This article is downloaded from

http://researchoutput.csu.edu.au

It is the paper published as:

**Author:** Guisard, Y., Hyde, S., Skinner, K., Simpson, M.

**Title:** Evidencing physiotherapy students’ preparedness for practice

**Journal:** Advances in Health Sciences Education

**ISSN:** 1382-4996

**Year:** 2019

**Volume:** 20

**Issue:** 1

**Pages:** 19-35

**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.
Abstract

In health professions, self-evaluation and reflection are necessary professional practice competencies. Significantly, the development of capacity in students to reflect on their preparedness to transition from an undergraduate student into a graduate professional necessitates that Higher Education Institutions scaffold these skills in their programs. To evaluate this, a valid and reliable assessment tool is crucial. This work revisits a recent study in which student preparedness for practice was compared in a traditional and a problem-based learning (PBL) curriculum in an undergraduate physiotherapy degree. A modified version of the Preparedness for Hospital Practice (PHPQ) questionnaire was used as a self-assessment tool in that earlier study. In the current study, we applied a combination of Rasch and Factor Analyses to the original data in order to ascertain the validity of the underlying construct, preparedness; and to assess the effect of PBL on students’ perceived preparedness for practice. Our analyses show that the PHPQ should be considered as a set of subscales rather than an instrument which measures a single construct. Some sub-scale s were found to be valid and evidenced a significant effect of the teaching pedagogy. However, the “collaboration” sub-scale could not be validated and several others were only partially validated and require further refinement. We concluded that although it is a useful instrument, the PHPQ requires further development in order to be used confidently as a self-assessment tool for students to evaluate their preparedness for practice as physiotherapists.
Keywords
PHPQ, Preparedness, Problem Based Learning, PBL, Rasch, self-report

Introduction

The Physiotherapy practice thresholds which describe the threshold competence required for initial and continuing registration as a physiotherapist in both Australia and Aotearoa New Zealand (Physiotherapy Board of Australia & Physiotherapy Board of New Zealand, 2015) identify “Reflective practitioner and self-directed learner” as one of 7 core roles required of registered physiotherapists. As such, students’ perceptions of how well their University course has prepared them for self-directed learning and their ability to reflect and have the capacity for insight about their capabilities is of significant interest to educators and to the profession.

The successful preparation of health professional students for transition into practice is a prime concern for educators and professional registration organisations in Australia and globally to ensure patient safety. Concurrent with this is a requirement grounded in health professional standards (Table 1) for graduates to have sufficient capacity for self-assessment, to recognise their own limitations and reflect on areas in need of improvement. In addition, times of transition between formal, structured learning programs and the clinical context are often stressful for new graduates (Smith and Trede, 2013; Yew and 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.

The traditional model of problem-based learning (PBL) comprises small groups of six to eight students, a tutor whose role is to facilitate discussion rather than to deliver content knowledge, and the use of authentic written or audiovisual cases based upon the problems in which patients present within a variety of health professional contexts such as the emergency room or in the general practice setting (Muhamad et al., 2016). The problem, or trigger, comes first, without advance readings, lectures, or preparation, serving as a stimulus for the need to know (Wijnia et al., 2014). For some hybrid PBL models, learning in PBL tutorials is supplemented by lectures, laboratory sessions, visits to clinical sites, and clinical skills sessions in relation to the ‘case of the week’. In PBL, students collaborate within the group to construct and apply their own knowledge and understanding around the case through a series of tutorials which can vary in frequency from one to three times a week, depending on the model of PBL being used. 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 et al., 2016).

The role of the tutor is especially unique in this type of learning and instruction. In the PBL context, the tutor facilitates and guides learning without contributing directly to the solution of the problem or being the primary source of information (Mennin and Majoor, 2000). The tutor uses questions to explore and stimulate student thinking, helps the group set standards for depth and breadth of knowledge, develop reasoning ability, and enhance communication skills. This collaborative approach to learning and teaching therefore creates many opportunities for self-directed learning and reflection to occur.

A key aspect of PBL is time dedicated within the final tutorial for reviewing the group process and providing peer feedback and self-evaluation. These processes facilitate students' personal and professional development by promoting students' reflections on their actions and the development of strategies for improved performance.

Evidence from a number of PBL studies shows that students learn how to learn and develop skills in academic regulation such as metacognitive knowledge monitoring, self-evaluation, goal setting, and self-directed learning (e.g. Blumberg, 2000; Hadwin, 1996; Ryan, 1997; Yew and Schmidt, 2009; Wijnen et al., 2016). 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 and Kong, 2001; Peterson, 1997). Furthermore, evidence is emerging that reflective thinking is affected by key variables such as: supportive environment (intellectual and emotional), authentic context, mentoring, group discussion, support and free expression of opinions (Mann et al., 2009) all of which are present in the PBL approach.

Self-directed learning and reflection are just two of the many outcomes PBL has been shown to facilitate. Overall preparedness for practice as a course outcome is becoming a common evaluation benchmark which can be used as an indicator of graduate preparedness for the transition from student to early career professional. Health professions are being confronted with new and expanding practice domains and there is a need for ongoing evaluation and curriculum development to enhance the wide-ranging skills students need to be equipped with (Hodgetts et al, 2007).

The Preparation for Hospital Practice Questionnaire (PHPQ) is a tool specifically designed to measure preparedness for practice in evaluating the outcomes of a new course delivery model (Hill et al., 1998). This instrument has previously been validated and used within other health professional degree programs and was especially valued for the currency and stability of items. Although the PHPQ has been widely used in medicine, dentistry, and occupational therapy, there is very little published literature on the use of this instrument in an undergraduate physiotherapy degree program. To investigate the utility of the instrument for our purposes of evaluating course outcomes, and justification of an
integrated PBL curricula, we decided to further validate the instrument using Rasch analysis to strengthen the basis on which to make core curricula decisions.

The Physiotherapy practice thresholds framework (Physiotherapy Board of Australia and Physiotherapy Board of New Zealand, 2015) supports the establishment of additional performance indicators and rating scales for valid measurement of physiotherapists’ competence for different purposes, in different settings and across different chosen fields of practice. We argue that standards requiring deep skills such as self-reflection, ongoing self-evaluation and ongoing critical self-appraisal are difficult to assess using a static assessment (such as an oral appraisal or a written essay) and could be usefully complemented by using relevant and well validated self-evaluation questionnaires on a periodic basis to assist with both learner self-reflection and as a tool to regularly assist in the quality improvement of the course as part of a formal evaluation cycle.

**Aims:** To investigate the validity and reliability of a self-report measure of preparedness for practice in two cohorts of Physiotherapy students after adoption of a curriculum innovation. More broadly, this study aims to contribute towards a curriculum design discourse, suggesting that robust and validated survey instruments are valuable evaluation strategies to assist course designers in identifying gaps in student preparedness, and long-term, to assist with student self-reflection and to inform an ongoing course quality improvement strategy.

**Background of the survey instrument:**

The Preparedness for Hospital Practice Questionnaire (PHPQ) is a survey instrument commonly used to evaluate medical students’ clinical capabilities for internship (Hill et al., 1998). It has forty-one items with eight recognised sub-scales: Interpersonal skills; Confidence and coping; Collaboration; Patient management and practical skills; Understanding science; Prevention; Holistic care, and Self-directed learning (Hill et al., 1998). 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 sub-scales, indicative of high internal consistency or relatedness within each sub-scale.

The PHPQ is widely used and continues to be utilised in the assessment of medical graduates in many nations (Dean et al., 2003; Kassim et al., 2016; MacCarrick et al., 2010; Scicluna et al., 2014; Scicluna et al., 2012), in nursing (Blodgett et al., 2016; Christensen et al., 2016; Liaw et al., 2014) and physiotherapy (Hess and Frantz, 2014; Skinner et al., 2016).
Table 1 Examples of professional standards in 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.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Standard</th>
<th>Sub Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nursing and Midwifery¹</td>
<td>Standard 1: Thinks critically and analyses nursing practice</td>
<td>1.2 develops practice through reflection on experiences, knowledge, actions, feelings and beliefs to identify how these shape practice</td>
</tr>
<tr>
<td>Paramedicine²</td>
<td>3.3. Practices within an approved scope of practice</td>
<td>utilises a range of integrated skills and self-awareness to manage clinical challenges effectively in unfamiliar circumstances or situations</td>
</tr>
<tr>
<td></td>
<td>9.1 Monitors and reviews the ongoing effectiveness of their practice and modifies it accordingly</td>
<td>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 reflects on practice and the application of such reflection to their future practice</td>
</tr>
<tr>
<td></td>
<td>1.4: Maintain and extend professional competence</td>
<td>Enabling Competency 1. Adopt a scope of practice consistent with competence. General Level: Recognises and responds to situations outside own competence.</td>
</tr>
<tr>
<td>Physiotherapy⁴</td>
<td>Role 4. Reflective practitioner and self-directed learner</td>
<td>4.1 assess their practice against relevant professional benchmarks and take action to continually 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</td>
</tr>
</tbody>
</table>

Background of the educational program:

At Charles Sturt University (CSU), physiotherapy has been taught since 1998, initially at Albury campus, taking a traditional pedagogical didactic approach in which students attended lectures for theoretical content and practical classes for clinical skills. 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 at both the Albury and Orange campuses with a transition to an integrated Problem Based Learning (PBL) course, in which key professional practice subjects in every year used a regular, consistent, and scaffolded experiential, small group learning approach to provide opportunities for students to: practice interpersonal and communication skills, reflection and self-evaluation, and receive and respond to feedback; collaborate with their peers in a professional context; and develop skills in self-directed learning.

It would be tempting to assume that exposing students to PBL experiences over the 4 years of their undergraduate degree would result in graduate physiotherapists who are confident in their capacity to become lifelong learners and reflective practitioners. However, as suggested by Lusardi, Levangie and Fein, (2002), there is a need for more research to better understand the impact of PBL on students’ transition to clinical practice and lifelong learning. In order to investigate students’ confidence in their ability to take on the roles required of them, the original authors chose to use a modified version of the PHPQ (Skinner et al, 2016).

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 on the reliability of the questionnaire (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 is “overall” aligned with a single trait (the “preparedness for hospital practice”); or

b) an instrument that is an aggregate of a number of closely related traits (sub-scale s) that “together” align with a broader multidimensional construct.

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), 58 students were sampled in the final week of study with a paper copy of the instrument.

The Rasch model is one of the simplest mathematical formulations 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 “ability” of the person and the “difficulty” of the item. Uniquely, the Rasch model, assumes that the responses provided by persons are influenced by a single, latent trait. The original survey included a “Don’t Know” category. In this study, such responses were coded as missing data. We argue here that “Don’t Know”, as in the PHPQ scale, is conceptually different from the traditional “Not Applicable” category used in discussions relating to missing data in analyses of scales. Rasch analysis is robust to missing data.

Rasch models (Rasch, 1960) rest upon a set of four assumptions:

- **Unidimensionality** (A single, or small number of related traits are assessed in the analysis)
- **Conditional independence** (Over the whole sample, there are correlations between items, but at any given ability level, the responses to the items are not correlated)
- **Sufficiency** (The sum of all scored items by all respondents contains all the information necessary to define the person ability and item difficulty), and
- **Monotonicity** (As the ability of a person increases, so does the probability response. This relationship is not linear but logistic)

Satisfying these assumptions induces a range of properties to the instrument, 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. These assumptions are validated during, rather than before the analysis as is the case with traditional statistical procedures. In this study, the RUMM2030 (RUMM Laboratory Pty Ltd, Version 5.4., Perth, Australia) software was used.

Factor Analysis is a statistical technique 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** (there are at least some correlations amongst the variables so that coherent factors can be identified), and
- **Appropriate sample size** (The sample size should be large enough to yield reliable estimates of correlations among the variables)

In our research, the first three assumptions were respected, but the sampling 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 response patterns (ie consistently responded using a “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 carried out in this study and the rationale for carrying it out.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Nature of the Analysis</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Original full model</td>
<td>Validation of the published PHPQ scale</td>
</tr>
<tr>
<td>2</td>
<td>Re-scored full model and removed extreme persons</td>
<td>Analysis 1 suggested that the original full model displayed disordered thresholds and extreme behaviours</td>
</tr>
<tr>
<td>3</td>
<td>Original sub-scale s</td>
<td>The second analysis displayed multi dimensionality</td>
</tr>
<tr>
<td>4</td>
<td>Re-scored sub-scale s</td>
<td>The third analysis suggested that the original full model displayed disordered thresholds</td>
</tr>
<tr>
<td>5</td>
<td>FA on the original dataset</td>
<td>An investigation of the sub-scale s suggested in this dataset using a methodology close to that published in the original paper</td>
</tr>
</tbody>
</table>

Results and Discussion

Analyses 1 and 2 – Full and re-scored models

The analysis of the original dataset (the entire PHPQ instrument) 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 was good (>0.7). The analysis displayed a large amount of local dependencies (>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 (cohort) factor considered. Items thresholds were however very poor and only 11 of the 41 items displayed ordered thresholds. This suggested that respondents had difficulty identifying specific response categories. For example, they may have had difficult conceptualising the difference between “Very Inadequate” and the adjacent “Inadequate” category.

The analysis of the residuals of the full model provided evidence of multidimensionality. This suggests that we were measuring more than one trait of ‘preparedness’.

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 (assumed to be “preparedness to practice”), 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” (or [1,2]) categories and a smaller number of items required a combination of the “Neutral” and “adequate” (or [2,3]) categories (see Table 4 for a
small sample of rescored items). The rescored full model, displayed similar model fit characteristics as 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 inter-dependent constructs potentially aligned with the sub-scales.

In view of the issue of multi dimensionality and poor targeting, the items were split, and an analysis of the original sub-scale s proposed by Hills et al (1998) was carried out.

Table 3 Summary of the models fit to the data. Cronbach Alphas are not provided here as some of the data was missing. PSI is an approximation of Cronbach 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 sub-scale s. For acceptable values and computations, see text.

<table>
<thead>
<tr>
<th>Original Models</th>
<th>Item</th>
<th>Item Mean</th>
<th>Items Fit Res. SD</th>
<th>Persons Mean</th>
<th>Persons Fit Res. SD</th>
<th>Chi Square pr.</th>
<th>PSI (with Extrm)</th>
<th>Number of Extreme Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>0</td>
<td>0.6</td>
<td>2.6</td>
<td>1.4</td>
<td>0.5</td>
<td>0.9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>0</td>
<td>0.5</td>
<td>1.2</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>0</td>
<td>0.7</td>
<td>2.2</td>
<td>1</td>
<td>0.5</td>
<td>0.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td>0</td>
<td>0.5</td>
<td>3.5</td>
<td>0.6</td>
<td>0.9</td>
<td>0.7</td>
<td>7 (ID 30; 2; 55; 14; 44; 35)</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>0</td>
<td>0.5</td>
<td>2.7</td>
<td>0.9</td>
<td>0.3</td>
<td>0.5</td>
<td>3 (ID 25; 45; 55)</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>0</td>
<td>0.2</td>
<td>2.5</td>
<td>0.9</td>
<td>0.6</td>
<td>0.4</td>
<td>4 (ID 27; 37; 2; 28)</td>
<td></td>
</tr>
<tr>
<td>Prevention</td>
<td>0</td>
<td>0.6</td>
<td>3.5</td>
<td>0.9</td>
<td>0.9</td>
<td>0.6</td>
<td>2 (ID 43; 58)</td>
<td></td>
</tr>
<tr>
<td>Holistic Care</td>
<td>0</td>
<td>0.4</td>
<td>3.2</td>
<td>1.1</td>
<td>0.8</td>
<td>0.8</td>
<td>3 (ID 14; 44 28)</td>
<td></td>
</tr>
<tr>
<td>Self-Directed Learning</td>
<td>0</td>
<td>0.8</td>
<td>3.9</td>
<td>1</td>
<td>0.01</td>
<td>0.7</td>
<td>3 (ID 35; 45; 55)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Re Scaled Models</th>
<th>Item</th>
<th>Item Mean</th>
<th>Items Fit Res. SD</th>
<th>Persons Mean</th>
<th>Persons Fit Res. SD</th>
<th>Chi Square pr.</th>
<th>PSI (with Extrm)</th>
<th>Number of Extreme Persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>0</td>
<td>0.6</td>
<td>2.9</td>
<td>1.5</td>
<td>0.08</td>
<td>0.9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>N/A</td>
<td>Thresholds were appropriate in original analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>0</td>
<td>1</td>
<td>2.3</td>
<td>1.1</td>
<td>0.7</td>
<td>0.6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Collaboration</td>
<td>0</td>
<td>0.8</td>
<td>1.9</td>
<td>1.2</td>
<td>0.5</td>
<td>0.4</td>
<td>3 (ID 25; 45; 55)</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>0</td>
<td>0.6</td>
<td>2.2</td>
<td>1.1</td>
<td>0.9</td>
<td>0.5</td>
<td>5 (ID 2; 27; 28; 37; 57)</td>
<td></td>
</tr>
<tr>
<td>Science</td>
<td>0</td>
<td>0.6</td>
<td>3.6</td>
<td>1</td>
<td>0.9</td>
<td>0.6</td>
<td>2 (ID 43; 58)</td>
<td></td>
</tr>
<tr>
<td>Prevention</td>
<td>0</td>
<td>0.4</td>
<td>3.5</td>
<td>0.8</td>
<td>0.8</td>
<td>0.7</td>
<td>9 (ID 14; 27; 28; 30; 31; 35; 44; 54; 57)</td>
<td></td>
</tr>
<tr>
<td>Holistic Care</td>
<td>0</td>
<td>0.8</td>
<td>4</td>
<td>1.1</td>
<td>0.008</td>
<td>0.8</td>
<td>3 (ID 35; 45; 55)</td>
<td></td>
</tr>
</tbody>
</table>

Analyses 3 and 4 – Original and re-scored sub-scale s

The analysis of the individual original sub-scale s yielded a good data fit to the model for all sub-scale s (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 felt that they displayed a large amount of the measured latent traits in each scale.

The sub-scale “Interpersonal skills” did not need re-scoring. It can therefore be used as proposed. In the context of this study, this sub-scale displayed a positive PBL cohort effect (i.e. PBL students endorsed the items significantly more than non PBL students; Table 5). Conclusions drawn by Skinner et al (2016), that there appear to be benefits for students.
in the areas of interpersonal skill development 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. Both cohorts perceiving themselves to be not quite adequately prepared, with the mean rating falling between neutral and adequate. The implications of these results for employers are that graduands’ development of interpersonal skills may need 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. Support and mentoring in this domain may enhance the transition to work.

The sub-scale “Confidence” also displayed a PBL cohort effect (i.e. PBL students endorsed the items significantly more than non PBL students) but was multi-dimensional, suggesting that the responses to the questions in this domain, at least for the physiotherapy students surveyed, did not on the face of it relate to a single concept previously associated with “confidence”.

None of the other sub-scale s displayed a significant difference between the two cohorts (ie, no factor effect was detected in the other sub-scale s).

All sub-scale s, except “Interpersonal skills”, required a partial to full re-scoring of items. For an example of re-scoring, see Table 4. Similar to the re-scoring process of the full model, the sub-scale s mostly required re-scoring in the “Very inadequate” and “Inadequate” categories. Several items required 2 sets of re-scoring across the 5 categories. Finally, all sub-scale s displayed a significant number of local dependencies (suggesting the removal of 1 or more items).

Table 4 Examples of re-scoring structure of some items. Brackets indicate the collapse of categories into a single one.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Re-scoring structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Model</td>
<td>2. Cope with stress caused by my work</td>
<td>0, [1, 2], 3, 4</td>
</tr>
<tr>
<td></td>
<td>3. Recognise my own clinical limitations</td>
<td>0, [1, 2], 3, 4</td>
</tr>
<tr>
<td></td>
<td>4. Carry out basic surgical procedures</td>
<td>0, 1, [2, 3], 4</td>
</tr>
<tr>
<td></td>
<td>7. Handle most clinical emergencies</td>
<td>0, 1, [2, 3], 4</td>
</tr>
<tr>
<td>Confidence</td>
<td>2. Cope with stress caused by my work</td>
<td>[0, 1], 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>3. Recognise my own clinical limitations</td>
<td>[0, 1], 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>37. Approach confidently senior staff for help in interpreting investigations</td>
<td>0, [1, 2], 3, 4</td>
</tr>
</tbody>
</table>

Following re-scoring of the seven sub-scale s, the sub-scale “Collaboration” did not converge (see 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” sub-scale 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 further refinement of the tool is needed for use in this specific context, before we can
make any conclusions around students’ self-perceptions of their capacity for self-directed learning.

For the other six sub-scales, the data displayed a good Items fit of the model (Chi Squared >0.05). The data displayed a good Person fit but did include some extreme behaviours. It also displayed a strong skew towards positive values for person means, suggesting that the questions included in the subscales may need further refinement with 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 the physiotherapy discipline, by using more contextualised and powerful words, may in fact be a more appropriate discrimination strategy. Furthermore, social desirability is known to positively influence self-reporting (Holtgraves, 2004) in at least 2 of a 5 stage response sequence process (“retrieving information” and “making a judgment”) (Sudman et al., 1996). These skewed responses are well documented (Kruger and 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 found to be unidimensional but displayed a significant number of local dependencies (ie, the response to one item is correlated with the response to another; Table 5). This violates the assumption that in Rasch analysis, 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 within a sub-scale, thus reducing the number of questionnaire items to be more concise.

The mean locations for some items displayed an “agree/disagree” (dichotomous) behaviour, rather than taking advantage of the granularity of a 5 point scale. These behaviours suggest that from a students’ perspective, some questions did not require answering using a paradigm of gradual endorsement of the trait measured in the sub-scale. It may also be reflective of a degree of overconfidence. It may therefore be useful for scales such as the PHPQ to identify items that could be correlated with a secondary assessment 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, 19, 35 and 38 loaded below 0.5 against their associated factor and should therefore be considered with caution. Overall, item loading against these factors displayed variable levels of targeting when compared to the original PHPQ sub-scale s. For example, Factor 1 included 4 items (out of 6) that were part of the “Confidence” scale. “Holistic Care” and “Self-Directed Learning” included 3 items from the original sub-scale, and “Management” and “Prevention” included 2 items from the original sub-scale. By contrast, “Interpersonal Skills”, “Collaboration” and “Science” sub-scale s did not display, within the limitations associated with a small sample size, any more than one item loading on one of the 12 factors.

It may be proposed that our dataset was small and as a result could be invalid. This would be correct 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.
<table>
<thead>
<tr>
<th>Item</th>
<th>Sub-scale</th>
<th>Original Models</th>
<th>Re-scored Models</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>20.</td>
<td>Interpersonal Skills</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>Confidence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>Collaboration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>Management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Prevention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Holistic Care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Self-Directed Learning</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5 Summary of the performance of each item.** 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 sub-scale (✓ 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.
Conclusions

The Rasch analysis carried out in this study suggests that the PHPQ should be considered a set of sub-scale s rather than a single trait. In the context of our study, the “Interpersonal skills” scale was validated but the “Collaboration” sub-scale was rejected. All other sub-scale s were partially validated, where items were found to be excessively easy to be endorsed by the students and 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 to better understand the underlying cognitive constructs of responses.

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 (ie, 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 more contextualised and powerful words) 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, alongside the need to refine the subscales depending on the purpose it will be used for. Further refinement, as befits the purpose, will assist in enhancing the validity of the tool for reporting on specific sub-scale 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.
References


