Paper Presented at the AARE Annual Conference, Canberra 2009

CAV091057

Quantifying student engagement in classroom learning: Student learning capabilities and the expectations of their learning

Rob Cavanagh (PhD) and Penelope Kennish (BA Hons)
Curtin University of Technology, Perth, Western Australia

Paper submitted to the Assessment and Measurement Special Interest Group at the 2009 Annual Conference of the Australian Association for Research in Education: Canberra.

Abstract

This paper reports on a quantitative investigation of secondary school student engagement in classroom learning. The empirical investigation was informed by a theoretical model in which *student engagement in classroom learning* was viewed as a function of *student capability for learning* and the *expectations placed on their learning*. The investigation was part of the second phase in a large scale ARC Linkage project. The first phase focussed on epistemological issues in defining *student engagement in classroom learning*. The second phase used interviews to collect qualitative and quantitative data from a representative sample of Western Australian secondary school students. The third phase currently underway is applying a self-report rating scale instrument to collect data from a large number of students to enable analysis of interactions between engagement variables.

In Phase Two, a researcher-completed rating scale instrument was developed using operational definitions of constructs identified in Phase One. The instrument was administered by two researchers to rate 195 secondary students on aspects of their engagement in classroom learning. Data on five aspects of *student capability for learning* and six aspects of *expectations of student learning* were tested separately against the Rasch Rating Scale Model. For this study, data on these two constructs were analysed separately. This was notwithstanding that a conjoint analysis was possible and that this procedure has been used to investigate the engagement of particular groups of students (e.g. LOTE students). Testing data-to-model fit using RUMM20220 estimated or generated summary test-of-fit statistics, category probability curves/threshold locations, individual item fit statistics, item characteristic curves/differential item functioning, person-item threshold distributions, and Varimax location loadings from factor analysis of residuals.

The rationales for the RUMM2020 estimates are explained and the results are used to confirm the compliance of data with the requirements for objective measurement that underpin the Rasch Rating Scale Model. In this way, the resulting measures are considered to manifest the latent trait of *student engagement in classroom learning*.

Address correspondence to:

Associate Professor Rob Cavanagh,
School of Education,
Curtin University of Technology
GPO Box U1987
Western Australia 6845
Email: R.Cavanagh@curtin.edu.au
Phone: +61 8 9266 2162
Fax: +61 8 9266 2547

The research was conducted as part of an Australian Research Council funded Linkage Project between Curtin University of Technology and the Participation Directorate of the Western Australian Department of Education and Training.

Quantifying student engagement in classroom learning: Student learning capabilities and the expectations of their learning

Introduction

The preliminary phase in an investigation of the engagement of secondary school students in Western Australia advanced a view of classroom engagement based on extant literature and intentions to measure the phenomenon (Cavanagh, Kennish & Sturgess, 2008). The key constructs and sub-constructs in an hypothesised model of *student engagement in classroom learning* were then investigated empirically. This report describes the instrumentation used to quantify this phenomenon including the analytic procedures applied to the data. The final section presents the results of the analyses for each of the key constructs. Data were analysed using the Rasch Rating Scale Model (Andrich, 1978a, b & c).

Theory

Cavanagh, Kennish and Sturgess (2008) reviewed the literature on engagement, engagement with schooling and classroom engagement. They proposed that the theorising about classroom engagement in learning could be advanced by consideration of Flow Theory about experiential conditions (see Csikszentmihalyi 1990a &1990b; Hekter, Schmidt & Csikszentmihalyi, 2007; Massimini, Csikszentmihalyi & Carli, 1987). Investigations into school engagement and participation have utilised Flow Theory (see Parr, Montgomery & deBell, 1998; Shernoff, Csikszentmihalyi, Schneider & Shernoff, 2003; Schweinle, Meyer & Turner, 2006). In Flow Theory, the Experience Fluctuation Model (Csikszentmihalyi & LeFevre, 1989) uses measures of a person's skill level and the challenge they face in a particular situation to plot experiential conditions on a Cartesian plane. Cavanagh et al. (2008, p. 7) applied this model to the notion of engagement and proposed that "…students who are engaged within a particular situation will have a balance between the perceived level of the challenge being faced and their perceived capability to meet the incumbent requirements". In cognisance of the nature of classrooms, teaching and learning, classroom challenges were defined as *expectations of learning* and capability/skills as *learning capabilities*. A student highly engaged in a particular activity was anticipated to perceive high expectations of his/her learning in conjunction with perceptions of high capability. Thus it is postulated that engagement in learning is a function of *learning capabilities* and *expectations of learning*. This can be represented diagrammatically - see Figure One below. It is proposed that measures of *learning capabilities* and *expectations of learning* can be plotted on a Cartesian plane to locate coordinates that indicate *engagement in learning*. For example, the balance that characterises engagement is shown by: high capabilities in conjunction with high expectations; moderate capabilities in conjunction with moderate expectations; or low capabilities in conjunction with low expectations.

In Figure 1, the space where there is a balance between *learning capabilities* and *expectations of learning* is labelled the *Zone of Engagement in Learning*. The amount of learning associated with engagement is expected to vary with more learning happening when the student has high *learning capabilities* in conjunction with high *expectations of learning*. Alternatively, when *expectations of learning* are low and *learning capabilities* are low, less learning is anticipated.

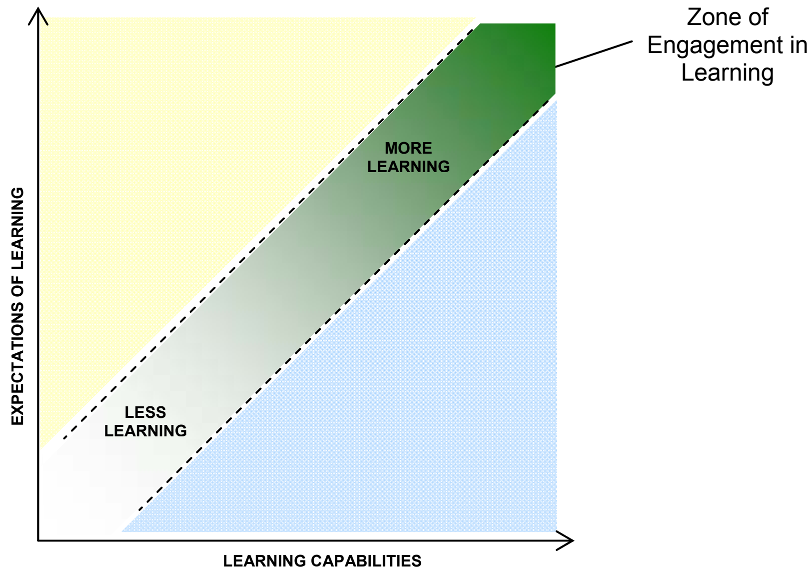

Figure 1. Hypothesised model of student engagement in classroom learning

The construct of *learning capabilities* was hypothesised to comprise two broad attributes of students - the *expressive self* as typified in studies of *self-esteem* and *self-concept* and the *managerial self* as typified in studies of *self-regulation* and *self-efficacy* (see Martin, 2007). Additionally, *resilience* was included in the taxonomy due to the prevalence of this notion in the research on school engagement and participation. The conceptual structure of *learning capabilities* is shown in Table 1 below.

Table 1.
Construct of learning capabilities and hypothesised sub-constructs

	Learning capabilities				
	Expressive self			Managerial self	
Sub-constructs:	Self-esteem	Self-concept	Resilience	Self-regulation	Self-efficacy

The construct of *expectations of student learning* was viewed in terms of learning for understanding. The framework used was the Wiggins and McTighe (2001) six facets of understanding (see Table 2 below).

Table 2.
Construct of expectations of learning for understanding and hypothesised sub-constructs

	Expectations of learning					
Sub-constructs:	Explanation	Interpretation	Application	Perspective	Empathy	Self-knowledge

Research objective and questions

The objective was to develop instrumentation to measure two constructs proposed to constitute *student engagement in classroom learning*. The research questions were:

1. Can a linear scale that measures student *learning capabilities* be constructed?
2. Can a linear scale that measures *expectations of student learning* be constructed?
3. Do measures of student *learning capabilities* and *expectations of student learning* conjointly indicate the higher order construct of student *engagement in classroom learning*?

Methodology

Instrumentation

The two theoretical frameworks presented in Tables 1 and 2 were elaborated by writing five hierarchically ordered proficiency levels for each of the eleven sub-constructs. These described how 'more' of the sub-construct would be seen in a student and also how 'less' of the sub-construct would be seen. Table 3 presents the five descriptors written for *self-esteem*. The proficiency levels for all eleven sub-constructs are presented in Appendices One and Two.

Table 3.
Behavioural descriptors for self-esteem

Level	**Descriptor**
5 High	*Has positive self image*
	Sees very little in self that needs to improve
	Is highly confident
4	*Has confidence to make decisions*
	Is confident to make choices about how to do things
	Is confident to make choices about what to do
3	*Has pride in self*
	Is proud of his/her achievements
	Thinks he/she is good compared to others
2	*Trusts self to act*
	Trust self to do what is right for self
	Has faith in own ability
1 Low	*Sees worth in self*
	Is happy with self
	Sees some good qualities in self

Data on the constructs and sub-constructs were obtained by interviewing students. The interview schedule comprised one question for each sub-construct. The students were asked questions about aspects of themselves which related to their learning and they were also asked about the expectations of their learning.

The interviews were of approximately 30 minutes duration and were conducted by two researchers. Students were asked to report on either their favourite subject or a non-favourite subject. The interviewers then applied the frameworks presented in Appendices One and Two to assign a numerical rating for each of the sub-constructs. Students were rated zero to five with the zero rating being assigned when there was minimal evidence of the sub-construct. Post-interviewing moderation was applied to gain consensus on the scores that were recorded for each student.

Sample

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 Society and Environment (see Table 4).

Table 4.
Sample characteristics

Student variables	**Count**	**Percentage**
Gender		
Females	102	52
Males	93	48
Year of schooling		
Yr 8	21	11
Yr 9	32	16
Yr 10	54	28
Yr 11	58	30
Yr 12	30	15
Subject reported		
English	53	27
Maths	58	30
S&E	41	21
Science	43	22
School location		
Metropolitan	57	29
Country	128	66
Remote	10	5
Total sample	195	100

Data analysis

The data available for analysis were scores from zero to five for each the eleven sub-constructs in *learning capabilities* and *expectations of learning for understanding*. Additionally, student membership of a year cohort, student gender, subject being reported, and whether or not this was a favourite subject of the student were coded and recorded. These categorical person data and the scores were entered into RUMM2020 (Andrich, Sheridan, Lyne & Luo, 2003) and several analyses were performed.

RUMM was used to conduct independent analyses of data on *learning capabilities* and *expectations of learning for understanding* to examine the psychometric properties of data from the two scales. Separate analyses were performed because the theoretical model presented in Figure 1 has two axes – one for each construct.

In the two analyses, RUMM2020 was used to generate:

(a) Summary test-of-fit statistics;
(b) Category probability curves and threshold locations;
(c) Individual item fit statistics;
(d) Item characteristic curves - differential item functioning;
(e) Person-item threshold distributions; and
(f) Varimax location loadings - factor analysis of residuals.

The reasons for conducting the six estimations are summarised in Table 5 below.

Table 5.
RUMM estimations and respective applications

Estimation	**Application**
(a) Summary test-of-fit statistics	Testing whether the data for persons and items fit the measurement model
(b) Category probability curves and threshold locations	Testing whether the data fit the hypothesised order of the behavioural proficiency levels for each sub-construct
(c) Individual item fit statistics	Testing data-to-model fit for individual sub-constructs
(d) Item characteristic curves - differential item functioning	Testing whether the item functions differently for particular groups of persons – e.g. boys and girls
(e) Person-item threshold distributions	Comparing distributions of person scores and item difficulties
(f) Varimax location loadings - factor analysis of residuals	Testing whether the data are uni-dimensional

Results

1. The *learning capabilities* scale

(a) RUMM2020 summary test-of-fit statistics

The summary test-of-fit statistics are presented in Table 6. The item-student interaction indicates the degree to which the data fit the measurement model and are based on residuals. Residuals are the differences between actual and expected responses as calculated from the measurement model parameters. When the data fit the model, the distribution of residuals, the ‘fit statistic’, has a mean near zero and a standard deviation near 1.0. A negative fit statistic indicates that the data fit the model very closely. A positive fit statistic indicates that some ‘noise’ is present. The fit statistic means of 0.33 with SD = 1.39 for the item data and -0.29 with SD = 0.99 for the student data indicate an acceptable global fit to the measurement model.

The item-trait interaction indicates the consistency of the sub-construct ‘difficulties’ across the range of different student *learning capabilities* measures on the scale: That is, how well the raters agreed on the ‘difficulties’ of the items along the scale. When the data fit the model, the item-trait interaction (a chi square) has a probability more than 0.05. However, chi square is sensitive to sample size and is not to be taken too literally for large samples. Notwithstanding this sensitivity, for these data, the chi square was 0.04 ($p < 0.05$) indicating the scale was possibly not measuring a uni-dimensional trait.

Table 6.
RUMM summary test-of-fit statistics - learning capabilities scale

```
                        ITEM-PERSON INTERACTION
====================================================================
                         ITEMS                            PERSONS
                 Location  Fit Residual      Location  Fit Residual
--------------------------------------------------------------------
Mean               0.00       0.33              1.40      -0.29
SD                 0.35       1.39              1.74       0.99
====================================================================
        ITEM-TRAIT INTERACTION           RELIABILITY INDICES
--------------------------------------------------------------------
Total Item Chi Squ          18.7       Separation Index 0.89
Total Deg of Freedom        10.0
Total Chi Squ Prob          0.04
====================================================================
        POWER OF TEST-OF-FIT
--------------------------------------------------------------------
Power is EXCELLENT
[Based on SepIndex of 0.89]
--------------------------------------------------------------------
```

RUMM estimates two reliability indices, Cronbach's Alpha and, the Separation Index (RUMMLab, 2004). Cronbach's Alpha, the proportion of observed variance considered true, should be close to 1.0. For this analysis, Cronbach's Alpha could not be estimated due to missing data. The Separation Index indicates the degree to which the calibrated scores of students (person location logits) are spread across a continuum. For an ideal distribution across this continuum this index will be close to 1.0. For these data, the Separation Index was 0.89 because the students with the higher learning capabilities obtained higher scores on items and those with lower learning capabilities tended to obtain lower scores on items. The power of the test-of-fit was excellent.

Conclusion: The data fits the Rasch Rating Scale model well showing that a measure has been constructed.

(b) Category Probability Curves

RUMM generated category probability curves to show the probability of students being assigned a particular rating as a function of their calibrated score (the location on the horizontal axis). The category probability curve for Sub-construct 3 – *resilience* is displayed in Figure 2.

For proficiency level 0 (minimal evidence of *resilience*), the probability is 0.7 for students with location -5.0 logits and this progressively decreases to a probability of zero for students with location -1.0 logits. Whereas for 1, the highest probability is for students with location -3.0 logits progressively decreasing to a probability of zero for students with location -0.4 logits. The coordinate on the person location scale for the intersection of pairs of curves is the threshold. For example, the proficiency level 0 and proficiency level 1 curves intersect at -4.07 logits which is the threshold between proficiency level 0 and proficiency level 1. The proficiency level 1 and proficiency level 2 curves intersect at -1.73 logits. These two threshold locations are ordered because the 0:1 threshold location is lower than the 1:2 threshold location.

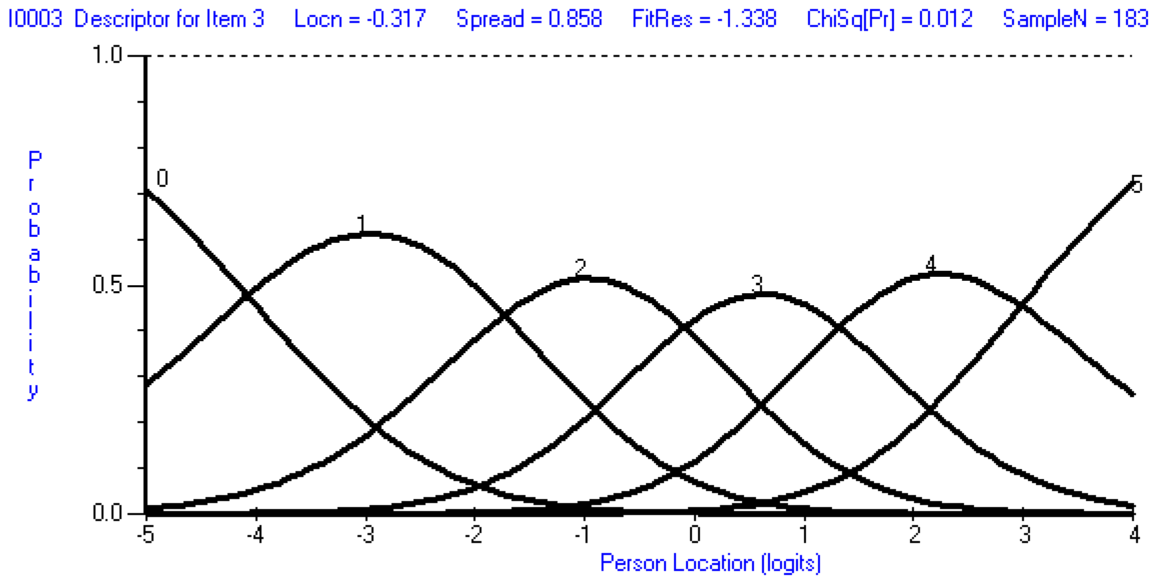

Figure 2. Category probability curve for *resilience*

The threshold locations (un-centralised) of the ratings for each sub-construct are summarised in Table 7 below. For example, the location at which students have an equal probability of being assigned a rating of '0' or '1' for *resilience* was -4.07 logits; the location with equal probability of being assigned a '1' or '2' is -1.73 logits. The threshold location logits are ordered from low to high because the students who were assigned higher ratings for the sub-construct were also students with the higher scores/locations.

Table 7.
Un-centralised threshold locations - learning capabilities scale

Thresholds	**1**	**2**	**3**	**4**	**5**
Sub-constructs	Location logits				
1. Self-esteem	-4.35	-0.40	0.41	0.84	3.63
2. Self concept	-1.78	-0.23	0.22	0.98	3.50
3. Resilience	-4.07	-1.73	-0.09	1.32	2.98
4. Self regulation	-3.25	-0.96	0.22	1.23	3.00
5. Self efficacy	-2.85	-0.94	0.08	0.73	1.51

Conclusion: Data on the respective proficiency levels for the five sub-constructs of *learning capabilities* were hierarchically ordered in line with increasing student levels of *learning capabilities*.

(c) Individual item fit statistics

RUMM2020 tests the fit of data from individual items to the measurement model. The fit of data is shown by the difference between the score predicted by the model and the actual score - a residual.

The data fit the model better when the residual is small. RUMM2020 identifies low residuals by assigning a default value of ±2.5. Another index for the test of fit is the chi square with the probability showing how well the data fit the model. RUMM2020 applies the Bonferroni adjustment to calculate an adjusted probability for each item and identifies probabilities below these values. The fit statistics for data on the five *learning capabilities* sub-constructs are presented in Table 8. The fit residuals and Chi Square probability values show the data for individual sub-constructs fit the measurement model well.

Table 8.
Individual item fit statistics - learning capabilities scale

Sub-constructs	Fit Residual	Chi square	Degrees of Freedom	Probability
1. Self-esteem	2.33	3.21	2	0.20
2. Self concept	0.91	1.29	2	0.52
3. Resilience	-1.34	8.82	2	0.01
4. Self regulation	0.10	3.18	2	0.20
5. Self efficacy	-0.37	2.21	2	0.33

Conclusion: The individual item fit statistics for each of the *learning capabilities* sub-constructs show the data from each sub-construct strongly contributes to measuring *learning capabilities*.

(d) Differential item functioning (DIF)

Tennant and Pallant (2007, p. 1082) noted: “If an item measures the same ability in the same way across groups then, except for random variations, the same success rate should be found irrespective of the nature of the group”. Items display differential item functioning (DIF) when the success rate differs for two or more groups – e.g. between males and females. DIF can be obvious from differences in scores between groups but can also be masked. For example, consider when high ability males have a higher success rate on an item than higher ability females but lower ability males have a lower success rate than lower ability females. In this case, the success rate for all the males and the success rate for all the females could be similar.

RUMM2020 can display DIF by generating an Item Characteristic Curve that plots the expected value for an item against the person location for particular categories of persons (see Figure 3).

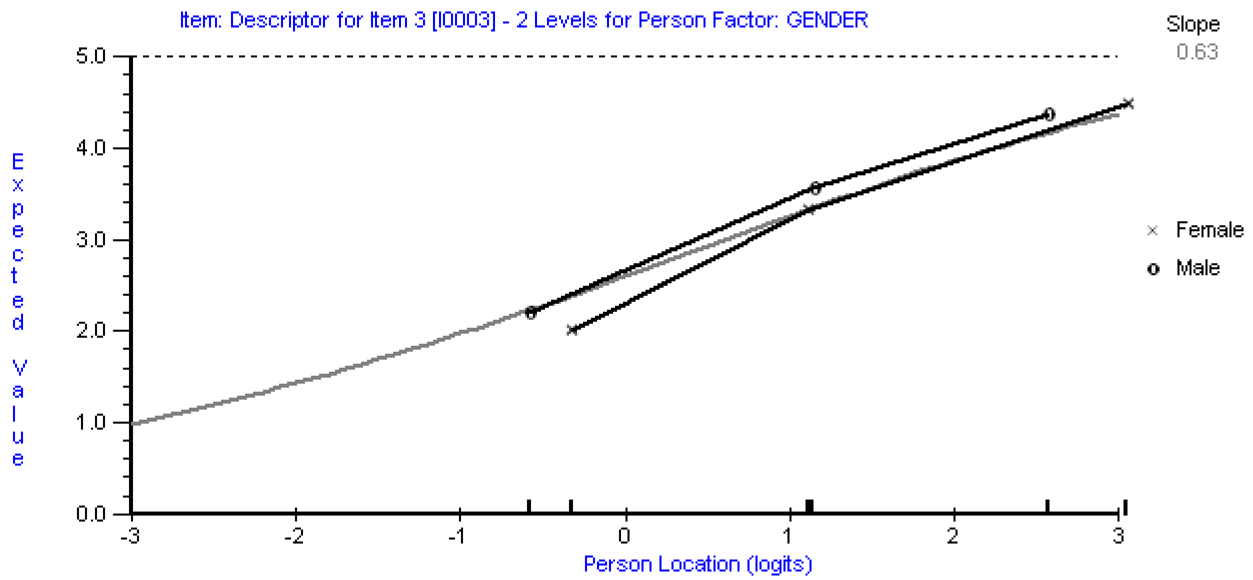

Figure 3. Item characteristic curve for *resilience* – person factor gender.

In Figure 3, the expected values for *resilience* are plotted against the student location logits for females and for males of the same learning capabilities score. The gender plot shows the coordinates for three class intervals of each gender. While the observed scores for the females are higher than expected and the males are lower than expected for all three class intervals, the differences may or may not be significant. RUMM2020 uses a one-way analysis of variance (ANOVA) to test whether membership of the different groups accounts for significant variance in the scores. For the *resilience* data, the difference between females and males was not significant (DIF [Gender]: F= 3.47, p>0.05). RUMM also improves the precision of the probability value using the Bonferroni adjustment. When the Bonferroni adjustment was made, the F-Ratios and probabilities for all five learning capability sub-constructs did not reveal DIF due to student gender. DIF due to the students' year of schooling was also examined but was not found for any of the five sub-constructs.

Conclusion: The success rate of boys compared with girls with the same level of learning capabilities, and between students from different year cohorts with the same level of learning capabilities did not differ in data from the five sub-constructs.

(e) Person-item threshold distribution

The person-item threshold distribution plots person location logits and item thresholds on the same scale. The distribution of calibrated student scores for learning capabilities and the five thresholds in the sub-construct data are presented in Figure 4. The distribution of thresholds (bottom plot) matches the distribution of student scores (top plot) quite well, although 22 of the students had high scores (more than four logits above the zero point), the highest item threshold location was not this high. The scale could be improved by refining the instrumentation to make the higher order proficiency levels more difficult to affirm.

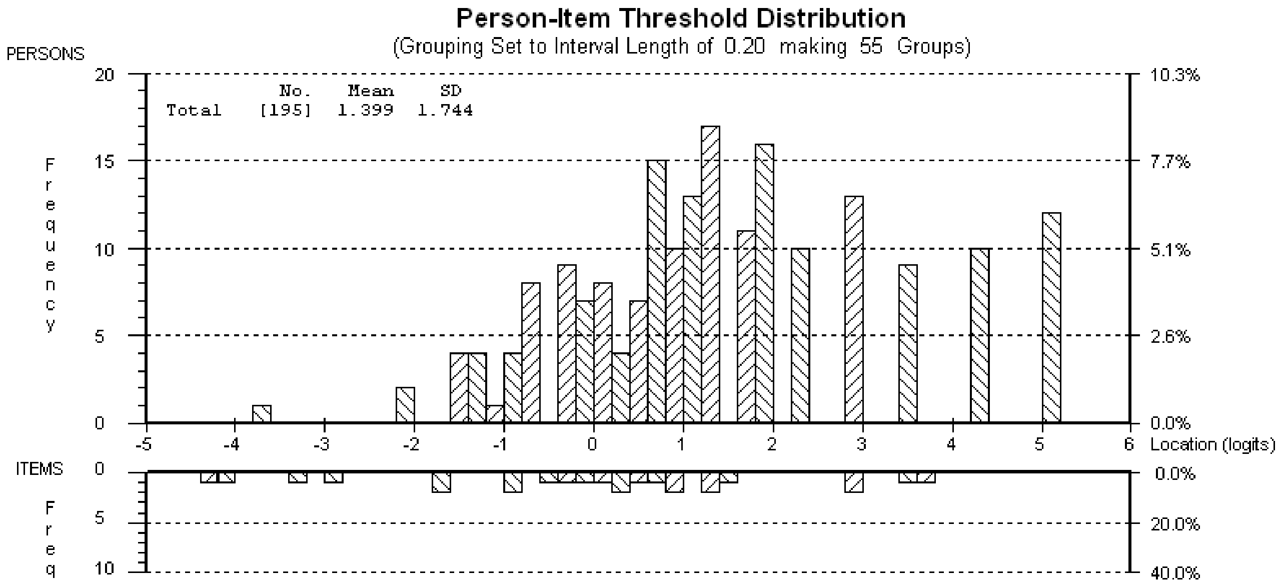

Figure 4. Person-item threshold distribution for *learning capabilities*

Conclusion: The items (respective sub-constructs proficiency levels) elicited data commensurate with the student levels of ability/scores for *learning capabilities*.

(f) Varimax Location Loadings - factor analysis of residuals

After the linear Rasch measure has been extracted from the data set, the residuals that remain might or might not have common variance. That is they could be noise or alternatively they could be structured due to the data measuring a construct different from the primary measure. A factor analysis of the residuals will reveal the presence or absence of such structure and in turn, inform judgements about the uni-dimensionality of the data. The factor loadings after Varimax rotation are presented in Table 9. The strong loadings (bold type for >0.30) for the various sub-constructs are distributed across different factors with no multiple loadings on a single factor.

Table 9.

Varimax rotation loadings from principal components factor analysis of residuals - learning capabilities scale

	Fac1	Fac2	Fac3	Fac4	Fac5
1. Self-esteem	**0.51**	**0.57**	-0.43	-0.42	0.24
2. Self concept	-0.95	0.17	-0.11	-0.24	-0.01
3. Resilience	0.08	0.09	**0.99**	-0.07	-0.01
4. Self regulation	0.18	-0.97	-0.13	-0.07	0.00
5. Self efficacy	0.19	0.03	-0.07	**0.98**	-0.02

Conclusion: There is minimal evidence of the presence of a second measure that could be extracted from the data set.

2. The *expectations of learning for understanding* scale

(a) Summary test-of-fit statistics

The summary test-of-fit statistics are presented in Table 10. The item fit residual standard deviation of 0.57 is low due to a lack of variance in the item data. The person fit residual means and item and person standard deviations are close to those for an ideal distribution. The item-trait interaction Chi Square probability value of 0.78 suggests a uni-dimensional trait is being measured. The separation index of 0.84 is high and the power of test-of-fit is good. These fit statistics generally indicate the data fits the Rasch Rating Scale Model well showing that a measure has been constructed.

Table 10.
RUMM summary test-of-fit statistics - expectations of learning scale

```
                        ITEM-PERSON INTERACTION
====================================================================
                         ITEMS                            PERSONS
                Location  Fit Residual      Location  Fit Residual
--------------------------------------------------------------------
Mean              0.00       0.30              0.52      -0.40
SD                0.36       0.57              1.61       1.17
====================================================================
        ITEM-TRAIT INTERACTION            RELIABILITY INDICES
--------------------------------------------------------------------
Total Item Chi Squ           8.03      Separation Index 0.84
Total Deg of Freedom        12.00
Total Chi Squ Prob           0.78
====================================================================
        POWER OF TEST-OF-FIT
--------------------------------------------------------------------
Power is GOOD
[Based on SepIndex of 0.84]
--------------------------------------------------------------------
```

(b) Threshold locations

The un-centralised threshold locations for the proficiency levels and sub-constructs comprising *expectations of learning* are presented in Table 11. The thresholds are all ordered because data on the respective proficiency levels for the six sub-constructs of *expectations of learning* were hierarchically ordered in line with increasing student levels of *expectations of learning*.

Table 11.
Un-centralised threshold locations - expectations of learning scale

Thresholds	1	2	3	4	5
Sub-constructs	Location logits				
1. Explanation	-2.75	-1.70	-0.12	1.97	4.58
2. Interpretation	-5.61	-2.67	-0.39	1.68	4.01
3. Application	-4.01	-1.45	0.03	1.49	3.96
4. Perspective	-2.28	-1.21	-0.27	1.10	3.45
5. Empathy	-1.61	-1.06	-0.19	1.14	3.06
6. Self knowledge	-3.05	-1.29	-0.25	0.81	2.63

(c) Individual item fit statistics

The fit statistics for data on the six expectations of learning sub-constructs are presented in Table 12. The data from individual items fits the model well - the residuals are low and the chi square probability values are >0.05. The data from the six sub-constructs contributes to the measure of *expectations of learning*.

Table 12.

Individual item fit statistics - expectations of learning scale

Sub-constructs	Fit Residual	Chi square	Degrees of Freedom	Probability
1. Explanation	0.40	0.55	2	0.76
2. Interpretation	-0.63	1.52	2	0.47
3. Application	0.98	0.72	2	0.70
4. Perspective	-0.07	0.58	2	0.75
5. Empathy	0.45	3.96	2	0.14
6. Self knowledge	0.66	0.70	2	0.71

(d) Differential item functioning (DIF)

With the Bonferroni adjustment to the ANOVA, neither student gender nor membership of a particular year cohort accounted for statistically significant differences in the variance of data for any of the six sub-constructs comprising *expectations of learning*.

(e) Person-item threshold distribution

The distribution of calibrated student scores for *expectations of learning* and the thresholds of the six sub-constructs are presented in Figure 5 below. The distribution of thresholds matches the distribution of student scores quite well although three of the students had scores more than five logits above the zero point. The items (respective sub-constructs proficiency levels) elicited a range of data commensurate with the students' levels of ability/scores for *expectations of learning*.

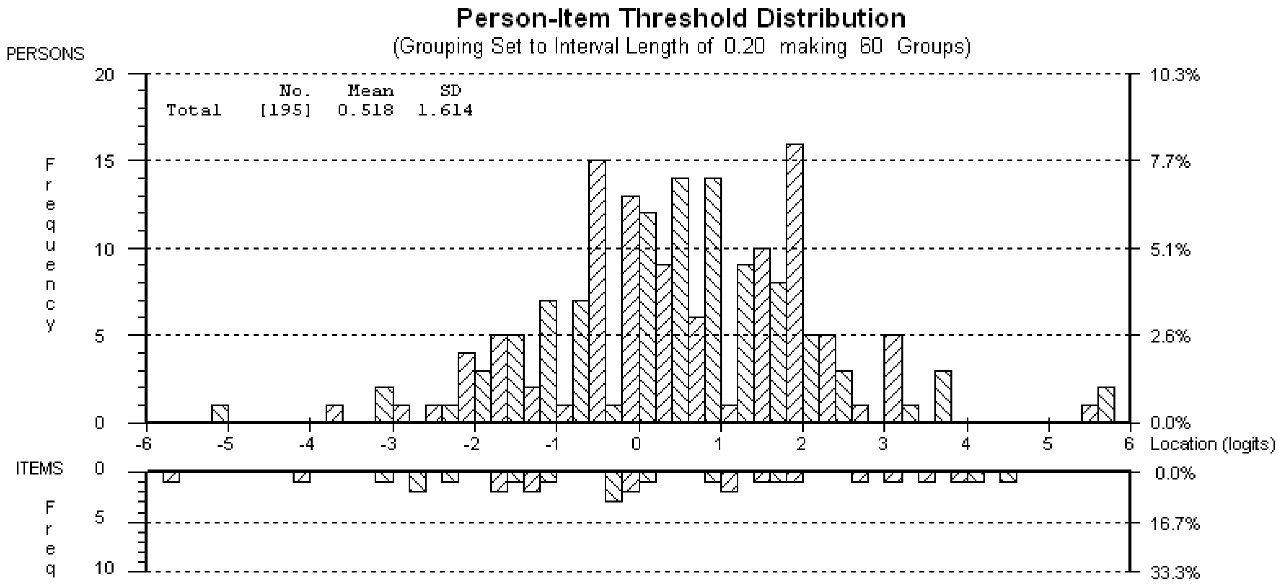

Figure 5. Person-item threshold distribution for *expectations of learning*

(f) Varimax Location Loadings - factor analysis of residuals

The factor loadings after Varimax rotation are presented in Table 13.

Table 13.

Varimax rotation loadings from principal components factor analysis of residuals – expectations of learning scale

	Fac1	Fac2	Fac3	Fac4	Fac5	Fac6
1. Explanation	0.08	0.16	**1.00**	-0.07	-0.03	0.00
2. Interpretation	0.02	0.06	-0.07	**1.00**	-0.15	0.00
3. Application	-0.97	0.14	-0.12	-0.05	-0.29	0.00
4. Perspective	0.23	-0.15	-0.02	-0.15	**0.96**	0.00
5. Empathy	0.15	-0.97	-0.21	-0.10	0.12	0.00
6. Self knowledge	**0.53**	**0.57**	-0.42	-0.42	-0.33	0.00

The strong loadings (bold type) for the various sub-constructs are distributed across different factors with no multiple loading on a single factor. There is minimal evidence of the presence of a second measure that could be extracted from the data set.

3. The relation between *learning capabilities* and *expectations of learning* – a plot of *engagement*

The scores of each student for *learning capabilities* and *expectations of learning* along with the respective standard errors were used to produce a scatter-plot (see Figure 5). A common person equating procedure was applied with the standard errors from the Rasch analyses estimating the

95% confidence band for these data (see Bond & Fox, 2007, pp. 96-99). The distribution of coordinates shows a direct relation between the two constructs consistent with the Figure 1 construct model of engagement.

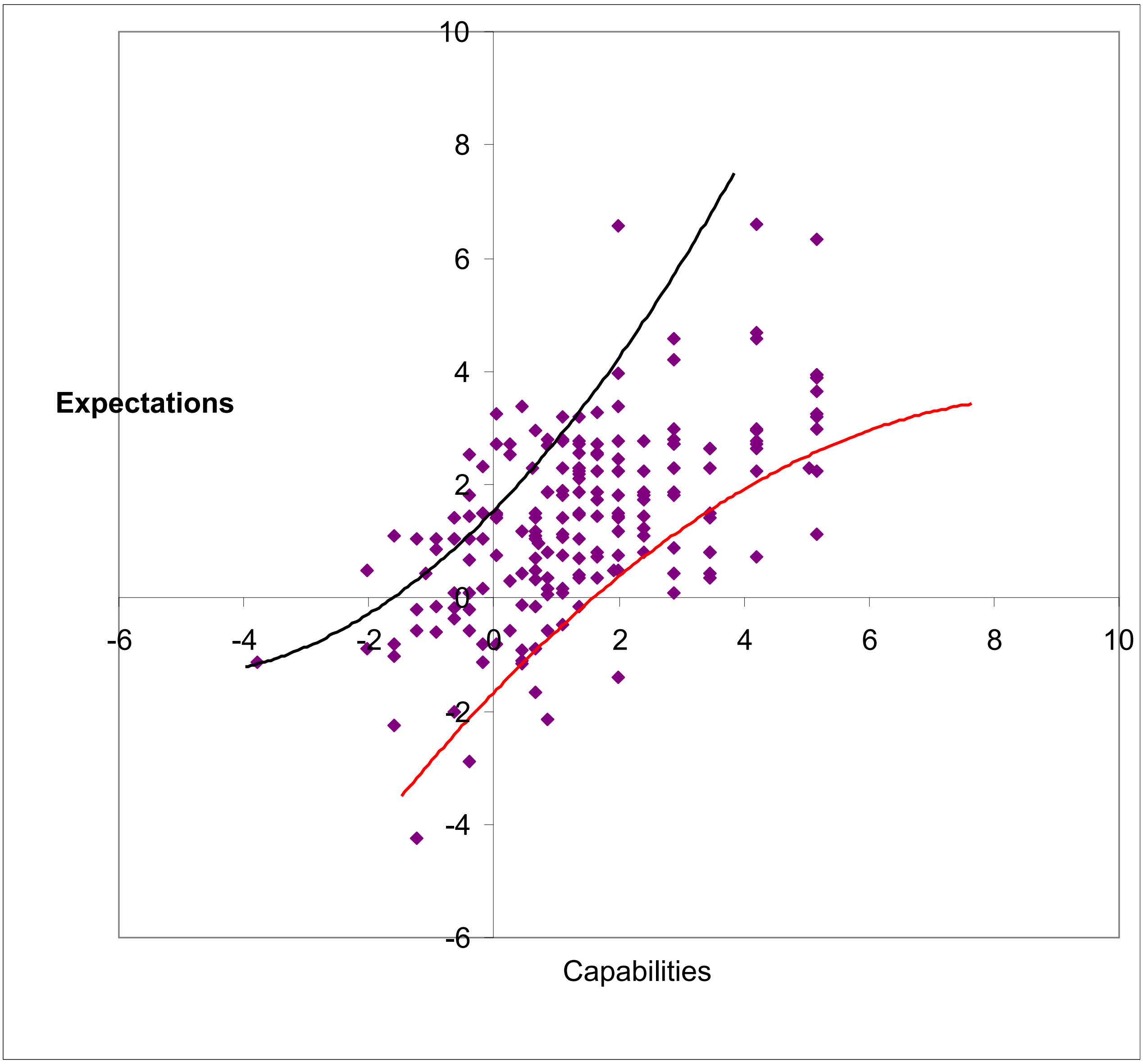

Figure 5. Scatter-plot of *learning capabilities* versus *expectations of learning*

Conclusion

Two measures were created and data from each conformed to the requirements of the Rasch Rating Scale Model (Andrich, 1978a, b & c). The measures elicited interval data on two constructs proposed to constitute *student engagement in classroom learning*. The construction of the two scales and the testing of data from these two scales will advance empirical quantitatively-based understanding of student engagement.

The time required to administer a researcher-completed instrument to individual students could be reduced by using a student self-report instrument. The theoretical frameworks used in developing the researcher-completed instrument could be easily applied in construction of a self-report instrument.

This report has focussed on technical issues with scale development and testing. However, the results of the various tests do convey significant information about the phenomenon of *engagement in classroom learning* and the students who were studied. A more detailed examination and discussion of the meaning of the scale and data are warranted. This will be the subject of other reports.

References

Andrich, D., Sheridan, B., Lyne, A., & Luo, G. (2003). *RUMM: a windows-based item analysis program employing Rasch unidimensional measurement models*. Perth: Murdoch University.

Andrich, D. (1978a). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594.

Andrich, D. (1978b). Rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Andrich, D. (1978c). Scaling attitude items constructed and scores in the Likert tradition. *Educational and Psychological Measurement, 38*(3), 665-680.

Cavanagh, R.F., Kennish, P., & Sturgess, K. (2008) *Development of a theoretical framework to inform measurement of secondary school student engagement with learning*. Paper presented at the 2008 Annual Conference of the Australian Association for Research in Education: Brisbane.

Csikszentmihalyi, M. (1990a). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.

Csikszentmihalyi, M. (1990b). Literacy and intrinsic motivation. *Daedalus, 119*(2), 115-140.

Csikszentmihalyi, M., & LeFevre, J. (1989). Optimal experience in work and leisure. *Journal of Personality and Social Psychology, 56*(5), 815-822.

Hektner, J.M., Schmidt, J.A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life.* Thousand Oaks, CA: Sage Publications.

Martin, J. (2007). The selves of educational psychology: Conceptions, contexts, and critical considerations. *Educational Psychologist, 42*(2), 79-89.

Massimini, F., Csikszentmihalyi, M., & Carli, M. (1987). The monitoring of optimal experience. A tool for psychiatric rehabilitation. *Journal of Nervous and Mental Disease, 175*(9), 545-549.

Parr, D.G., Montgomery, M., & deBell, C. (1998). Flow theory as a model for enhancing student resilience. *Professional School Counselling; 1*(5), 26-31.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: MESA Press.

Schweinle, A., Meyer, D.K., & Turner, J.C. (2006). Striking the right balance: students' motivation and affect in elementary mathematics. *Journal of Educational Research, 95*(5), 271-293.

Shernoff, D.J., Csikszentmihalyi, M., Schneider, B., & Shernoff, E.S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158-176.

Tennant, A., & Pallant, J.F. (2007). DIF matters: A practical approach to test if differential item functioning makes a difference. *Rasch Measurement Transactions,* 20(4), 1082-1084.

Appendices

Appendix One: Framework of *leaning capabilities*

Learning capabilities	Expressive self		Managerial self		
	Self-esteem	**Self-concept**	**Resilience**	**Self-regulation**	**Self-efficacy**
5 High The student:	*Has positive self image* Sees very little in self that needs to improve Is highly confident	*Strives to be perfect* Even though he/she knows does very well, still looks for ways to improve Knows self very well	*Has unqualified expectations of coping* Expects will always be OK Doesn't face any unfixable problems	*Takes responsibility for learning* Is in control of own learning Is very competent	*Has perseverance in face of adversity* Keeps trying when things go seriously wrong Never gives up
4 The student:	*Has confidence to make decisions* Is confident to make choices about how to do things Is confident to make choices about what to do	*Motivated by self reflection* Thinking about self makes he/she feel good Thinking about self helps he/she do better	*Can deal with failure* Things going wrong is not an issue for him/her Believes things will eventually work out well	*Makes improvement in own learning* Is continuously improving Changes how he/she learns	*Has determination* Believes can overcome most difficulties Expect to succeed in difficult situations
3 The student:	*Has pride in self* Is proud of his/her achievements Thinks he/she good compared to others	*Self reflecting* What he/she does shapes his/her view of myself Thins about yself when he/she need to	*Expects success* Expects if he/she works at problems they will be solved Expects to eventually succeed	*Understands own learning* Knows how to learn better Knows how he/she learns best	*Recognises contextual influences* Recognises some situations present more difficulty than others Knows when and where he/she can succeed
2 The student:	*Trusts self to act* Trusts self to do what is right for self Has faith in own ability	*At ease comparing self with others* How he/she feels about self comes from how others see he/she I am comfortable comparing myself with others	*Overcomes small setbacks* Considers overcoming small problems is possible Can deal with small hassles	*Assesses own learning* Learns from mistakes Builds on what he/she can do well	*Has expectations of self* Needs to be successful Doesn't give up easily
1 Low The student:	*Sees worth in self* Is happy with self Sees some good qualities in self	*Compares self with others* Compares self with others Checks own progress against that of others	*Is aware of problems* Accepts a little difficulty is OK Is aware that things go wrong sometimes	*Thinks about learning* Thinks about own learning Questions self	*Has goal orientation* Hope to achieve his/her goals Sets goals for self that are achievable

Appendix Two: Framework of *expectations of learning for understanding*

Expectations	Explanation	Interpretation	Application	Perspective	Empathy	Self-knowledge
5 High The student is expected to:	*Sophisticated* Bring together many ideas to explain something in a new way Develop original (new) explanations of what was taught	*Profound* Show a deep and very clear understanding of the of the work Find simple explanations for complicated things	*Masterful* Find new ways to use knowledge and skills Be flexible in how knowledge and skills are used	*Insightful* Make sure own feelings don't cloud judgements Carefully and fairly evaluate the views of others	*Mature* Be willing to see things the way others do Seek out views highly different from my own	*Wise* Make serious decisions based on knowing what he/she has learnt Make serious decisions based on knowing what he/she has understood
4 The student is expected to:	*In-depth* Understand the work in a way that is different from what was taught Find connections between different parts of what was learnt	*Revealing* Compare different ways of understanding the work Explain the differences between ways of understanding the work	*Skilled* Use knowledge and skills to perform well in a range of situations Use knowledge and skills to perform well in different situation	*Thorough* Be critical of the views of others in a fair way Balance own views against the views of others	*Sensitive* See things in ways similar to others Develop attitudes similar to others	*Circumspect* Have a clear understanding of both strengths and weaknesses Clearly see the strengths and weaknesses of others
3 The student is expected to:	*Developed* Include a range of own ideas when explaining what was learnt Explain what was learnt using own words	*Perceptive* Correctly explain to others how work should be done Help others understand why what is being learnt is important	*Able* Use skills to perform well in some situations Use knowledge to perform well in some situations	*Considered* Understand the views of others Think carefully about the views of others	*Aware* Know that others feel differently from self Be aware that others see things differently from self	*Thoughtful* Identify what he/she doesn't understand Spend time thinking about what he/she can and can't do
2 The student is expected to:	*Intuitive* Explain what was learnt by including extra information Include some of own ideas when explaining what was learnt	*Interpreted* Show correct understanding of the work Explain the importance of what was learnt	*Apprentice* Use the same ways of doing things in different situations Use routines that help get jobs done	*Aware* Show awareness of differences in what others value Reconsider own point of view after listening to others	*Developing* Force self to make sense of ideas that seem strange to him/her Discipline self to understand attitudes different to own	*Unreflective* Accept that others can help he/she see what he/she need to know Let others tell him/her what he/she needs to know
1 Low The student is expected to:	*Naive* Use the words of others when explaining things Use the ideas of others when explaining things	*Literal* Repeat what he/she has been told Repeat what he/she has read	*Novice* Use what was learnt with help from others Follow instructions to complete tasks	*Uncritical* Not ignore points of view different from own Use own views to be critical of things or people	*Egocentric* Try to make sense of ideas that seem strange to him/her Try to understand attitudes different to own	*Innocent* Think about what he/she knows Be aware of things he/she should know