Associations between the Classroom Learning Environment and Student Engagement in Learning 2: A Structural Equation Modelling Approach

Allen G. Harbaugh & Robert F. Cavanagh
Murdoch University (Perth) & Curtin University (Perth)

Abstract

This report is about the second of two phases in an investigation into associations between student engagement in classroom learning and the classroom-learning environment. Whereas the first phase utilized Rasch modelling (Cavanagh, 2012), this report uses latent variable modelling to explore the data. The investigations in both phases of this study employed a novel model of engagement in classroom learning based on flow theory and bio-ecological frameworks. An 85-item survey item was used to collect data from 1760 secondary-school students. Comparable to the findings of the first phase, there was strong evidence for the psychometric properties of the instrument measuring the latent constructs of student engagement in classroom learning and the characteristics of the classroom-learning environment. Furthermore, classroom-learning environment characteristics had direct effects on students' self-esteem and had direct and indirect effects on students' expectations of the classroom environment. Classroom characteristics directly influencing students' self-esteem included the educational values, learning outcomes, classroom learning and parental support. Classroom characteristics directly influencing students' expectations included the educational values, learning outcomes, classroom learning, support from fellow students and expectations of the teacher. Among other benefits, the two phases of this study enable comparison of contemporary analytic approaches: Rasch and Structural Equation Modelling.

Introduction

Student engagement in learning is an important concept to be considered when studying various aspects of the learning environment. This report summarises the findings of the second of two phases in this research study. (For a more detailed exploration of the theory and data collection methods described below, please see the report from the first phase of the study: *Associations between the Classroom Learning Environment and Student Engagement in Learning 1: A Rasch Model Approach* [Cavanagh, 2012].)

This study was part of a large scale Australian Research Council project to investigate the participation and engagement of Western Australia secondary students. Two research questions are examined here. First, does the survey instrument employed in the first phase of this study demonstrate strong psychometric properties for latent constructs of student engagement in classroom learning and the characteristics of the classroom-learning environment? Second, what relationships exist between student engagement and classroom environment?

Theoretical Framework

Student engagement is best understood as a multifaceted construct (Fredricks, Blumenfeld & Paris, 2004). In this analysis, self-esteem, actions attributed to positive self-esteem, and overall expectations of the classroom will be examined as proxy measures for sub-facets of the broader concept of student engagement. Intellectual development and academic performance can be viewed through a bio-ecological framework (Ceci, Rosenbaulum, DeBruyn & Lee, 1997). Such models posit that *proximal processes* describe reciprocal influences of a person with the environment (including setting, other individuals, tools and language). While a bio-ecological framework accounts for more of the environmental interactions, another theory addressing the reciprocal interaction of the person with personally contextualized elements of the environment is Flow theory (Csikszentmihalyi, 1990).

This model suggests that the reciprocal relationship between person and environment is optimal when the situation (environmental factor) poses a challenge at a sufficiently high level to match the person's skill level.

Combining these two theories, Cavanagh, Kennish and Sturgess (2008) proposed a model of student engagement in classroom learning. Parallel to the concepts of skill and challenge in Flow theory, the model suggests that perceived learning capabilities (the skill) and expectations about learning in the given environment (the challenge) interact in a comparable manner to create an optimal experience (engagement) for students. (See the report for the first phase of this study for more detailed information about this model [Cavanagh, 2012].) Based on this model, Cavanagh and Waugh (2004) created a survey instrument to measure the learning environment constructs (e.g., learning outcomes, attitudes of students, teachers and parents). Kennish & Cavanagh (2011) created a survey to measure the student engagement constructs. As the validity and reliability of these scales had already been established, they were chosen for this research.

One of the key features considered in the choice of these measurement scales was the requirement for a uni-dimensional scale of measurement. First, this allowed for an item response theory (IRT) or Rasch model analysis. Second, it aligned well with the Flow theory concept that skill and challenge need to be measured on the same interval-level scale. With IRT, this analysis was possible. (See Cavanagh [2012] for a detailed analysis of this aspect of the survey instrument.)

While the concept of student engagement may be measured on a uni-dimensional interval scale that in turn can be used to measure perceptions of the classroom environment, this does not forego the possibility of underlying latent constructs within the response patterns to the survey instrument. As there have been numerous surveys used to measure classroom characteristics with factor analysis techniques (see Cavanagh [2012] for a detailed list), it seemed appropriate to examine the results of this survey in a similar manner. While the underlying theory of the data generation process is different between Rasch modelling & SEM, it is not impossible for both analyses to provide insight into the nature of the relationships among the data.

Methods

Building from previously established measurement scales for learning environment (Cavanagh & Waugh, 2004) and student engagement (Kennish & Cavanagh, 2011), an 85-item survey was created. The student engagement section had two intended subscales: learning capabilities (12 items) and expectations of learning (15 items). The section measuring student perceptions of the classroom learning environment had eight intended subscales: educational values (5 items), learning outcomes (9 items), classroom learning attitudes and behaviours (11 items), classroom and peer support (11 items), classroom discussion (5 items), classroom planning (3 items), expectations and support from teacher (9 items) and parental involvement (5 items). (The complete survey is available in the appendix of the report of the first phase of this study [Cavanagh, 2012].) Responses were marked on a 3-point positively-packed Likert-type scale (strongly agree, agree and disagree). Support for such a measure has been validated when responses are more often expected to be positive than not (Bond & Hwang, 1986; Lam & Klockars, 1982).

Survey responses were obtained from 1760 secondary school students. This represented 17 schools in the Perth metropolitan region (1323 surveys) and 6 rural/remote schools (437 surveys). Gender indicated for 802 boys (46%) and 951 girls (54%). Regarding the subject matter references in the survey, 354 (20%) indicated it was their favourite subject and 1046 (80%) indicated otherwise. The school years represented were 384 Year-8 students (22%), 348 Year-9 (20%), 535 Year-10 (30%) and 489 Year-11 (28%). Nearly comparable proportions were obtained for the different subject areas: 437 in Mathematics, 451 in English, 434 in Science and 438 in Society & Environment. The stratified sampling process resulted in sample demographics proportional to those of the sampling frame. Thus, the sample is taken as representative of the population of Western Australia. Relationships between participant demographics and responses were examined in detail in the first phase of this study. Though minor variations among the measurement models were detected for some of the subgroups, the overall conclusions remained the same and are not elaborated here.

Space constraints also prohibit an exploration of possible moderation effects via demographic variables for the final model.

Data were primarily analysed using the sem-package (Fox, Nie, & Byrnes, 2012) in R (R Development Core Team, 2012). As needed, additional analyses were conducted using AMOS (Arbuckle, 2009). The choice to examine the data with SEM was motivated by past research with the survey instruments used to measure these constructs. While the IRT focus is on a uni-dimensional interval scale, it is still possible that the items measured against such a scale can demonstrate multidimensional clustering. Such a finding does not call into question the value of the uni-dimensional scale, but it does suggest that multiple latent constructs might be concurrently measured against the same scale. The detection and validation of these psychometric constructs permits the analysis of relationships between the latent variables.

Results

Survey items were measured on a 3-point Likert-type scale with a dominant positive valence. Thus, this raises concerns about using standard SEM analyses as the resulting distributions for the manifest variables can clearly not be multivariate normal. Consequently, analyses were assessed using both regular maximum likelihood estimation (MLE) and robust estimation. As might be expected for a sample size as large as this, results were comparable for all analyses. For ease of presentation, only MLE results are presented below.

All analyses reported are the result of split-half testing. Model adjustments were made on random subsets of the complete data set; adjusted models were tested for goodness-of-fit with the unused subset of the data. In all instances, adjustments were supported by this protocol. Consequently, all measures of model fit are reported for the complete data set only.

The data set did contain a substantial amount of missing data (19% of respondents were missing one or more responses). One of the advantageous of Rasch analysis is its ability to obtain estimations despite incomplete responses. While SEM is not as well suited for dealing with missing data, AMOS provides a rigorous estimation method for missing data. Under the assumption that data are missing completely at random (MCAR), this method has been shown to be quite robust (Byrne, 2010). All analyses were confirmed to be comparable when using the subset of complete responses as when using an imputed complete data set. Thus, results reported are based on the analysis of the completely non-missing data subset (after list-wise deletion for the analysis under consideration).

Measurement Models

The survey instrument employed for this research had 85 items intending to measure student expectations, self-esteem and classroom characteristics. Based on the theoretical support for the construction of the survey, nominal groupings naturally emerged for numerous subsets of the items (two subscales for engagement and eight subscales for classroom characteristics). With this as the grounding for each model, item sets were tested for the presence of predicted latent constructs. In most cases, 5 or more items were expected to measure an intended construct. This allowed flexibility in obtaining measures in that items demonstrating poor fit could be removed while retaining a sufficient number of items to measure the intended construct with appropriate psychometric qualities. Thus, the protocol for model adjustment was to remove items that demonstrated poor fit (weak loadings, cross loadings or appearance in multiple modification indices). Items were considered for removal until there were only three items measuring the construct. With the theoretical model as the underlying support for the *a priori* groupings, no items were considered for cross-loading or transition of the loading to a different construct. Prior to this step in the model adjustment, principal component analysis was run to assess the number of latent constructs. In most cases, a single construct was indicated. In a few instances, there was support for separating the items into 2 possible constructs. Finally, error correlations between manifest items were considered only if a clearly evident rationale was present (e.g., similar phrases appearing in the survey items). Unless indicated otherwise, no error correlations were added to the models presented below. After the model had been established, internal reliability estimates among the items were obtained with Cronbach's alpha.

The large number of survey items and the potentially large number of constructs required the measurement models to be assessed piecemeal. When reasonably possible, multiple constructs were assessed collectively. When the size of the covariance matrix made this difficult due to convergence issues, models containing fewer constructs were confirmed.

There were 12 items intended to measure self-esteem. Two constructs were detected with $n = 1705$. The first construct appeared to measure perceptions of self-esteem (e.g., "I am pleased with myself"). Three items loaded on this factor (labelled SE-P). The second construct appeared to measure actions supported by self-esteem (e.g., "Big challenges bring out the best in me"). After dropping one item for very weak loading, 8 items loaded on this factor (labelled SE-A). The model fit for these constructs was high with $\chi^2(43) = 332.0$, $p < .001$, CFI = 0.953, RMSEA = .063 in a 90% confidence interval of (.057, .069). The reliability measures were $\alpha = .79$ for SE-P and $\alpha = .82$ for SE-A. As would be expected, the interfactor correlation was high ($r = .816$).

There were 15 items intended to measure perceived expectations in the classroom with $n = 1679$. While the final model was less conventional than a single-factor model, five predicted sub-constructs were detected, and all loaded on a second-order latent construct. Due to space limitations, the sub-constructs will not be discussed here. The larger construct appeared to measure expectations (e.g., "Be positive towards learning about things that are new for me"). Three items loaded on each of the 5 sub-constructs, and all loaded on the second-order factor (labelled EXP). No items were dropped, and one error correlation between two items was included in the model due to a common phrase ("outside of the class", $r = .176$). The model fit for this construct was high with $\chi^2(84) = 480.6$, $p < .001$, CFI = 0.960, RMSEA = .053 in a 90% confidence interval of (.048, .058). The reliability measure for the second-order factor was $\alpha = .90$. In support of the use of a second-order factor model, the first-order construct standardized loadings were high and ranged from $\lambda = .78$ to .93.

The remaining items were intended to measure various characteristics of the classroom environment. The next measurement model analysed was to capture students' educational values and learning outcomes. When analysed together, three constructs were detected with $n = 1683$. The first construct appeared to measure educational values (e.g., "I gain satisfaction from learning new things"). Five items loaded on this factor (labelled EV). The second construct appeared to measure learning outcomes relating to grades (e.g., "My test scores are high"). Six items loaded on this factor (labelled LO-G). The third construct in this group appeared to measure learning outcomes relating to actions (e.g., "I meet homework requirements"). Three items loaded on this factor (labelled LO-A). No items were dropped, and one error correlation between two items was included in the model due to a distinct concept from the other items (success as a student vs. success on specific assessments, $r = .219$). The model fit for these constructs was high with $\chi^2(73) = 574.6$, $p < .001$, CFI = 0.955, RMSEA = .064 in a 90% confidence interval of (.059, .069). The reliability measures were $\alpha = .82$ for EV, $\alpha = .88$ for LO-G and $\alpha = .76$ for LO-A. Supporting both convergent and divergent validity of the items, the interfactor correlation was high between LO-G and LO-A ($r = .741$), and moderate between EV and LO-G ($r = .576$) and EV and LO-A ($r = .656$).

The next measurement model analysed was to capture students' perceptions of classroom learning, discussions and planning. When analysed together, three constructs were detected with $n = 1639$. The first construct appeared to measure perceptions of classroom learning (e.g., "Learning is really important in this class"). One item was dropped, and one error correlation between two items was included in this part of the model due to a common concept (specific reference to learning "in this class", $r = .152$). Ten items loaded on this factor (labelled CL). The next construct appeared to measure perceptions of classroom discussions (e.g., "We have discussions about what we should be learning"). No items were dropped, and one error correlation between two items was included in this part of the model due to word choices (these items use the verb "discuss" while the other items use the verb "talk", $r = .150$). Five items loaded on this factor (labelled CD). The third construct appeared to measure perceptions of classroom planning (e.g., "We are given assessment tasks or tests when we are ready"). Three items loaded on this factor (labelled CP). The model fit for these constructs was high with $\chi^2(130) = 675.2$, $p < .001$, CFI = 0.957, RMSEA = .051 in a 90% confidence interval of (.047, .054). The reliability measures were $\alpha = .87$ for CL, $\alpha = .86$ for CD and $\alpha = .81$ for CP. As would be expected, the interfactor correlations were moderately high and ranged from $r = .667$ to .732.

There were 11 items intended to measure perceived supports in the classroom. Two constructs were detected with $n = 1652$. The first construct appeared to measure perceptions of classroom support via collaboration (e.g., "Students are willing to help each other when problems arise"). Five items loaded on this factor (labelled CS-C). The second construct appeared to measure perceptions of classroom support via harmonious interaction (e.g., "Students are not nasty towards each other"). Six items loaded on this factor (labelled CS-H). No items were dropped. The model fit for these constructs was high with $\chi^2(43) = 353.9$, $p < .001$, CFI = 0.970, RMSEA = .066 in a 90% confidence interval of (.060, .073). The reliability measures were $\alpha = .88$ for CS-C and $\alpha = .88$ for CS-H. As would be expected, the interfactor correlation was high ($r = .872$).

The final measurement model analysed was to capture students' perceptions of support from the teacher and support from parents. When analysed together, three constructs were detected with $n = 1596$. The first construct appeared to measure perceptions of the teacher as helpful (e.g., "The teacher helps students who get into trouble around the school"). One item was dropped, and one error correlation between two items was included in this part of the model due to a common concept (specific reference to teacher and family, $r = .206$). Five items loaded on this factor (labelled YT-H). The next construct appeared to measure perceptions of the teacher as holding high expectations (e.g., "The teacher sets high standards"). No items were dropped, and three items loaded on this factor (labelled YT-E). The third construct appeared to measure perceptions of parental support (e.g., "My parent(s) take an interest in my progress"). No items were dropped, and one error correlation between two items was included in this part of the model due to a common concept (these items did not refer specifically to communication with teacher whereas the other items did, $r = .198$). Five items loaded on this factor (labelled YP). The model fit for these constructs was high with $\chi^2(60) = 399.4$, $p < .001$, CFI = 0.958, RMSEA = .060 in a 90% confidence interval of (.054, .065). The reliability measures were $\alpha = .82$ for YT-H, $\alpha = .83$ for YT-E and $\alpha = .82$ for YP. Interfactor correlations were moderate and ranged from $r = .319$ to .531.

Structural (Path) Models

With the confirmation of psychometrically sound measures, the next step was to explore potential relationships between the constructs. The principal hypothesis was that the classroom characteristics would influence students' self-esteem and expectations. With this guiding the analysis, combinations of the characteristics were used to predict the self-esteem and expectation measures. The protocol was to free parameters for estimation (inclusion in the model) based on sufficiently large Lagrange Multiplier modification indices. Once items had been added to the model, the Wald test was used to determine if items with insignificant regression parameters could be dropped. Once evidence was obtained to indicate which items appeared to provide predictive relationships for the three dependent variables, all of the influential characteristics were combined into one model, and the process of adding and removing parameters was repeated. This process was conducted on the factor scores derived from the measurement models[1]. Once a final model had been obtained and cross-validated with split-half random samples, the final step was to run the model with the complete set of relevant manifest variables.

The most noteworthy observation about this process was that every analysis with the classroom characteristics indicated that predictive relationships should be included from general perceptions of self-esteem (SE-P) to actions related to positive self-esteem (SE-A), and from SE-A to perceived classroom expectations (EXP). Along with every analysis indicating the same relationships in the same directions, none of the analyses indicated the addition of an indirect effect from SE-P to EXP. While there is nothing to support a causal relationship in these findings, the consistency of the findings in the different scenarios does support a correlational predictive relationship between the variables with the indirect effects well accounted for by the classroom-characteristic variables in the model. Furthermore, it should be noted that these findings align well with the underlying theory and with a commonsense understanding of the concepts relating to student engagement.

[1] This method is not desirable, but the size of the matrices involved in the calculations proved too challenging for the SEM package in R. While the models would run, the memory requirements to obtain modification indices were too large. As the focus would be on paths between the latent constructs and none of the other manifest or latent variables, it was decided comparable information could be obtained from an analysis of the estimated factor scores.

Regarding the analysis of the classroom characteristics, this process resulted first in the removal of three of the classroom characteristic constructs: classroom discussions (CD), classroom planning (CP) and harmonious classroom support among students (CS-H). When examined separately with the dependent variables (self-esteem and expectations), there was no predictive power for any of these constructs. When the remaining eight constructs were combined into one model, the most noteworthy observation was the removal of one predictor. When combined with the other characteristics, perceptions of the teacher as helpful did not predict any of the dependent variables. Finally, while parameter values slightly changed, the general analysis from the individual models to the aggregate model were in strong agreement (i.e., very few paths were added or removed in the aggregate model that were not supported by the preliminary analyses of models with fewer predictors). Summary statistics and correlations of all latent constructs appearing in the final model are found in Table 1. The final structural model is shown in Figure 1.

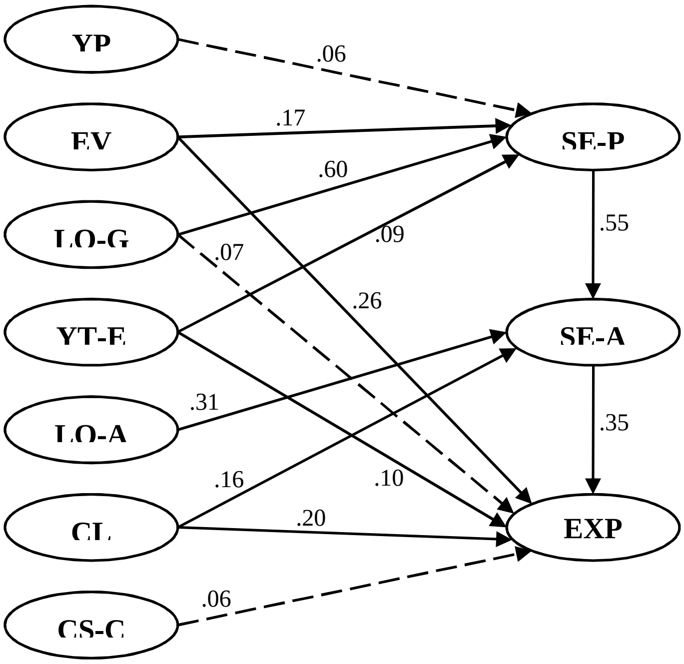
Table 1
Summary Statistics for Estimated Factor Scores for Latent Constructs Appearing in the Final Model.

			Correlations (Cronbach's α)									
	M (SD)	Range	LO-G	LO-A	EV	CL	CS-C	YT-E	YP	SE-P	SE-A	EXP
LO-G	0.01 (0.41)	[-0.76, 0.85]	*(.88)*									
LO-A	0.01 (0.49)	[-0.97, 1.02]	.84	*(.76)*								
EV	0.01 (0.44)	[-1.01, 0.81]	.65	.76	*(.82)*							
CL	0.00 (0.40)	[-0.82, 1.01]	.53	.63	.64	*(.87)*						
CS-C	0.00 (0.49)	[-1.00, 0.97]	.39	.46	.47	.65	*(.88)*					
YT-E	0.00 (0.49)	[-1.14, 0.72]	.26	.32	.31	.40	.34	*(.83)*				
YP	0.00 (0.34)	[-0.59, 0.68]	.38	.43	.44	.52	.35	.36	*(.82)*			
SE-P	0.01 (0.46)	[-1.18, 0.87]	.69	.64	.58	.47	.35	.30	.39	*(.79)*		
SE-A	0.01 (0.33)	[-0.92, 0.69]	.70	.69	.63	.55	.41	.34	.42	.90	*(.82)*	
EXP	0.00 (0.93)	[-2.74, 2.10]	.60	.64	.66	.64	.49	.39	.41	.59	.66	*(.90)*

Note: n = 1430; internal reliability measures (Cronbach's α) for each construct are listed on the diagonal and italicized. All correlations significant at α = .001.

Based on the factor scores, the predictive model included seven exogenous variables with three endogenous variables. As would be expected with an SEM using estimated factor scores, the model fit was excellent: $\chi^2(11) = 17.25$, $p = .101$, CFI = 0.999, RMSEA = .020 in a 90% confidence interval of (.000, .037). While the results obtained from the SEM using the estimated factor scores produced comparable findings as when the SEM was run using the manifest variables, a noticeable change in the parameters for the relationships between the endogenous variables was observed. As such, it seems appropriate to report both findings. The model fit for the manifest variables was good: $\chi^2(1847) = 6012$, $p < .001$, CFI = 0.908, RMSEA = .039.

Standardized path coefficients for the final models (one with estimated factor scores and one with the manifest variables) are reported in Table 2. For the first model, all paths were significant at $\alpha = .002$. For the second model, all paths were significant at $\alpha = .05$ (with all but three significant at $\alpha = .001$).



Note: Classroom characteristics appear on the left of the model, and student engagement measures appear on the right; n = 1484, and coefficients obtained using manifest variables; dashed lines indicate paths with regression parameters significant at α = .05 (all others significant at α = .001).

Figure 1. Structural (path) model relating classroom characteristics to student engagement.

Table 2
Standardized Regression Path Coefficients for the Final Model when Calculated with Estimated Factor Scores and with Manifest Survey-Item Variables.

	Factor Scores			Manifest Variables		
Paths from…	SE-P	SE-A	EXP	SE-P	SE-A	EXP
exogenous						
LO-G	0.52		0.10	0.60		0.07*
LO-A		0.15			0.31	
EV	0.18		0.24	0.17		0.26
CL		0.10	0.20		0.16	0.20
CS-C			0.07**			0.06*
YP	0.08			0.06*		
YT-E	0.08		0.10	0.09		0.10
endogenous						
SE-P		0.76			0.55	
SE-A			0.26			0.35

Note: For factor score model, n = 1430; for manifest variable model, n = 1484; significance level α = .05 indicated by *, significance level α = .005 indicated by **, all other values significant α = .001.

Table 3
Unanalysed Correlations among Exogenous Variables in the Final Model when Calculated with Estimated Factor Scores and with Manifest Survey-Item Variables.

	LO-G	LO-A	EV	CL	CS-C	YT-E	YP
LO-G		0.75	0.59	0.56	0.37	0.26	0.38
LO-A	0.84		0.67	0.70	0.47	0.34	0.46
EV	0.65	0.76		0.71	0.48	0.32	0.46
CL	0.53	0.63	0.64		0.68	0.42	0.54
CS-C	0.39	0.46	0.47	0.65		0.36	0.35
YT-E	0.26	0.32	0.31	0.40	0.34		0.31
YP	0.38	0.43	0.44	0.52	0.35	0.36	

Note: Correlations for factor score model are the lower-triangular matrix elements (n = 1430), and correlations for the manifest variable model are the upper-triangular matrix elements (n = 1484); all values significant at α = .001.

Discussion

The first research question for this phase of the project was to determine if the survey instrument demonstrated strong psychometric properties for the latent constructs of student engagement in classroom learning and the characteristics of the classroom-learning environment. For all of the measurement models examined, the goodness-of-fit indices were all strong (with the lowest CFI being .953—all greater than .95), the badness-of-fit indices were moderate (with the RMSEAs ranging from .051 to .066—all less than .08, but greater than .05), and the internal reliabilities of the items for each construct ranged from α = .76 to .90 (all moderate to strong). The presence of the latent constructs was confirmed, and the reliability of the instrument was demonstrated. The construct validity can be inferred from the observed relationships aligning well with theoretical predictions for hypothesized models between the constructs intended to be measured. Together, this suggests that this survey instrument is a viable psychometric tool for measuring student engagement and classroom environment.

While the findings do not perfectly parallel those of the first phase of this study (Cavanagh, 2012), a few noteworthy observations about the psychometric analyses can be made. First, the Rasch analysis of the first phase demonstrated a uni-dimensional interval scale for each construct. This does not mean that there is a uni-dimensional nature to the data. In particular, the uni-dimensional scale can be used to measure multiple sub-constructs. The observed differences in the mean difficulties for each sub-construct from the first phase of the study coupled with the observed latent factor structure from the second phase of the study provides compelling evidence for the presence of multiple sub-constructs in the classroom characteristics. Furthermore, the size of the estimated correlations between the classroom characteristics (the exogenous variables) from the final model are moderate to strong (see Table 3). This provides additional support for a uni-dimensional link between each of the classroom characteristics. In addition, it was noted that the weaker correlations were predicted for the two constructs that were not retained in the Rasch analysis of the first phase (teacher and parental support).

The second research question for this phase of the project was to determine what relationships exist between student engagement and classroom environment. In this study, the sub-facets of engagement were operationalized as self-esteem, actions relating to positive self-esteem and overall learning expectations. As this exploration attempted to locate a model that fit the data, all conclusions should be approached with caution. However, the strength of the resulting models does warrant support for future exploration of these findings. As already noted, a predictive relationship emerged for self-esteem and actions related to positive self-esteem. As this aligns with the concept of thought preceding action, this finding appears reasonable. Next, the actions related to positive self-esteem predicted positive expectations of the learning. This finding coincides nicely with

theories of student engagement. While no indirect effects were detected for these constructs, the spurious and unanalysed effects were accounted for via the classroom characteristics.

The finding that harmonious classroom support among students and perceptions of the teacher as helpful did not influence students' self-esteem or learning expectations was curious. While this might suggest that engagement is more related to a personal trait than verbal support or encouragement, it would have to be explored further. Additionally, classroom discussions and classroom planning was not found to be predictive of student engagement. This could be attributed to the variety of quality of discussions or levels of inclusive planning among the students and classroom surveyed for this study. Again, as the nature of this analysis was exploratory, further research into this relationship is still warranted.

Regarding the final model, the fit indices, modification indices and significance of parameter estimates all indicated excellent fit for the factor score analysis. As might be expected, learning outcomes relating to grades had the largest influence on perceptions of self-esteem. As self-esteem in this context is strongly associated with each student's academic self-concept, a highly predictive relationship between the two constructs is reasonable. Additionally, educational values had a relationship with both self-esteem and learning expectations. In line with the traditional expectancy-value motivation theory, these findings seem reasonable. Classroom learning had a relationship with both actions related to positive self-efficacy and with learning expectations. This suggests that the learning focus of the environment might influence the learning expectations of the students. This finding aligns well with the predicted relationship of classroom influence on student engagement. Finally, parental support influenced self-esteem, but did not directly influence learning expectations, whereas teachers' expectations influenced both self-esteem and student expectations for learning. Each finding coincides with the appropriate sphere of influence, and appears reasonable.

Two findings were noticeably different between the analysis of the final model with the factor scores (supported by the psychometric qualities demonstrated for the survey) and with the manifest variables (the responses to the survey items). The path from actions related to learning outcomes and actions related to positive self-esteem noticeably increased from the model using factor scores to the model using the manifest variables. As both of these constructs tap into the underlying notion of student-initiated action (e.g., "I make an effort" from the engagement measure and "I perform to the best of my ability" from the classroom measure), it is possible the path coefficient is inflated in the manifest model as it may be subsuming part of the covariance related to similarly-worded items. On the other hand, the path from the endogenous variables self-esteem to actions related to positive self-esteem noticeably decreased from the model using factor scores to the model using the manifest variables. This could be a demonstration of the increase in correlations from distinct measures to aggregated measures (Triola, 2001). That is to say, correlations between averages of items tend to be larger than correlations between the individual items. With the fact that comparable values were obtained on all other parameters for the final model, it may be argued that the more conservative (smaller) parameter values for these two paths may be more representative of the underlying relationships in question.

Conclusion

The two phases of this study provide compelling evidence that a rich picture can be obtained from data analyses when complementary methods are employed to examine the data. In this case, SEM and IRT findings supported the conclusion of psychometric reliability, and at the same time, each model provided slightly different insight into the relationships between the constructs being measured.

This investigation is an important contribution to knowledge and theorizing about student engagement and classroom learning environments. This second phase of the research complements the findings of the first phase. The first phase used Rasch modelling to demonstrate the presence of a uni-dimensional scale with the possibility of a dominant construct for classroom learning. This analysis used SEM to elaborate the multiple sub-constructs that comprised the larger classroom-learning construct. The fit of the model to the data was quite strong.

Further research should be conducted to confirm predictive relationships observed between sub-constructs of classroom characteristics and student engagement. Additionally, demographic variables should be analysed in future research to ascertain the presence of moderating effects on student engagement. Of particular interest for motivation researchers would be to determine if a different model emerges for students responses when referencing their favourite subject compared to their non-favourite subjects. Finally, it would be beneficial for future research to examine the model in different populations, including primary and tertiary students.

References

Arbuckle, J. L. (2009). AMOS (Version 18.0) [Computer Program]. Chicago: SPSS.

Bond, M.H., & Hwang, K.K. (1986). The social psychology of Chinese people. In M.H. Bond (Ed.), *The psychology of Chinese people* (pp. 213-266). Oxford, UK: Oxford University Press.

Byrne, B.M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cavanagh, R. F. (2012, December). *Associations between the classroom learning environment and student engagement in learning 1: A Rasch modelling approach.* Paper presented at the Australian Association for Research in Education Annual Conference, Sydney, NSW, Australia.

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

Cavanagh, R.F., & Waugh, R.F. (2004). Secondary school renewal: The effect of classroom learning culture on improving educational outcomes. *Learning Environments Research, 7*(1), 245-269.

Ceci, S.J., Rosenbaulum, T., DeBruyn, E., & Lee, D.Y. (1997). A bioecological model of human development. In R.J Sternberg & E.L. Grigorenko (Eds.), *Intelligence, heredity, and environment* (pp. 303-322). Cambridge, UK: Cambridge University Press.

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

Fox, J., Nie, Z., & Byrnes, J. (2012). sem: Structural Equation Models. R package version 3.0-0. http://CRAN.R-project.org/package=sem

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 51-109.

Kennish, P., & Cavanagh, R.F. (2011). The engagement i classroom learning of years 10 and 11 Western Australian students. In R.F. Cavanagh & R.F. Waugh (Eds.), *Application of Rasch Measurement in Learning Environments Research* (pp. 285-304). Rotterdam: Sense Publishers.

R Development Core Team (2012). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Triola, M.F. (2001). *Elementary statistics,* (8th ed.). Boston, MA: Addison Wesley Longman.