (<i>n</i> = 294)	119 (40.5)	170 (57.8)	5 (1.7)
Expert – Focus on Analysis: Should we keep 1.3, 1.4, and 1.5 on one rating scale? (<i>n</i> = 290)	112 (38.6)	171 (59.0)	7 (2.4)
Expert – Focus on Intervention: Should we keep 1.6, 1.7, and 1.8 on one rating scale? (<i>n</i> = 287)	144 (50.2)	140 (48.8)	3 (1.0)
Scholarly Practitioner: Should we keep it on one rating scale? (<i>n</i> = 278)	223 (80.2)	51 (18.3)	4 (1.4)
Communicator – Would you have been able to accurately separate this student's performance into each of the three competencies? (<i>n</i> = 271)	120 (44.3)	149 (55.0)	2 (0.7)
Collaborator – Would you have been able to accurately separate this student's performance into each of two competencies? (<i>n</i> = 272)	68 (25)	202 (74.3)	2 (0.7)
Manager – Would you have been able to accurately separate this student's performance into each of three key competencies? (<i>n</i> = 272)	98 (36.0)	169 (62.1)	5 (1.8)
Professional – Would you have been able to accurately separate this student's performance into each of the three key competencies? (<i>n</i> = 271)	92 (34.0)	178 (65.7)	1 (0.4)
Advocate – One rating scale would have been adequate to assess the student's performance in this area. (<i>n</i> = 270)			
Strongly agree	119 (44.1)		
Agree	126 (46.7)		
Neutral	13 (4.8)		
Disagree	7 (2.6)		
Strongly disagree	4 (1.5)		
No answer	1 (0.4)		

The majority of CIs thought there should be one rating scale for each key competency within the roles of Communicator (55.0%), Collaborator (74.3%), Manager (62.1%), Professional (65.7%), and Advocate (90.8%). See Table 4 for detailed results.

General questions about the ACP

Although 88% of respondents said that this measure would allow them to adequately identify a student who was performing poorly, an overwhelming 95% felt it would allow them to adequately identify a student who was performing very well. General comments reflected an appreciation for the Canadian content as well as the categories and organization of the new measure, and some respondents were hopeful that it would be available online and take less time than the current measure:

I have been waiting for some time for a Canadian evaluation measure to be developed. I think the current measure, while able to evaluate students, had many components that were irrelevant to many areas of practice in Canada or within different areas of practice. I think this measure is more specific to the competencies we test for during the national examination process.

The new draft does appear to be more suitable to Canadian CIs.

I do like the categories better.

I like that the rating system is more clearly defined. The competencies and examples listed under each role are comprehensive. This scale looks like it will be more user-friendly than the CPI.

I would like it to be available in electronic format so comments can be typed in.

I feel like this is much more user friendly than the current measure and appears that [it] will be quicker to complete!

We asked for CIs' input on the name of the new measure and provided a list of six potential names. Although 14% expressed no preference, the name *Canadian Physiotherapy Assessment of Clinical Performance (ACP)* received the highest number of votes (25%), and no other option received more than 20% support.

Changes to the draft ACP as a result of this study are summarized in the last two columns of Table 1.

DISCUSSION

Gathering input and feedback through broad consultation with key stakeholders is an essential step in this phase of measure development. Because these stakeholders are often the people who will need to complete the measure, response burden is an important consideration.¹³ In our case, we sought the opinions of CIs in an effort to make the measure as acceptable for its real-world application as possible. Through broad consultation, we found that CIs preferred a rating scale with 10 boxes or a continuous-line adjectival rating scale, would prefer one rating scale for the Scholarly Practitioner role and one rating scale for each remaining key competency, and believed that the measure would help them adequately identify a student who was performing very well or performing poorly.

Sample

We aimed to recruit CIs from a variety of practice areas across Canada. Representation from diverse practice areas is important to ensure content and face validity across a variety of settings in which physical therapists work and students complete clinical internships. In 2011, the Canadian Institute for Health Information (CIHI) published a comprehensive snapshot of the Canadian PT workforce.¹¹ Our sample is similar to the CIHI data in some respects: Of CIs, 77.6% were female, and the average age was 41.7 years; most (91.7%) worked in an urban centre, and one third (33.1%) worked in Ontario. In other respects, our sample differed from the data reported by CIHI: Of Canadian physiotherapists, 22.5% work in Quebec, but only 7.9% of our sample did so; fewer CIs in our sample practised in musculoskeletal areas of practice (22.2% vs. 42.6%), but our sample included more CIs practising in neurology and cardiorespiratory care and more CIs working in hospital or rehabilitation facilities (64.1% vs. 38.1% in the CIHI data). These differences reflect the curriculum goals of Canadian PT programmes, which work to ensure that students graduate as generalists with experience in a variety of areas of practice.^{12,14}

Rating scale

One of our goals was to develop a rating scale that is objective and easy to interpret, with clearly defined anchors, which CIs had identified as important in a previous study.¹⁵ Gathering input on CIs' preferred student assessment rating scale was critical, both because we value CIs' opinions and because we hoped to encourage stakeholder buy-in by involving them in the process of selecting the rating scale to be used in the ACP.

CIs' preferences were almost evenly split between the 10-box adjectival scale (Likert type) and the continuous-line rating scale (visual analog scale, or VAS). When the VAS and Likert scales are compared, simple VAS and numeric VAS had comparable reliability.¹⁶ Similarly, in a study comparing a 7-point Likert scale with a VAS to assess quality of life in a cohort of 20 people with heart disease, the two scales had comparable responsiveness and validity, but the 7-point Likert scale was easier to administer and interpret and was therefore recommended.¹⁷ Guyatt and colleagues,¹⁸ who compared a 7-point Likert scale with the VAS to assess change in quality of life for a small cohort of 28 people with chronic lung disease after an inpatient rehabilitation programme, also concluded that the two methods showed comparable responsiveness. Their analysis found that the improvement was not significantly different as measured by the two types of scale; however, although the VAS showed greater improvement than the Likert scale, it also showed variability in the improvement. Thus, the VAS and Likert-type scales appear to have similar psychometric properties.

As well as considering the psychometric properties of rating scales, we also considered their utility aspects. We wanted a rating scale that would yield the same information whether it was completed on paper or electronically. An electronic sliding scale is difficult to create, and photocopying can create variations in the actual length of the line. We also wanted our rating scale to provide data that are easy to analyze and to minimize sources of error. Interpreting the VAS on the current CPI, for example, requires three measurements (total length of the line, the midpoint mark, and the final point mark) and is not efficient. Therefore, considering both measurement properties and utility features, we chose the 10-box ordinal scale for the next draft of the measure.

Item refinement

Respondents expressed a preference for a distinct rating scale for each key competency in the areas of Expert—Focus on Assessment and Expert—Focus on Analysis. For example, although key competency items 1.1 and 1.2 relate to the subjective and objective components of assessment within the Expert role, CIs preferred having the option to assess the student on each key competency rather than on one item that combined several key competencies. Rating the student on distinct key competencies has several advantages. First, it allows the CI to distinguish and provide more specific feedback on the student's performance in each area of assessment. Second, it simplifies what the CI needs to consider when assessing the student's performance; one comprehensive review on sources of rater bias found that raters have difficulty in considering multiple factors when assessing performance.¹⁹

With respect to the role of Expert—Focus on Intervention, the responses were split between using one rating scale for key competencies 1.6, 1.7, and 1.8 and using a rating scale for each key competency. Because this is a critical area of development for students, and given the advantages of allowing the CI to assess each item separately, as stated earlier, we decided to include a rating scale for each key competency for the section Expert—Focus on Intervention.

In contrast, CIs indicated that one rating scale for the role of Scholarly Practitioner would suffice for key competencies 6.1, 6.2, and 6.3. This finding may reflect the fact that these key competencies are more closely related and thus harder to differentiate, at least in the context of student internships. Therefore, only one rating scale was retained for the Scholarly Practitioner role.

Overall, responses to the survey affirmed our proposed layout with respect to items to be assessed with rating scales, agreeing that one rating scale for each remaining key competency was preferred.

General impressions

Respondents' agreement that the measure would help them adequately identify a student who was performing

very well or performing poorly can be considered as establishing its face validity. This element is critical because both CIs and academic faculty want to be able to reinforce when a student is performing very well and because it is important to identify when a student is not performing well, to allow for educational intervention, remediation, and opportunities for improvement.

Limitations

This study has several limitations. First, the response rate is difficult to determine because multiple methods were used to invite participation. Approximately 10,000 CPA members would have received the association's newsletter, and physiotherapists (who may or may not have been CPA members) were also invited to participate by email through each university PT programme's Director of Clinical Education or Academic Coordinator of Clinical Education. Although 390 CIs consented to participate and met the inclusion criteria, there were only 274 responses to the question on geographical practice setting (urban or rural) and 259 responses to the question on CI's year of birth. Some participants may not have completed the survey because it required cross-referencing with the ECP and attention to detail. However, we placed the question about the rating scale first because we wanted to ensure the maximum number of responses to that question, and we placed the demographic questions at the end. In addition, the online survey was not constructed to control for the possibility of multiple entries by the same person. Second, our survey was available only in English, which may have discouraged CIs from Quebec from participating. The next phase of testing will ensure broader representation in both English and French as we pilot test the ACP across Canada.

CONCLUSIONS

This study set out to continue investigating the face and content validity of the new measure through broad consultation with CIs across Canada, a key component in the development process, to help ensure that the measure is valid, acceptable, and meaningful. CIs indicated their preferences in terms of the design, organization, and naming of the ACP and were confident that it would allow them to identify students across the performance continuum.

Readers should note that the description of the ACP in Table 1 should be considered preliminary until additional validation testing is complete. The next phase of development will be pilot testing the ACP in English and French with students and CIs across a variety of areas of practice, regions of Canada, and stages of the students' programme. Data from this study phase will be used to analyze the reliability, validity, and practicality of the ACP across a variety of clinical education settings in Canada.

KEY MESSAGES

What is already known on this topic

Assessment in clinical education is an important part of the physiotherapy students' development into a competent entry-level clinician. Face and content validity are important measurement properties of an assessment measure.

What this study adds

Physiotherapy clinical instructors across Canada have provided input into and feedback on the development of the Canadian Physiotherapy Assessment of Clinical Performance, the new measure to assess physiotherapy students' clinical performance behaviours in the clinical education setting.

REFERENCES

1. National Physiotherapy Advisory Group. Essential Competency Profile for Physiotherapists in Canada. Toronto: Canadian Physiotherapy Association; 2009.
2. Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assessment Res Eval*. 2007;12(10).
3. Mori B, Brooks D, Norman KE, et al. Development of the Canadian Physiotherapy Assessment of Clinical Performance: a new tool to assess physiotherapy students' performance in clinical education. *Physiotherapy Canada* 2015; 67(3):281–289; doi:10.3138/ptc.2014-29E
4. 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>.
5. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50(1):1–73. <http://dx.doi.org/10.1111/jedm.12000>.
6. Beaton DE, Boers M, Tugwell P. Assessment of health outcomes. In: Firestein GS, Budd RC, Gabriel SE, et al, editors. *Kelley's textbook of rheumatology*. 9th ed. Philadelphia: Saunders Elsevier; 2012. p. 462–75.
7. Messick S. Validity. In: Linn RL, editor. *Educational measurement*. 3rd ed. New York: American Council on Education and Macmillan; 1989. p. 13–103.
8. Finch E, Brooks D, Stratford PW, et al. *Physical rehabilitation outcomes measures: a guide to enhanced clinical decision-making*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 2002.
9. American Physical Therapy Association. *Physical therapist clinical performance instrument*. Alexandria, VA: The Association; 1997.
10. Task Force for the Development of Student Clinical Performance Instruments, American Physical Therapy Association. The development and testing of APTA clinical performance instruments. *Phys Ther*. 2002;82(4):329–53. Medline:11922850
11. Canadian Institute for Health Information. *Physiotherapists in Canada*, 2010. Ottawa: The Institute; 2011.
12. Norman KE, Booth R, Chisholm B, et al. Physiotherapists and physiotherapy student placements across regions in Ontario: a descriptive comparison. *Physiother Can*. 2013;65(1):64–73. <http://dx.doi.org/10.3138/ptc.2011-63>. Medline:24381384
13. Auger C, Demers L, Swaine B. Making sense of pragmatic criteria for the selection of geriatric rehabilitation measurement tools. *Arch Gerontol Geriatr*. 2006;43(1):65–83. <http://dx.doi.org/10.1016/j.archger.2005.09.004>. Medline:16257067
14. National Physiotherapy Advisory Group. *Entry-to-practice physiotherapy curriculum: a companion document*. Clinical education guidelines for Canadian university programs. Council of Canadian Physiotherapy University Programs; 2011.

15. 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
16. van Laerhoven H, van der Zaag-Loonen HJ, Derkx BH. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr.* 2004;93(6):830–5. <http://dx.doi.org/10.1111/j.1651-2227.2004.tb03026.x>. Medline:15244235
17. Jaeschke R, Singer J, Guyatt GH. A comparison of seven-point and visual analogue scales: data from a randomized trial. *Control Clin Trials.* 1990;11(1):43–51. [http://dx.doi.org/10.1016/0197-2456\(90\)90031-V](http://dx.doi.org/10.1016/0197-2456(90)90031-V). Medline:2157581
18. Guyatt GH, Townsend M, Berman LB, et al. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis.* 1987;40(12):1129–33. [http://dx.doi.org/10.1016/0021-9681\(87\)90080-4](http://dx.doi.org/10.1016/0021-9681(87)90080-4). Medline:3680471
19. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med.* 2003;15(4):270–92. http://dx.doi.org/10.1207/S15328015TLM1504_11. Medline:14612262