Evidencing physiotherapy students' preparedness for practice: A validation study

Y. Guisard[1], S. Hyde[2], K. Skinner[3] & M. D. Simpson[4]

Abstract

Introduction: In the health professions, self-assessment and reflective practice are required professional competencies. Student capacity to reflect on and self-assess their preparedness for practice as they transition from an undergraduate student into a graduate health professional requires scaffolding these skills in their academic programs. Drawing on medical education, we evaluated the usefulness of a previously validated tool to measure student perceptions of preparedness in a problem-based undergraduate physiotherapy degree.

Methods: Rasch and factor analyses were applied to a modified version of the Preparedness for Practice Questionnaire (PHPQ) to ascertain the construct validity of the instrument and to assess the effect of teaching method on students' perceived preparedness for practice.

Results: The PHPQ should be considered as a set of subscales rather than an instrument that measures a single construct. Some subscales were found to be valid and evidenced a significant effect of the teaching pedagogy. However, the "collaboration" subscale could not be validated, and several others were only partially validated and require further refinement.

Conclusions: This study has implications for the future use of the PHPQ in similar contexts in terms of student self-assessment of preparedness and suggests that students are really self-assessing a number of capabilities rather than an overall sense of preparedness. Although this reflection is still useful for practice, it lacks face validity at the moment when the current PHPQ is used. The PHPQ requires further refinement in order to be used confidently as a self-assessment tool for students to evaluate their preparedness for practice as physiotherapists.

1 Faculty of Science, Charles Sturt University, Orange, NSW, Australia
2 Three Rivers University Department of Rural Health, Charles Sturt University, Orange, NSW, Australia
3 School of Community Health, Charles Sturt University, Orange, NSW, Australia
4 School of Biomedical Sciences, Charles Sturt University, Orange, NSW, Australia

Correspondence
Yann Guisard
Course Director
Charles Sturt University
Leeds Parade
Orange, NSW 2800
Australia
Tel: +61 2 6365 7889
Email: yguisard@csu.edu.au

ISSN 1442-1100

Keywords: PHPQ; preparedness; problem-based learning; Rasch; self-report.

Introduction

"Reflective practitioner and self-directed learner" is one of seven core roles required of registered physiotherapists in both Australia and Aotearoa New Zealand (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015). As such, students' perceptions of how well their university course has prepared them for self-directed learning, and their ability to reflect, is of significant interest to educators and the profession.

Possessing sufficient capacity for self-assessment and the ability to recognise their own limitations and reflect on areas in need of improvement is a requirement for graduates that is grounded in national health professional standards (Table 1) and contributes to ensuring patient safety. The translation of the ability to self-assess and reflect on learning in the clinical environment does not always occur, especially at times of transition between formal structured learning programs and the clinical context, which is often stressful for new graduates (Smith & Trede, 2013; Yew & Goh, 2016). Self-evaluation of students' confidence in a specific professional skill set, as measured by self-report instruments, is one way to measure the preparedness of graduates for practice as a self-directed learner and health professional.

Student capacity to reflect on and self-direct their learning can be enhanced in curricula through the use of problem-based learning pedagogy (e.g., Loyens, Magda, & Rikers, 2008). Problem-based learning (PBL) uses small teaching groups of six to eight students, a tutor whose role is to facilitate discussion rather than to deliver content knowledge and authentic cases (Muhamad, Henry, & Ramli, 2016). In PBL, students collaborate within the group to construct and apply their own knowledge and understanding of the case through a series of tutorials, which can vary in frequency from one to three times a week.

Within the tutorial, students are guided by the initial case presentation, hypothesise about the cause of the patient's problem, discuss what they already know in relation to the problem, identify areas for further learning and reconvene at a later time in the week to discuss what has been learnt in the interim between tutorial sessions. This discussion in PBL evokes ideas for individuals that they might not have generated had they been working on their own (Skinner, Hyde, McPherson, & Simpson, 2016). This collaborative approach to learning and teaching, therefore, creates opportunities for self-directed learning (SDL) and reflection.

A number of PBL studies provide evidence that PBL students learn "how to learn" and display increased skills in academic regulation, such as metacognitive knowledge monitoring, self-evaluation, goal setting and self-directed learning, when compared to their non-PBL peers (e.g., Blumberg, 2000; Hadwin, 1996; Ryan, 1997; Wijnen, Loyens, Smeets, Kroeze, & van der Molen, 2016; Yew & Schmidt, 2009). The achievement of specific learning outcomes from PBL, including self-directed learning and reflection, is dependent, however, on the success of the group and on interactions within the group (Holen, 2000; Ngeow & Kong, 2001; Peterson, 1997). Furthermore,

ISSN 1442-1100

Table 1
Examples of Professional Standards[1]

Discipline	Standard	Sub Standard
Nursing and midwifery[2]	Standard 1: Thinks critically and analyses nursing practice	1.2 develops practice through reflection on experiences, knowledge, actions, feelings and beliefs to identify how these shape practice
Paramedicine[3]	3.3. Practices within an approved scope of practice	Utilises a range of integrated skills and self-awareness to manage clinical challenges effectively in unfamiliar circumstances or situations
	9.1 Monitors and reviews the ongoing effectiveness of their practice and modifies it accordingly	Monitors and evaluates the quality of practice and the value of contributing to the generation of data for quality assurance and improvement programs
		Considers feedback from colleagues about and critically reflects on their own Paramedic practice
	9.2 Audits, reflects on and reviews practice	Reflects on practice and the application of such reflection to their future practice
Pharmacy[4]	1.1: Uphold professionalism in practice	Enabling Competency 4. Accept professional responsibility and accountability. • Reflects on professional performance with a view to improving outcomes.
	1.4: Maintain and extend professional competence	Enabling Competency 1. Adopt a scope of practice consistent with competence. General level: • Recognises and responds to situations outside own competence.
	4.1: Show leadership of self	Enabling Competency 1. Display emotional awareness and effective self-regulation of emotions. General level: • Recognises and manages the impact of assumptions, values, beliefs, attitudes and behaviours on self and others. • Displays control of emotions, in particular the ability to manage disruptive emotions and impulses.
		Enabling Competency 2. Apply reflective skills for self-assessment. General level: • Uses self-reflection to assist with continuous self-development and growth. • Reflects on leadership effectiveness and adopts an empathetic and adaptive leadership style.
Physiotherapy[5]	Role 4. Reflective practitioner and self-directed learner	4.1 Assess their practice against relevant professional benchmarks and take action to continually self-directed learner improve their practice
		4.2 Evaluate their learning needs, engage in relevant continuing professional development and recognise when to seek professional support, including peer review

Key to table on following page

ISSN 1442-1100

Table 1
Examples of Professional Standards[1] (contd.)

1 From a range of health professions relevant to this study, identified using a word search on keywords "own", "self", "reflect" and "awar(e, ness…)". The standards were noted only when the nature of the standard relates to self-assessments or self-awareness of a desirable professional trait.
2 http://www.nursingmidwiferyboard.gov.au/Codes-Guidelines-Statements/Professional-standards.aspx
3 http://www.caa.net.au/images/documents/accreditation_resources/Paramedic_Professional_Competency_Standards_V2.2_February_2013_PEPAS.pdf
4 http://www.acp.edu.au/imis15/documents/ACP/National_Competency_Standards_Framework_for_Pharmacists_in_Australia_2016.pdf
5 https://physiocouncil.com.au/media/1020/physiotherapy-board-physiotherapy-practice-thresholds-in-australia-and-aotearoa-new-zealand-6.pdf

evidence is emerging that reflective thinking is affected by key variables such as a supportive environment, authentic context, mentoring, group discussion, support and free expression of opinions (Mann, Gordon, & MacLeod, 2009), all of which are present in the PBL approach.

The Preparation for Hospital Practice Questionnaire (PHPQ) is a tool designed to measure students' preparedness for practice prior to practice (Hill, Rolfe, Pearson, & Heathcote, 1998). This instrument has been used in other health professional degree programs and was especially valued for the currency and stability of items. To investigate the utility of the instrument for our purpose of evaluating course outcomes and providing justification for an integrated PBL curricula, we decided to further validate the instrument using a combination of Rasch and factor analyses to strengthen the basis on which to make core curricula decisions. We suggest that professional standards such as self-reflection, ongoing self-evaluation and ongoing critical self-appraisal are difficult to evidence using traditional static assessments. We propose that the periodic use of a well-validated self-evaluation questionnaire may assist students in developing these non-cognitive skills. Furthermore, it would provide the teaching institution with data to inform the development of support strategies for students.

Aims

To investigate the validity of a modified PHPQ in two cohorts of physiotherapy students (before and after PBL was introduced) and to contribute towards a curriculum design discourse.

Background of the survey instrument

The Preparedness for Hospital Practice Questionnaire (PHPQ) has 41 items with eight recognised subscales (Hill et al., 1998): interpersonal skills, confidence and coping, collaboration, patient management and practical skills, understanding science, prevention, holistic care and self-directed learning. Items are anchored by a 7-point Likert scale, ranging from 0 (don't know) to 6 (very adequately). The original instrument is perceived as valid and reliable, with Cronbach's alpha coefficient from 0.78 to 0.88 across the eight subscales, indicative of high internal consistency or relatedness within each subscale. The PHPQ is widely used in the assessment of medical graduates in

ISSN 1442-1100

many nations (Dean, Barratt, Hendry, & Lyon, 2003; Kassim, McGowan, McGee, & Whitford, 2016; MacCarrick, Kelly, & Conroy, 2010; Scicluna, Grimm, Jones, Pilotto, & McNeil, 2014; Scicluna et al., 2012), in nursing (Blodgett, Blodgett, & Bleza, 2016; Christensen et al., 2016; Liaw et al., 2014) and, more recently, in physiotherapy (Hess & Frantz, 2014; Skinner et al., 2016).

Background of the educational program

At Charles Sturt University, physiotherapy has been taught since 1998, using a traditional pedagogical didactic approach in which students attended lectures for theoretical content and practical classes for clinical skills. Until 2010, the students gained relevant clinical information and skills but had limited opportunities for self-directed learning and the development of skills in reflective practice. In 2010, a significant change was made to the undergraduate physiotherapy degree with the introduction of integrated PBL.

Background and aims of the study

The PHPQ was administered to final-year physiotherapy students in two consecutive cohorts, one from the final traditional course cohort and one from the first PBL cohort. Minor modifications were made to 13 of the 41 items to reflect physiotherapy practice. These modifications did not change the essence of the questions, and as such, the authors felt that these changes would minimally impact the reliability of the questionnaire (see Skinner et al., 2016, p. 25).

Our intention in the current study is to re-analyse Skinner et al.'s (2016) data with a focus on indicators that support or reject the use of the PHPQ as either:

a) an instrument that aligns with a single latent trait (the "preparedness for hospital practice") and can therefore be represented using a single achievement score, or

b) an instrument that is an aggregate of a number of closely related traits (subscales), whereby each subscale requires its own score.

Methods

To investigate the statistical validity of the PHPQ, we re-analysed Skinner et al.'s (2016) original data using a combination of Rasch and factor analysis (FA) procedures. As described by Skinner et al. (2016), a sample of 58 students were administered a paper copy of the instrument in the final week of study.

The Rasch model is the simplest mathematical formulation belonging to a family of models aligned with the item response theory (IRT). It computes the probability of individuals (persons) responding correctly to a question (item) on the basis of the relationship between the "ability" of the person and the "difficulty" of the item. The Rasch model assumes that the responses provided by persons are influenced by a single latent trait. Rasch models rest upon a set of four assumptions (unidimensionality, conditional independence, sufficiency and monotonicity) (Rasch, 1980). If these assumptions are satisfied, the user can be confident that the instrument has the

ISSN 1442-1100

listed properties, including the separability of the persons and items, objectivity (the instrument is sample independent) and the fact that person and item scales use the same units to measure ability and difficulty. In this study, the RUMM2030 (RUMM Laboratory Pty Ltd, Version 5.4., Perth, Australia) software was used. In view of the small sample size, a single factor (PBL/traditional teaching method) was tested.

Factor analysis is commonly used in the process of scale design, validation and dimension reduction. Factor analysis also rests upon four assumptions: normality of the variables, linear relation between variables, factorability and appropriate sample size. In our study, the first three assumptions were respected, but the sample size was smaller than considered appropriate (the variable:factor ratio was 1:1.4; recommended minimum values 1:5). We used SPSS (IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.) for our factor analysis. The procedure used included a principal component analysis, combined with varimax rotation. The number of retained factors was not limited, and all factors with eigenvalues greater than 1 were retained for assessment. The Kaiser–Meyer–Olkin (KMO) test was used to assess sample adequacy.

The original dataset (n = 58 students) was used for a preliminary Rasch analysis. Two respondents were considered "extreme" in their responses (i.e., consistently responded using the "very adequately" category). Extreme students do not yield significant information in Rasch analyses and were eliminated from further analyses. Using this slightly reduced dataset (n = 56 students), five analyses were computed (Table 2). Theoretically, only the first analysis would be necessary to validate a published model, however our results suggested that further investigation was warranted.

Table 2
Description of the Analyses From This Study and Associated Rationale

Analysis	Nature of Analysis	Rationale
1	Original full model	Validation of the published PHPQ scale
2	Re-scored full model and removed extreme persons	Analysis 1 suggested that the original full model displayed disordered thresholds and extreme behaviours
3	Original subscales	The second analysis displayed multidimensionality
4	Re-scored subscales	The third analysis suggested that the original full model displayed disordered thresholds
5	Factor analysis on original dataset	An investigation of the subscales suggested in this dataset using a methodology close to that published in the original paper

ISSN 1442-1100

Results

Analyses 1 and 2: Full and re-scored models

The analysis of the original dataset included all respondents and all items in the questionnaire. Overall, the data was found to fit the model well (chi squared > 0.05, Table 3). Residual standard deviations were in an acceptable range (< 1.5) for both items and persons. The Person Separation Index (PAI) was good (> 0.7), but the analysis displayed a large amount of local dependencies (r > 0.2), suggesting that some items could be removed from the scale. There was no differential item functioning (DIF), indicating that respondents answered the survey questions in similar manners, regardless of the teaching method. Items thresholds were very poor, and only 11 of the 41 items displayed ordered thresholds. This suggested that respondents had difficulty identifying specific response categories, particularly the difference between "very inadequate" and "inadequate".

Finally, item targeting was poor. The persons mean was 2.6, indicating that the respondents easily displayed a significant amount of the measured latent trait, and the person–item map (not shown) indicates that the spread of items matched the person spread very poorly. The data displayed a strong item skew towards "easy to answer" items scores and "have a lot of the trait" person scores.

Thirty items were re-scored by collapsing categories. Most re-scored items combined the "inadequate" and "neutral" categories. The re-scored full model displayed similar model fit characteristics to the full model and fixed all threshold issues, however the multi dimensionality and person skew remained. These results suggest that the latent trait measured by the scale is not a single "preparedness for practice" trait but rather a small subset of interdependent constructs potentially aligned with the subscales.

Analyses 3 and 4: Original and re-scored subscales

The analysis of the individual original subscales yielded a good data fit to the model for all subscales (chi squared > 0.05, Table 3). Items- and persons-fit standard deviations were good (< 1.5), however persons means were strongly positively skewed, indicating that the respondents easily displayed a large amount of the measured latent traits in each scale.

The subscale "interpersonal skills" did not need re-scoring, displaying a positive PBL cohort effect (i.e., PBL students endorsed the items significantly more than non-PBL students; Table 4). It can, therefore, be used as proposed. Conclusions drawn by Skinner et al. (2016) are, thus, supported by the current analysis. Skinner et al. do note, however, that despite improvements with this approach, interpersonal skills still remain a challenge for new graduates, with both cohorts perceiving themselves to be not quite adequately prepared, and the mean rating falling between neutral and adequate. The implications for employers are that graduates' development of interpersonal skills needs

ISSN 1442-1100

Table 3
Summary of Models' Fit[1]

	Item Mean SD	Items Fit Res.	Persons Mean SD	Persons Fit Res. pr.	Chi Square	PSI (with Extrm)	Number of Extreme Persons
Original Models							
Full model	0	0.6	2.6	1.4	0.5	0.9	0
Interpersonal skills	0	0.5	1.2	1	0.8	0.6	0
Confidence	0	0.7	2.2	1	0.5	0.6	0
Collaboration	0	0.5	3.5	0.6	0.9	0.7	7 (ID 30; **2**; **55**; 3; **14**; **44**; **35**)
Management	0	0.5	2.7	0.9	0.3	0.5	3 (ID 25; **45**; **55**)
Science	0	0.2	2.5	0.9	0.6	0.4	4 (ID 27; 37; **2**; **28**)
Prevention	0	0.6	3.5	0.9	0.9	0.6	2 (ID 43; 58)
Holistic care	0	0.4	3.2	1.1	0.8	0.8	3 (ID **14**; **44**; **28**)
Self-directed learning	0	0.8	3.9	1	0.01	0.7	3 (ID **35**; **45**; **55**)
Re-Scaled Models							
Full model	0	0.6	2.9	1.5	0.08	0.9	0
Interpersonal skills	N/A – Thresholds were appropriate in original analysis						
Confidence	0	1	2.3	1	0.7	0.6	0
Collaboration	Did not converge						
Management	0	0.8	1.9	1.2	0.5	0.4	3 (ID 25; **45**; **55**)
Science	0	0.6	2.2	1.1	0.9	0.5	5 (ID 2; **27**; **28**; 37; **57**)
Prevention	0	0.6	3.6	1	0.9	0.6	2 (ID 43; 58)
Holistic care	0	0.4	3.5	0.8	0.8	0.7	9 (ID 14; **27**; **28**; 30; 31; **35**; 44; 54; **57**)
Self-directed learning	0	0.8	4	1.1	0.008	0.8	3 (ID **35**; **45**; **55**)

1 PSI is an approximation of Cronbach's alpha when missing data is acceptable.
Abbreviations: Fit Res. SD = Fit residuals standard deviation; PSI (with Extrm) = person separation index (with extreme persons); Bold ID of extreme person indicate that the person was extreme on two or more subscales.

formal support during their transition into work, with particular emphasis on support in developing skills necessary to work with distraught patients and dealing with death and dying.

The subcale "confidence" also displayed a PBL-cohort effect (PBL students endorsed the items significantly more than non-PBL students) but was again multidimensional. None of the other subscales displayed a significant difference between the two cohorts (i.e., no factor effect was detected in the other subscales).

ISSN 1442-1100

Table 4
Summary of Each Item's Performance

Original Models									**Re-scored Models**							**Factor**	
Item	Subscale	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	N	IL
20	Interpersonal	-1.1	0.7	✓	✓	22; **30**; 36 C1 = 0.4 C2 = 1.4	✓	$p = 0.007$	No need to rescore							4	0.67
22		1	-0.2	✓	✓	20; **30**; **36**										6	0.53
30		-0.8	0.2	✓	✓	**20**; **22**										8	0.74
36		1	-0.4	✓	✓	20; **22**										4	0.51
2	Confidence	1.3	0	✗	✓	17; 26; **37**	✗	$p = 0.005$ C1 = 1.5 C2 = 2.5	1.9	-0.3	✓	✓	**17**; 26; **37**	✓	$p = 0.005$ C1 = 1.5 C2 = 2.5	1	0.68
3		-1.6	-0.3	✗	✓	17; 37			-1.5	-0.4	✓	✓	17; 26; 37			1	0.72
6		0.8	0.1	✓	✓	17; 26; 37			0.9	0.1	✓	✓	17; 26; 37			4	0.59
17		-0.3	0.2	✓	✓	2; 3; 6			-0.3	0.1	✓	✓	**2**; 3; 6			1	0.60
26		-0.4	-0.6	✓	✓	2; 6			-0.4	-0.8	✓	✓	2; 3; 6			1	0.74
37		0.2	1.5	✗	✓	**2**; 3; 6			-0.6	2.1	✓	✓	**2**; 3; 6			2	0.36
28	Collaboration	0.1	0.1	✗	✓	**40**; **41** no SD	✓	$p = 0.06$	Could not converge							3	0.68
33		0	0	✗	✓	**40**; **41**										10	0.64
40		0	0.5	✗	✓	**28**; **33**										4	0.85
41		-0.2	-0.8	✗	✓	**28**; **33**										9	0.64
4	Management	-0.6	0.1	✗	✓	11; **31**	✓	$p = 0.5$	0	0	✓	✓	11; **31**	✓	$p = 0.4$ no SD	3	0.77
7		0	-0.1	✗	✓	11; 31			0.1	0.1	✓	✓	11; **31**			4	0.77
11		-0.5	0	✗	✓	4; 7; 31			0.6	0.4	✓	✓	4; 7; 31			6	0.71
25		-0.8	-0.1	✗	✓	**31**			-1.9	-1	✓	✓	**31**			6	0.57
31		1.9	1.1	✗	✓	**4**; 7; 11; **25**			1.1	1.2	✓	✓	**4**; **7**; 11; **25**			4	0.61

Contd. next page

ISSN 1442-1100

Table 4
Summary of Each Item's Performance (Contd.)

Original Models									**Re-scored Models**							**Factor**	
Item	Subscale	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	N	IL
8	Science	-0.5	0	✗	✓	12; **29**	✓	$p = 0.8$ no SD	0.5	0.4	✓	✓	12; 19; **29**	✓	$p = 0.9$	8	0.54
12		-0.5	-0.2	✗	✓	8; 29			-0.5	-0.2	✓	✓	8; **29**			6	0.62
19		-0.8	0	✗	✓	**29**			-1.4	-0.7	✓	✓	8; **29**			3	0.45
29		1.8	0.2	✗	✓	**8**; 12; **19**			1.5	0.6	✓	✓	**8**; **12**; **19**			2	0.85
5	Prevention	0.3	0.9	✓	✓	13; 18; 32; 34	✓	$p = 0.7$ no SD	0.2	0.7	✓	✓	13; 18; 32; 34	✓	$p = 0.6$ no SD	3	0.56
9		-0.2	-0.2	✗	✓	13			-0.6	0	✓	✓	**13**			11	0.61
13		0	1.2	✗	✓	5; 9; 18; 32			0.3	1.3	✓	✓	5; **9**; 18; 32			1	0.58
18		0.5	0.3	✓	✓	5; 13; 34			0.6	0.2	✓	✓	5; 13; 34			8	0.56
32		0	0.3	✗	✓	5; 13			0.2	-0.4	✓	✓	5; 13			11	0.72
34		-0.6	0.1	✗	✓	5; 18			-0.7	0	✓	✓	5; 18			9	0.50
1	Holistic care	-0.9	0.7	✗	✓	**21**; 24	✓	$p = 0.3$ no SD	-0.5	0.3	✓	✓	15; 21; 35	✓	$p = 0.2$ no SD	5	0.57
15		-0.6	0.3	✗	✓	21; **24**			-0.2	-0.5	✓	✓	1; 16; 24; 35			2	0.50
16		-0.5	0.3	✓	✓	**21**			-0.1	0	✓	✓	15; **21**; 24			5	0.60
21		2.4	0.3	✗	✓	**1**; 15; **16**			1.3	0.4	✓	✓	1; **16**; 35			2	0.76
24		0.5	-0.2	✗	✓	1; **15**			0.1	-0.5	✓	✓	15; 16			2	0.53
35		-0.9	-0.3	✗	✓	21			-0.7	-0.2	✓	✓	1; 15; 21			9	0.48

Contd. next page

ISSN 1442-1100

Table 4
Summary of Each Item's Performance (Contd.)

Original Models									**Re-scored Models**							**Factor**	
Item	Subscale	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	Loc	Fit Res	T	DIF	LD	U	Cohort Factor	N	IL
10	Self-directed learning	-0.5	1.8	✗	✓	14; 23; 27; 28	✓	$p = 0.07$ no SD	-0.5	1.8	✓	✓	14; 23; 27; 28	✓	$p = 0.06$ no SD	9	0.74
14		-0.4	0.3	✓	✓	10; 27; 39			-0.4	0.2	✓	✓	10; 27; 39			7	0.86
23		0.6	0	✗	✓	10; 27; 39			0.6	-0.1	✓	✓	10; 27; 39			4	0.74
27		-0.4	0.6	✗	✓	10; 14; 23			-0.4	0.5	✓	✓	10; 14; 23			12	0.66
38		-0.5	-0.4	✗	✓	10			-0.6	-0.4	✓	✓	10			12	0.45
39		1.2	0.1	✓	✓	14; 23			1.3	0	✓	✓	14; 23			12	0.73

Abbreviations:
Loc = logit location
Fit Res = fit residual
T = threshold behaviour (✓ indicates that all thresholds have clear categories)
DIF = differential item functioning (✓ indicates that items do not DIF)
LD = local dependency (Items listed display a correlation greater than 0.2 with the item studied; bold items indicate correlations of 0.4 or above)
U = unidimensionality of the *subscale* (✓ indicates that the scale is statistically unidimensional at $p = 0.05$)
Cohort factor = cohort (PBL/Non PBL) factor effect
N = factor associated with the item loading in the factor analysis
IL = item loading on the factor in the factor analysis
For details, please see text.

ISSN 1442-1100

All subscales, except "interpersonal skills", required a partial to full re-scoring of items. Similar to the re-scoring process of the full model, the subscales mostly required re-scoring in the "very inadequate" and "inadequate" categories. Several items required two sets of re-scoring across the five categories. Finally, all subscales displayed a significant number of local dependencies (suggesting the removal of one or more items).

Following re-scoring of the seven subscales, the subscale "collaboration" did not converge (Table 3). This means that the responses from participants did not follow the expected pattern of response of the model. Further computations could, therefore, not be carried out because the basic Rasch assumptions were violated.

Similarly, the "self-directed learning" subscale could be considered a poor fit to the model but would require further investigation before being discarded. Given that this domain was one of the key aspects for which a PBL cohort effect might have been expected, this may lead us to question the conclusion of "no effect" and suggest that for use in this specific context, further refinement of the tool is needed before we can make any conclusions around students' self-perceptions of their capacity for self-directed learning.

For the other six subscales, the data displayed a good items fit of the model (chi squared > 0.05). The data also displayed a good persons fit but did include some extreme behaviours. In addition, it also displayed a strong skew towards positive values for persons means, suggesting that the questions included in the subscales may need further refinement, with the addition of new items that discriminate further between students (i.e., don't make it so "easy" for the students to endorse the trait), in order to develop a more robust version of the original PHPQ for this student group. It may be possible that grounding the items further in physiotherapy language may be a more appropriate discrimination strategy. Furthermore, social desirability is known to positively influence self-reporting (Holtgraves, 2004) in at least two of a five-stage response sequence process ("retrieving information" and "making a judgment") (Sudman, Bradburn, & Schwartz, 1996). These skewed responses are well documented (Kruger & Dunning, 1999), but their quantification is difficult. Validity and item refinement in the PHPQ scale may, therefore, be enhanced through the use of focus groups with the intended sample population to explore the cognitive interpretation of specific items.

All re-scored scales were unidimensional but displayed a significant number of local dependencies (i.e., the response to one item is correlated with the response to another; Table 4). This violates the assumption in Rasch analysis that each item contributes a small, yet unique, amount of information about a trait. These results suggest that there may be scope in our future use of the PHPQ to remove one or more of the items.

The mean locations for some items displayed an "agree/disagree" behaviour rather than taking advantage of the granularity of a 5-point scale. These responses may be reflective of a degree of student overconfidence. It may, therefore, be useful for scales such as the PHPQ to identify items that could be correlated with a secondary assessment

ISSN 1442-1100

source. An example of this could be that items such as "carry out basic musculoskeletal physiotherapy treatments", "understand the physiological basis of disease" and/or "apply an understanding of basic sciences to clinical conditions" could be cross validated with clinical educator reports of workplace learning assessments. The value of such a process would be twofold: firstly, to investigate what, if any, conclusions can reasonably be drawn from the students' self-reporting about their actual ability; and secondly, to use the findings as a tool for helping students to develop the skills to self-reflect on the meaning and learning opportunities provided by any mismatches between their own and the clinical educator's perceptions. This self-evaluation/feedback/self-evaluation cycle can become a powerful learning strategy.

Analysis 5: Factor analysis

The factor analysis yielded 12 significant factors with eigenvalues greater than 1. Factor 1 explained 27% of the variance and the other factors (2–12) explained between 2.5 and 8% of the variance each. The KMO measure of sampling adequacy was 0.6, suggesting that the responses provided in the sample are barely acceptable but that the analysis can proceed (with caution). The KMO value is understandable in the context of our previous comments on sample size.

Using the rotated component matrix, items loading above 0.5 against one factor only were assigned to this factor. Items 37 (approach confidently senior staff for help in interpreting investigations), 19 (apply an understanding of basic sciences to clinical conditions), 35 (treat the patient as a whole person) and 38 (identify my own educational needs) loaded below 0.5 against their associated factor and should be considered with caution. Overall, item loading against these factors displayed variable levels of targeting when compared to the original PHPQ subscales. For example, Factor 1 included four items (out of six) that were part of the "confidence" scale. "Holistic care" and "self-directed learning" included three items from the original subscale, and "management" and "prevention" included two items from the original subscale. By contrast, the "interpersonal skills", "collaboration" and "science" subscales did not display, within the limitations associated with a small sample size, any more than one item loading on one of the 12 factors.

Discussion

The Rasch analysis suggests that the PHPQ should be considered a set of subscales rather than describing a single trait. In the context of our study, the "interpersonal skills" scale was validated but the "collaboration" subscale was rejected. All other subscales were partially validated, where items were found to be excessively easy to be endorsed by the students and had a significant amount of local dependencies. These results suggest that further work to validate the PHPQ scale may include the removal of some items and the further refinement of others, perhaps using semi-structured interviews

ISSN 1442-1100

to better understand the underlying cognitive constructs of responses. Furthermore, when students or professionals use the PHPQ instrument to report on preparedness for practice, all valid subscales need to be scored independently. An average mean or aggregated score of the entire scale is not representative of the individual's ability.

Limitations

It is noted that our dataset was small and, as a result, could be invalid. This would be accurate if we were designing the scale for the first time, however this was not the situation. We were using a scale already widely published, well used and reputable, and as such, the size of the dataset does not matter, particularly in the context of Rasch analyses. It is further noted that a more holistic understanding of student preparedness for practice could involve consultation with clinical supervisors and employers after graduation to determine the accuracy of student self-assessments.

The factor analysis confirmed that the PHPQ may indeed measure several sub-constructs not necessarily aligned with the original PHPQ constructs and suggested that a reduced set of the original items, together with new items that discriminate further between students (i.e., don't make it so "easy" for the students to endorse the trait) could constitute a more robust version of the original PHPQ. It may be possible that grounding the items further in the discipline (by using words better grounded in the discipline) may in fact be a more appropriate discrimination strategy.

This analysis provides scope for health professional educators, and vocational training providers, to consider the further use of this or similar instruments for self-evaluative and/or outcomes evaluation purposes, however the need to refine the subscales depending on the purpose it will be used for needs to be addressed. Further refinement, as befits the purpose, will assist in enhancing the validity of the tool for reporting on specific subscale traits. We propose that further work cross validate some of the items with traditional assessments in the teaching program as a form of feedback to the students. This would be particularly beneficial when using a self-evaluation instrument in a scaffolded manner across a teaching program in order to address specific professional standards. Validity and item refinement may also be enhanced through the use of focus groups with the intended sample population, to explore interpretation of specific items.

Conclusions

The consideration of the utility of previously validated scales within particular disciplinary contexts needs to be more widely acknowledged and engaged with. This is especially the case when instruments have been in use for more than a decade, as is the case with the PHPQ, despite changes in the learning and teaching strategies displayed by students and teachers, respectively, across this timeframe. This was evidenced in this work by the lack of validation of several subscales of the widely used PHPQ.

ISSN 1442-1100

Health professions and health professions education are both rapidly evolving, and the training environments continue to be dynamic. This situation requires educational institutions to be responsive to the needs of the profession and more thoughtful about how students' sense of preparedness is measured. It can no longer be assumed that instruments validated more than a decade ago will continue to be valid for current training contexts, and so an enhanced understanding of measurement methodology and scale validation is needed to maintain currency within the training environment.

References

Blodgett, T. J., Blodgett, N. P., & Bleza, S. (2016). Simultaneous multiple patient simulation in undergraduate nursing education: A focused literature review. *Clinical Simulation in Nursing*, *12*(8), 346–355.

Blumberg, P. (2000). Evaluating the evidence that problem-based learners are self-directed learners: A review of the literature. In D. H. Evensen & C. E. Hmelo (Eds.), *A research perspective on learning interactions: Problem-based learning* (pp. 199–226). Mahwah, NJ: Lawrence Erlbaum Associates.

Christensen, M., Aubeeluck, A., Fergusson, D., Craft, J., Knight, J., Wirihana, L., & Stupple, E. (2016). Do student nurses experience imposter phenomenon? An international comparison of final year undergraduate nursing students' readiness for registration. *Journal of Advanced Nursing*, *72*(11), 2784–2793. doi:10.1111/jan.13034

Dean, S. J., Barratt, A. L., Hendry, G. D., & Lyon, P. M. (2003). Preparedness for hospital practice among graduates of a problem-based, graduate-entry medical program. *Medical Journal of Australia*, *178*(4), 163–166.

Hadwin, A. F. (1996, April). *Promoting self-regulation: Examining the relationships between problem-based learning in medicine and the strategic content learning approach*. Paper presented at the annual meeting of the American Educational Research Association, New York City, NY.

Hess, D., & Frantz, J. M. (2014). Understanding the learning styles of undergraduate physiotherapy students. *African Journal of Health Professions Education*, *6*(1), 45–47.

Hill, J., Rolfe, I. E., Pearson, S. A., & Heathcote, A. (1998). Do junior doctors feel they are prepared for hospital practice? A study of graduates from traditional and non-traditional medical schools. *Medical Education*, *32*(1), 19–24.

Holen, A. (2000). The PBL group: Self-reflections and feedback for improved learning and growth. *Medical Teacher*, *22*, 485–488.

Holtgraves, T. (2004). Social desirability and self reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, *30*(2), 161–172.

Kassim, S. S., McGowan, Y., McGee, H., & Whitford, D. L. (2016). Prepared to practice? Perception of career preparation and guidance of recent medical graduates at two campuses of a transnational medical school: A cross-sectional study. *BMC Medical Education*, *16*(1), 1.

ISSN 1442-1100

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Liaw, S. Y., Koh, Y., Dawood, R., Kowitlawakul, Y., Zhou, W., & Lau, S. T. (2014). Easing student transition to graduate nurse: A simulated professional learning environment (SIMPLE) for final year student nurses. *Nurse Education Today*, *34*(3), 349–355.

Loyens, S. M. M., Magda, J., & Rikers, R. M. J. P. (2008). Self-directed learning in problem-based learning and its relationships with self-regulated learning. *Educational Psychology Review*, *20*(4), 411–427. doi:10.1007/s10648-008-9082-7

MacCarrick, G., Kelly, C., & Conroy, R. (2010). Preparing for an institutional self-review using the WFME standards: An international medical school case study. *Medical Teacher*, *32*(5), 227–232.

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education*, *14*(4), 595–621.

Muhamad, Z., Henry, L. J., & Ramli, A. (2016). A narrative review of assessment appraisal in clinical performance of physiotherapy students. *International Medical Journal*, *23*(2).

Ngeow, K., & Kong, Y. (2001). Learning to learn: Preparing teachers and students for problem-based learning. *ERIC Digest* [ED 457 524].

Peterson, M. (1997). Skills to enhance problem-based learning. *Medical Education Online*, *2*(1). Retrieved from https://www.tandfonline.com/doi/abs/10.3402/meo.v2i.4289

Physiotherapy Board of Australia & Physiotherapy Board of New Zealand. (2015). *Physiotherapy practice thresholds in Australia and Aotearoa New Zealand*. Retreived August 7, 2017, from https://www.physioboard.org.nz/sites/default/files/PhysiotherapyPractice%20Thresholds3.5.16.pdf

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Expanded ed.). Chicago, IL: University of Chicago Press.

Ryan, G. L. (1997). *The development of problem solving and self-directed learning ability in problem-based learning*. Unpublished doctoral dissertation, University of Sydney.

Scicluna, H. A., Grimm, M. C., Jones, P. D., Pilotto, L. S., & McNeil, H. P. (2014). Improving the transition from medical school to internship: Evaluation of a preparation for internship course. *BMC Medical Education*, *14*(1), 14–23.

Scicluna, H. A., Grimm, M. C., O'Sullivan, A. J., Harris, P., Pilotto, L. S., Jones, P. D., & McNeil, H. P. (2012). Clinical capabilities of graduates of an outcomes-based integrated medical program. *BMC Medical Education*, *12*(1), 12–23.

Skinner, K. L., Hyde, S. J., McPherson, K. B., & Simpson, M. D. (2016). Improving students' interpersonal skills through experiential small group learning. *Journal of Learning Design*, *9*(1), 21–36.

ISSN 1442-1100

Smith, M., & Trede, F. (2013). Reflective practice in the transition phase from university student to novice graduate: Implications for teaching reflective practice. *Higher Education Research and Development*, *32*(4), 632–645.

Sudman, S., Bradburn, N. M., & Schwartz, N. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco, CA: Jossey Bass.

Wijnen, M., Loyens, S. M., Smeets, G., Kroeze, M., & van der Molen, H. (2016). Comparing problem-based learning students to students in a lecture-based curriculum: Learning strategies and the relation with self-study time. *European Journal of Psychology of Education*, *32*(3), 431–447.

Yew, E. H. J., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. *Health Professions Education*, *2*(1), 75–79. doi:10.1016/j.hpe.2016.01.004.

Yew, E. H. J., & Schmidt, H. G. (2009). Evidence for constructive, self-regulatory, and collaborative processes in problem-based learning. *Advances in Health Sciences Education*, *14*, 251–273.

ISSN 1442-1100