0416

Measuring student flow in the classroom

Penelope Kennish, Robert Cavanagh
Curtin University of Technology, Perth, Western Australia, Australia

Background

Csikszentmihalyi (1990) reported that when people described optimal experiences (situations which are highly enjoyable), they often used the term ‘flow'. According to Csikszentmihalyi and Schneider (2000, p. 97), flow refers to the "... spontaneous, seemingly effortless aspect of such experiences". A recurring aspect of flow experience descriptions is the balance between perceived high levels of challenge and high levels of skill - the task is demanding but the enjoyment of the experience also derives from having the skills necessary to complete the task (Massimini, Csikszentmihalyi and Carli, 1988).

Author, Author and Author (2008) proposed a two-dimensional model of student engagement in learning in which engagement at a given time and in a particular context was seen a balance between a student's capability to learn (skills) and the expectations of their learning (challenge). The model was empirically tested by Rasch Rating Scale model analyses of data comprising researcher ratings of aspects of student engagement (see Author, 2009). These analyses confirmed the two-dimensional flow-based engagement theory.

Concurrently, a12-item student self-report instrument, the Course Experience Survey (CES), was developed to provide criterion-related evidence of validity. The CES was based on the Shernoff, Csikszentmihalyi, Schneider and Shernoff (2003) construct of student flow (concentration, interest and enjoyment). This paper describe how data from the CES were collected an analysed.

Research Questions

The research sought to construct a parsimonious linear measure of flow experiences. Attaining this outcome required data fit the Rasch model. Specifically:

1. Does data from the CES fit the Rasch Rating Scale Model?

2. What evidence of validity is available from the Rasch analysis?

Methods

The CES is a 12-item rating scale instrument and students select from a five-category response scale scored 1 (never) to 5 (always). The sample was 195 Years Eight to Twelve students from Western Australian metropolitan, rural and remote schools. It included females and males from the subject areas of English, Mathematics, Science and Social Studies. Demographic student data were also collected on gender, year of schooling, subject being reported, and whether the subject was a favourite subject. Data were entered into RUMM2020 (Andrich, Sheridan, Lyne and Luo, 2003) and analysed with the Rasch Rating Scale model.

Frame

Contemporary fields of psychology and educational psychology recognise and emphasise the importance of connectedness, happiness, and indeed flow for personal well-being. The positive psychology movement has emerged as a viable alternative to psychologists and disciplinary psychology which centre on "... negative pathological features and accounts of human functioning" (Martin, 2006, p. 307). Using Flow Theory to understand student experiences and the learning environment provides perspectives on the learning process not possible with more traditional approaches.

The quantitative investigations of flow have typically applied Classical Test Theory and correlational analytic techniques. There is a paucity of objective measurement procedures and criteria in studies of flow, particularly in classroom research about student flow. There is a need to use measurement approaches in which scores are plotted on an interval scale, missing data can be handled, test properties are sample free, different scales can be placed on a common metric, and short targeted scales with high reliability are developed. The fit of CES data to the Rasch Rating Scale model shows the CES meets these requirements and is an objective measure.

Data fit to the Rasch model was assessed using five RUMM2020 of analyses/displays. These exemplified three aspects of validity evidence and instrument development activities that Wolf and Smith (2007a & 2007b) identified as requirements for validity arguments:

(a) Content aspect;

(b) Substantive aspect; and

(c) Generalisability aspect.

Research findings

(a) The content aspect of validity evidence can be assessed by estimation of statistics to show how well the distribution of observed values fits with values predicted by a measurement model. RUMM2020 estimates residuals for each item - the difference between the raw score and the score predicted by the Rasch Rating Scale model. RUMM2020 generates an Item Characteristic Curve (ICC). The observed scores for class intervals are then plotted over the ICC. The logits on the horizontal axis are the logarithmic odds of Item 5 being affirmed by the respondents. When the observed scores for three class intervals of students were plotted, these were close to the respective values predicted by the model. The fit residual was 2.03 ($<\pm 2.5$) due to the observed scores being close to the predicted scores. All twelve items had fit residuals $<\pm 2.5$.

Insert Figure 1 here

Figure 1. Item Characteristic Curve for Item 5

(b) The substantive aspect of validity concerns observed consistencies among item responses being explained by the theory or hypotheses informing an investigation. Previous research on classroom flow and engagement identified higher levels of flow when students described their favourite subjects (Shernoff, 2001; Shernoff, Csikszentmihalyi, Schneider and Shernoff, 2003). RUMM2020 generates a Person Frequency Distribution that plots the distribution of calibrated person scores for different groups of persons. In Figure 2, the respective frequencies of flow scores for students reporting on their favourite and non-favourite subjects are plotted. The mean score for students reporting on their favourite subject was 1.56 logits in comparison to a mean score of 0.35 logits for the others. The difference was statistically significant ($F=66.3$, $p<0.005$). This concurrence between theory-based expectations and flow scores is evidence of the substantive aspect of validity.

Insert Figure 2 here

Figure 2. Person Frequency Distribution

(c) "The generalisability aspect of validity addresses the degree to which measures maintain their meaning across measurement contexts" (Wolfe and Smith, 2007b, p. 215). For example, in terms of group independence, an item for which the success rate does not differ between females and males of similar ability. The Item Characteristic Curve presented in Figure 3 shows the observed scores for three class intervals of girls and boys for Item 2.

Insert Figure 3 here

Figure 3. Item Characteristic Curve for Item 2

The differences in observed scores between the genders for the three class intervals of students were not statistically significant ($F = 0.01$, $p > 0.05$). DIF due to gender was not revealed for any of the other items.