Studies in Educational Evaluation 56 (2018) 144–153

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

# A Rasch analysis of the Revised Study Process Questionnaire in an Australian osteopathy student cohort

Brett Vaughan[a,b,*]

[a] College of Health & Biomedicine, Victoria University, Melbourne, Australia
[b] Institute of Sport, Exercise and Active Living, Victoria University, Melbourne, Australia

ARTICLE INFO

*Keywords:*
Evaluation
Validity
Reliability
Item response theory
Medical education

ABSTRACT

There are a number of questionnaires designed to capture a students' self-reported learning approach. One example is the Study Process Questionnaire and its modified version, the 2-factor revised Study Process Questionnaire (R-SPQ-2F). Students in the first year of the 2014 and 2015 cohorts in the osteopathy program at (institution) completed the R-SPQ-2F at the commencement of their course. Rasch analysis was employed to evaluate the R-SPQ-2F internal structure. Reliability estimations were calculated for both subscales using Cronbach's *alpha* and McDonald's *omega*. Both the Deep and Surface subscales failed to fit the Rasch model at initial analysis. Modifications included item rescoring, removal of misfitting responses, and removal of one item from both subscales. After modification, both subscales fit the Rasch model and demonstrated unidimensionality. Reliability estimations were acceptable and consistent with previous literature. The study has provided evidence for the internal structure of the R-SPQ-2F in this cohort.

## 1. Introduction

Educational researchers have generally described approaches to learning as *deep* and *surface* (Biggs, Kember, & Leung, 2001; Entwistle & Entwistle, 2003; Kember, Biggs, & Leung, 2004). The deep learning approach is thought to be used by students who wish to develop an understanding of the material presented, and integrate this with their current knowledge base. The use of this approach is likely influenced by factors including, but not limited to, the topic being presented, the understanding of the topic, and desire to engage with the topic at a deeper level (Platow, Mavor, & Grace, 2013).

Previous research has also indicated that utilising a deep approach to learning positively correlates with assessment results (May, Chung, Elliott, & Fisher, 2012; Mok, Dodd, & Whitehill, 2009; Salamonson et al., 2013), grade point average (Liu, Carmen, & Yeung, 2015), reflection and critical reflection engagement (Leung & Kember, 2003), and satisfaction with problem-based learning (Gurpinar, Kulac, Tetik, Akdogan, & Mamakli, 2013). Conversely, the adoption of a surface learning approach and engaging in superficial approaches to learning (Vanthournout, Coertjens, Gijbels, Donche, & Van Petegem, 2013) appears to be related to an expectation of transmission of content from teacher to student (Platow et al., 2013), learning for the purpose of assessment rather than understanding (extrinsic motivation) (Vanthournout et al., 2013), and "…memorizing the facts without any concern to linking or integrating prior and new knowledge or fully understanding underlying mechanisms and principles" (Gurpinar et al., 2013, p. 85). Further, surface learning approaches are negatively correlated with assessment results (Salamonson et al., 2013), influenced by the educational context (Balasooriya, Toohey, & Hughes, 2009) and were found to predominate in early year medical students, but not in the later clinical years (Mirghani, Ezimokhai, Shaban, & van Berkel, 2014).

One tool that has been used to evaluate student deep and surface learning approach profiles is the Study Process Questionnaire (SPQ) (Biggs, 1987). The SPQ was developed to evaluate student learning strategies, and subsequently use this information as part of quality assurance programs, to identify students who may need assistance, and to evaluate innovations in teaching and assessment (Biggs et al., 2001). With regard to the health professions, there are numerous examples (Fox, McManus, & Winder, 2001; Gurpinar et al., 2013; Kusurkar, Croiset, & ten Cate, 2013; Kusurkar, Croiset, Galindo-Garré, & Ten Cate, 2013; Mattick, Dennis, & Bligh, 2004; Platow et al., 2013; Weller et al., 2013) where the SPQ and its more recent incarnation, the revised 2 factor version of the SPQ (R-SPQ-2F) (Biggs et al., 2001; Munshi, Al-Rukban, & Al-Hoqail, 2012; Salamonson et al., 2013), have been used. Students in medicine, nursing, osteopathy, and generic health sciences programs have participated in studies utilising either version. These studies have typically demonstrated there is little difference between

* Correspondence address: College of Health and Biomedicine, Victoria University, PO Box 14428, Melbourne, VIC 8001, Australia.
*E-mail address:* brett.vaughan@vu.edu.au.

https://doi.org/10.1016/j.stueduc.2017.12.003
Received 31 July 2017; Received in revised form 11 November 2017; Accepted 9 December 2017
0191-491X/ 

professions for surface learning approaches however deep learning approaches do vary (Salamonson et al., 2013).

A number of authors have explored the psychometric properties of the R-SPQ-2F (Justicia, Pichardo, Cano, Berbén, & De la Fuente, 2008; Socha & Sigler, 2014). There is little agreement in the literature as to the internal structure of the R-SPQ-2F beyond the presence of the two first order factors, Deep and Surface learning strategy. Both Justicia et al. (2008) and Socha and Sigler (2014) have provided extensive reviews of the properties of the R-SPQ-2F, and also evaluated its factor structure using classical test theory approaches. Readers interested in the classical test theory approaches to evaluating the internal structure of the R-SPQ-2F are encouraged to seek out the work by these authors.

To provide further evidence in support of the validity of the interpretation of the scores derived from R-SPQ-2F, a Rasch analysis (RA) can be performed. RA is an increasingly popular method for investigating the construct validity and dimensionality of questionnaires in the health sciences (Hagquist, Bruce, & Gustavsson, 2009; Retief, Potgieter, & Lutz, 2013; Tennant & Conaghan, 2007) and there are numerous examples of the application of the Rasch model in health professions education (Bourke, Wallace, Greskamp, & Tormoehlen, 2015; Dalton, Davidson, & Keating, 2011; Hecimovich, Styles, & Volet, 2014; Rodger et al., 2013; Tor & Steketee, 2011). Lamoureux et al. (2007) summarise the Rasch model as "… the probability of a given response affirming an item is a logistic function of the relative distance between the item's location and respondent's location on a linear scale." The location of the item or respondent is measured in log odds units called 'logits' (Meads & Bentall, 2008; Rodger et al., 2013). When compared to item response theory (IRT) models, Rasch analysis allows for the evaluation of the summated scale/raw score as a sufficient statistic (Andrich & Hagquist, 2012; da Rocha, Chachamovich, de Almeida Fleck, & Tennant, 2013; Hagquist & Andrich, 2004), that is, "…no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter (p. 310)" (Fisher, 1922). Further, the Rasch model is reported to have a number of advantages over classical test theory approaches to scale development. These include: conversion of ordinal data to interval data when the data fits the Rasch model (da Rocha et al., 2013; Wright & Mok, 2004); scale development is independent of the test sample used during development (Engelhard, 2013); and the capacity to locate both the items and persons responding to the scale on the same linear scale, representing the underlying or latent construct being measured (Chang & Engelhard, 2015). The purpose of the present study was to evaluate the internal structure of the R-SPQ-2F in an Australian osteopathy student population using Rasch analysis to provide evidence to support its use in the evaluation strategy for our teaching program.

## 2. Method

This study was approved by the Victoria University Human Research Ethics Committee. Participant responses were anonymous. Consent to participate in the study was implied upon return of the questionnaire, and participation was not required as part of any subject in the teaching program.

### 2.1. Participants

Students enrolled in the 2014 and 2015 cohorts for year 1 of the Bachelor of Science degree in the osteopathy program at Victoria University (VU, Australia) were invited to participate in a larger project evaluating the teaching, learning and assessment practices. Students were invited to complete a number of questionnaires in week 1 of the 1st teaching semester as part of their first practical skills class. The responses to these questionnaires contributed data to a larger project to evaluate the teaching, learning and assessment practices in the VU osteopathy program.

### 2.2. Measure

Participants completed the R-SPQ-2F and two demographic questions (age & gender). The R-SPQ-2F was developed by Biggs et al. (2001) based on the original version of the SPQ, and consists of 20 items spread across 2 first order factors (deep, surface) and 4 second order factors (deep motive, deep strategy, surface motive, surface strategy). Each item is rated on a five-point Likert scale (1 = Never or only rarely true for me to 5 = Always or almost always true for me). These authors have reported confirmatory factor analysis (CFA) statistics supporting the factor structure of the R-SPQ-2F, however the reliability estimations (Cronbach's alpha) for all but the *surface motive* subscale were below 0.70 (Biggs et al., 2001). Other authors have demonstrated higher alpha scores (> 0.80) for the deep and surface factors only (Immekus & Imbrie, 2009; Liu et al., 2015; Socha & Sigler, 2014) without dividing these into the second order factors. These mixed results suggest the reliability estimations for the R-SPQ-2F are likely context dependent.

### 2.3. Data analysis

Data were entered into Microsoft Excel. Descriptive statistics were generated in the *R* program (R Core Team, 2014) using the *psych* package (version 1.5.4) (Revelle, 2014).

#### 2.3.1. Rasch analysis

To perform the Rasch analysis, data were exported into RUMM2030 (Andrich, Sheridan, & Luo, 2010). The Deep and Surface subscales were analysed independently as the calculation of a total score for the R-SPQ-2F is not recommended (Justicia et al., 2008; Socha & Sigler, 2014). The following description articulates how the analysis was performed on each subscale using RUMM2030, and the criteria against which decisions were made.

#### 2.3.2. Overall Rasch model fit

Overall model fit provides a global picture as to how well the data fit the Rasch model. Overall model fit is evaluated using the chi-square statistic and Bonferonni-adjusted p-value. Non-significant chi-square results suggest overall fit to the Rasch model (Pallant & Tennant, 2007). A significant chi-square value indicates there is a discrepancy between the observed and expected values (da Rocha et al., 2013), and the data are measuring multiple latent constructs (Tor & Steketee, 2011). Fit residual standard deviations (SD) are used to evaluate the fit of the items and persons respectively (Pallant & Tennant, 2007). The fit residuals represent standardised differences between the observed responses and the responses expected by the Rasch model. Perfect fit of the persons and items respectively to the Rasch model is indicated by a standardised normal distribution mean of 0 and a standard deviation of 1 (Pallant & Tennant, 2007). Fit residual SD's outside of a range of ± 1.5 suggests that items or persons do not fit the Rasch model. Further, the Person Separation Index (PSI) provides an indication of the ability of the items to spread out the test sample (Lamoureux, Pallant, Pesudovs, Hassell, & Keeffe, 2006), and is interpreted in the same manner as Cronbach's alpha (Tor & Steketee, 2011). A related statistic is the number of strata or statistically different groups that this spread allows for the identification of (Wright & Masters, 2002). A PSI of 0.67–0.80 allows for the identification of 2–3 strata and a PSI of 0.81–0.90 allows for 3–4 strata to be identified (Fisher, 2007).

Overall fit to the Rasch model was indicated by a Bonferonni-adjusted non-statistically significant chi-square statistic. RUMM2030 calculates the fit residual standard deviation (SD) for the items and persons, and a fit to the Rasch model was demonstrated by a fit residual SD of ± 1.5. Internal consistency is reported as the Person Separation Index (PSI) and was interpreted in the same manner as other internal consistency statistics such as Cronbach's alpha (Tor & Steketee, 2011). Further analyses in RUMM2030 can identify items or persons that do