Curtin
UNIVERSITY OF TECHNOLOGY

A two-step approach to modelling student performance: A methodology that accounts for measurement and structural error

Beverley J Webster

Occupational Therapy Research Centre

Darrell L Fisher

Science and Mathematics Education Centre

Paper presented at the Australian Association for Research in Education

December 3-6, 2001 Fremantle Perth

Contact Details

Beverley Webster (Director) GPO Box U1987,

Occupational Therapy Research Centre of WA Perth WA 6845

Email: b.webster@curtin.edu.au Curtin University of Technology

Ph: +61 8 9266 3611

ABSTRACT

This paper describes the methodology used to investigate influences on student mathematics achievement and addresses the techniques of secondary analysis and its associated limitations and includes a two-step approach to modelling. A total of 57 Australian schools were surveyed and data collected from 620 teachers and 4645 students. The two-step model building approach consisted of the analysis of two conceptually distinct models. The first was an analysis of the measurement model, which specifies the relationships between the observed variables and the latent variables. The results of this analysis identified the measurement properties (reliabilities and validities) of the observed and latent variables. The second involved a structural equation model, which specifies the relationships among the latent variables as posited by theory and previous research and also describes those links between the latent variables and student science and mathematics achievement. This two-step approach to modelling student performance allows for clarity of model fit and identifies whether any source of poor fit is due to the measurement or the structural model. This paper then presents an explanatory model of student performance incorporating the student home background, student attitudes towards mathematics, success attribution, instructional practices and school-level environment factors.

POPULATION

Third International Mathematics and Science Study (TIMSS)

TIMSS is a cross-national survey of student achievement in mathematics and science that was conducted at three levels of the educational system.

- The two adjacent grades with the largest proportion of 9-year olds at the time of testing (third and fourth grades in many countries)
- The two adjacent grades with the largest proportion of 13-year olds at the time of testing (seventh and eighth grades in many countries)
- The final year of secondary education

Forty-five countries eventually agreed to participate in the survey. The students, their teachers, and the principals of their schools were asked to respond to questionnaires about their backgrounds, attitudes, experiences, and practices in the teaching and learning of mathematics and science.

Sample design for TIMSS

The basic sample design used in the TIMSS was a two-stage stratified cluster design of three different populations . The first stage consisted of a sample of schools; the second stage consisted of samples of intact mathematics classrooms from each eligible target grade in the sampled schools. The design required schools to be sampled using a probability proportional to size systematic method (PPS), as described by Foy, Rust, and Schleicher (1996), and class rooms to be sampled with equal probabilities . To account for differential probabilities of selection due to the nature of the design, TIMSS computed a sampling weight for each student that participated in the assessment and these weights were included in all analyses involving student data. The general weighting procedure for TIMSS required three steps. The first step for all target populations consisted of calculating a school weight. The second step consisted of calculating a classroom weight, which was calculated independently for each school and grade. The final step consisted of calculating a student weight independently for each sampled classroom. The overall sampling weight attached to each student record is the product of the three intermediate weights: school, classroom, and student.

Sample design for the Australian schools

All of the Australian schools that participated in the TIMSS study were invited to also participate in the school level environment study. This time, school selection was on a volunteer basis. Requests to principals were in the form of a letter explaining the aims of the research and seeking their cooperation. There were initially 161 letters posted to schools in all states and territories, all the schools who participated in the TIMSS study, asking them if they would allow their staff members to participate in the study. A total of 57 schools agreed to participate and from these schools 620 teachers returned completed questionnaires.

Table 1

Number of Schools from each State/Territory

State/Territory	Schools	Teachers	Classes	Students
Australian Capital Territory	3	28	9	265
New South Wales	8	65	25	510
Northern Territory	2	14	8	152
Queensland	11	147	43	1024
South Australia	7	68	29	627
Tasmania	3	23	9	218
Victoria	14	142	43	970
Western Australia	9	133	37	879

Total	57	620	203	4645

VARIABLES IN THE STUDY

Mathematics Achievement

The dependent outcome measure in this study was mathematics achievement and was available with the TIMSS data. The overall mathematics results from TIMSS were summarized using an item response theory (IRT) scaling method. The mathematics tests consisted of 36 multiple choice and free response items and the scales were constructed using the IRT procedure, Rasch Modelling . This procedure scores the test items and then estimates the student's ability on that test item as a function of the difficulty of the test item and the student responses to other test items. The purpose of the Rasch analysis was to make the achievement scores independent of the sample (student ability) and the item difficulty; some of the rotated test items may have been more difficult than other test items.

Socio-economic status

A family socio-economic index was calculated for each student as a composite of mother's education, father's education, English as a first language, number of books in the students home, and family size.

Student Attitudes

An overall attitude index was calculated for each student as a weighted composite of five items listed below which asked students to state their level of agreement on a scale from strongly agree (4) to strongly disagree (1) about their view of the utility of mathematics, their enjoyment of mathematics and their perception of the importance of mathematics.

Success Attribution

Items in this scale were designed to indicate the extent to which students believe they need various attributes or activities to do well in mathematics. Students were asked to respond on a scale of strongly agree (4) to strongly disagree (1) indicating whether they needed lots of natural ability, good luck, lots of hard work studying at home or memorising the textbook or notes to do well in mathematics. The higher scores in this scale indicate that students believe that success in mathematics is due more to internal attributes.

Instructional Practices

Instructional practices in the classroom were measured as dimensional, weighted composites of ten items. Students were asked to say how often in their mathematics classes each of the items occurred from almost always happens (4) to never happens (1). Some items required reversing of the responses so that the higher scores consistently reflected the more Teacher Centred approach and the lower scores reflecting the Student Centred approach.

School Level Environment

The instrument selected to measure the school environment in the study described in this thesis, was the School Level Environment Questionnaire (SLEQ). The version of the SLEQ used included 56 items that produced eight scales (Student Support, Affiliation, Professional Interest, Mission Consensus, Empowerment, Innovation, Resource Adequacy, Work Pressure) and each scale consisted of 7 items as described in Chapter Two. Each of the 56 items was scored on a five-point scale with responses of Strongly Disagree (1) to Strongly Agree (5). The development of the SLEQ and issues of reliability and validity have been documented in detail . A description of what each SLEQ scale is measuring is presented in Table 2.

Table 2

Description of the Scales in the SLEQ

Scale Name	Description of Scale	Sample Item
Student Support	There is good rapport between teachers and students, students behave in a responsible self-disciplined manner	There are many disruptive, difficult students in the school (-)
Affiliation	Teachers can obtain assistance, advice and encouragement, and are made to feel accepted by their colleagues	I feel that I could rely on my colleagues for assistance if I should need it (+)
Professional Interest	Teachers discuss professional matters, show interest in their work and seek further professional development	Teachers frequently discuss teaching methods and strategies with each other (+)
Mission Consensus	Consensus exists within the staff about the goals of the school	Teachers agree on the school's overall goals (+)
Empowerment	Teachers are empowered and encouraged to be involved in decision making processes	Decisions about the running of this school are usually made by the principal or a small group of teachers (-)
Innovation	The school is in favour of planned change and experimentation, and fosters classroom openness and individualisation	Teachers are encouraged to be innovative in this school (+)
Resource Adequacy	Support personnel, facilities, finance, equipment and resources are suitable and adequate	The supply of equipment and resources is inadequate (-)
Work Pressure	The extent to which work pressure dominate school environment	Teachers have to work long hours to keep up with the work load (+)

METHOD OF ANALYSIS

Secondary Analysis

The basic procedure used in this study was secondary analysis, a procedure that involves the analysis of data collected previously by other researchers. It is a very general label covering a wide range of analyses. The Third International Mathematics and Science Study (TIMSS) data-base, which provided the first stage of data collection, was extraordinarily large in terms of numbers of students and numbers of variables. Additional data were collected in a second stage and these data were merged with the TIMSS data and used in the analysis. The range of problems associated with this type of research and the skills required differ from those involved in smaller scale studies. In this research, adequate measures have been taken and procedures followed to allow an accurate and detailed interpretation of all the results presented.

A recognised advantage of secondary analysis is the availability of very large data sets not easily obtained by individual researchers, the low cost of these databases and the reduced student response burden on schools. The problems associated with secondary analysis can be that the statistical expertise necessary for this type of analysis is not available, the data are not in a useable form due to poor documentation and are often difficult to understand, it takes time to understand the data in terms of what information is present, and the data may not contain the precise variables of interest and be of limited quality. For example, in this present study, a significant limitation was the lack of a measure of students' prior mathematical ability, not allowing the concept of value added or distance travelled, to be explored. Cross-sectional studies such as this one do not allow for the measurement of growth and change . An additional limitation in this study included restrictions in using variables chosen previously which did not adequately address the issues of social and environmental influences on mathematics achievement.

Introduction to two-step approach to modelling

The two-step model building approach is recommended by Anderson and Gerbing (1988) and Joreskog and Sorbom. The first part is an analysis of the measurement model, which specifies the relationships between the observed variables (items) and latent variables or hypothetical constructs (factors). The results of this analysis identify the measurement properties (reliabilities and validities) of the observed and latent variables . This is done separately before fitting a structural equation model, to look at the relationship between the latent variables. Measurement models were defined for all the independent latent variables. The dependent variable, Mathematics Achievement, was developed by TIMSS and is used as it is in this analysis. The second part is the structural equation model, which specifies the relationships among the hypothetical constructs (latent variables) as posited by theory or previous research and describes the links between school-level environment, student background characteristics and student cognitive and affective outcomes.

This two-step approach allows the researcher to identify sources of poor fit of a full structural model and decide whether this poor fit is due to the measurement or structural model. Jöreskog and Sörbom (1996) believe that the testing of the structural model can be meaningless unless it is first established that the measurement model holds. They further noted that if the chosen observed variables for a construct do not measure that construct, the specified theory must be modified before it can be tested. Therefore, the measurement model should be tested prior to the structural relationships being tested.

Measurement model (confirmatory factor analysis)

The term confirmatory factor analysis (CFA) is also used to refer to the analysis of measurement models. CFA approaches attempt to test the viability of *a priori*structures, which have been identified based on theory or previous experience or research, and to examine whether or not existing data are consistent with a highly constrained *a priori* structure that meets conditions of model identification . In this study, two types of measurement models were assessed, namely one-factor congeneric models and multi-factor models. One-factor congeneric measurement model analysis was used to assess item reliability, determine scale reliability, and to generate factor score regression values for computing composite variables to be used in the structural model.

One factor-congeneric measurement model

A one-factor congeneric measurement model is one type of measurement model within which a single latent variable (factor) is measured by several observed variables (items). Latent (unobserved) variables, which represent abstract concepts or theoretical constructs, are not directly observable or measured but must be assessed indirectly or inferred . This is often accomplished by collecting responses for a number of items and then computing the latent (unobserved) variable. Such variables are often referred to as factors or constructs. For example, student attitude toward mathematics cannot be directly observed (e.g. through visual inspection of an individual) and thus there is no single, agreed upon definition or measure of attitude, however, it can be indirectly measured or inferred through observable or indicator variables (e.g. question items from the TIMSS student questionnaire).

Observed (indicator) variables are variables that are directly observable or such as items in a survey instrument. Schumacker and Lomax (1996) suggest a minimum of three items is required for fitting a congeneric model and computing a latent construct (factor). Four to five items per factor are recommended.

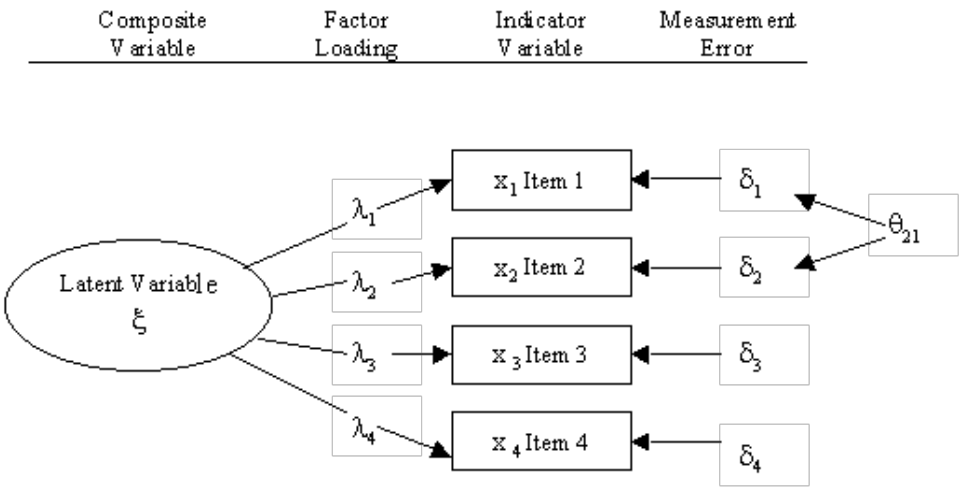

Figure 1. A generic one-factor congeneric model.

Figure 1 presents an example of a one-factor congeneric measurement model that defines an independent latent variable (x). This approach can also be used to define a dependent

latent variable (h) and is appropriate for analysing the latent variables related to factors affecting student outcomes. The processes of fitting both independent and dependent one-factor congeneric models are exactly the same, the only difference being the LISREL notation used in the pictorial representations.

Such a model was tested for each latent variable separately for student, class, and school components. LISREL 8.30 was used to test each hypothesised measurement model, such as the eight school level environment scales.

A one-factor congeneric model represents the regression of a set of observed variables on a single latent variable. In Figure 2 the following apply:

- x is the single latent independent variable (factor) in the model depicted by an ellipse
- x_i indicates the observed indicator variables (items) and are depicted by rectangles

- d_i are the errors associated with the measurement of x_i
- l_i are the regression coefficients (factor loadings) of the x in the regression against x_1
- q i indicates the error variances that are allowed to co-vary indicating that something in common is being measured, other than that measured by the items being estimated

For the observed variable, researchers are particularly interested in the extent they actually measure the hypothesised latent variable – i.e., how good is the SLEQ as a measure of school level environment? Which observed variable is the best measure of a particular latent variable – i.e. is item 5 a better measure of 'student support' than item 3? Or, to what extent are the observed variables actually measuring something other than the hypothesised latent variable – i.e., is the SLEQ measuring something other than school level environment.

The measurement model reflects the extent to which the observed variables are assessing the latent variables in terms of reliability and validity. The relationships between the observed variables and the latent variables are described by factor loadings l_1, and convergent validity is reflected in the magnitude of the factor loadings. In other words, the factor loadings provide us with information about the extent to which a given observed variable is able to measure the latent construct . If observed variables specified to measure a common underlying factor all have relatively high loadings on that factor, then this is evidence of convergent validity .

Measurement error (d_1) is the portion of an observed variable that is measuring something other than what the latent variable is intended to measure. It serves as a measure of reliability . A correlated measurement error (q_{21}) reflects the assumption that items 1 and 2 measure something in common that is not represented in the model .

Use of congeneric measurement models provides a number of benefits in structural equation modelling, for example, reducing the number of observed variables to a single latent variable, assessing item and composite reliability, improving the reliability and validity of composite variables, etc.

Multi-factor analysis

In previous school environment studies, the constructs have been aggregated to give overall environment scores for participating schools, however, by doing this environment is represented as a unidimensional concept. Yet it is believed that both classroom level and school level environments are multidimensional constructs. Teachers may be happy with some aspects of the environment and yet not be happy with other aspects of the environment, which in an aggregated arrangement for constructing factors negates the ability to distinguish between these possibilities. A multifactor model analysis allows researchers to test the multidimensionality of a theoretical construct. Moreover, a multi-factor analysis allows researchers to address the issues of convergent and discriminant validity. Discriminant validity refers to the distinctiveness of the factors measured by different sets of observed variables and can be supported if the estimated correlations between the factors are not excessively high.

Structural equation models

The structural models, which specify and simultaneously estimate the relationships among the latent dependent (endogenous: such as student outcome measures) and dependent (exogenous: such as school level environment measures) variables, describe the links between variables. Measurement models must first be specified and fitted to test that the latent variables (factors) are measured well (valid and reliable) by selected observed variables (items). In this study, in building structural equation models, both the independent and dependent latent-variable measurement models were used.

Based on the measurement model, the values of factor loadings and measurement error variances were then fixed when examining the latent-variable relationships in subsequent structural equation models . Since the factor loadings and the measurement error variances were fixed in the measurement part of the model, the parameters to be estimated are structural regression coefficients in the structural part of the model. The structural regression coefficients indicate the strength (i.e. weak or strong) and direction (i.e. positive or negative) of the relationships among the latent variables. Each structural equation also contains an error term that indicates the portion of the latent dependent variable that is not explained or predicted by the latent independent variables in that equation.

Model assessment and modification

The main purpose of conducting structural equation modelling analysis is to assess the extent to which a hypothesised model adequately describes the sample data. The steps discussed by Burne (1998) were used in this study to assess the adequacy of a hypothesised model and to detect any sources of mis-estimation in the model. This process includes the following steps:

1. adequacy of the parameter estimates. A non-significant parameter can be considered unimportant to the model and should be deleted .
2. adequacy of the measurement model. This can be determined from the squared multiple correlation (R^2) reported for each observed variable (which is an indication of item reliability with respect to its underlying latent construct) and the coefficient of determination reported for all the observed variables jointly (which is an indication of composite reliability for the individual measurement model). These values range from 0 to 1.00; values close to 1.00 represent good models. The scale coefficient of determination is equivalent to the maximised composite scale reliability coefficient used in this study .

3. adequacy of the model as a whole. It is recommended that assessment of model adequacy must be based on various goodness of fit criteria that take into account theoretical, statistical, and practical considerations . The following criteria were used to evaluate the adequacy of the model fit: Normed Chi square (c^2/df) < 3 ,the Root Mean Square Error of Approximation (RMSEA) < .05, Non-Normed Fit Index (NNFI) > .090, the Comparative Fit Index (CFI) > 0.90 and the Adjusted Goodness of Fit Index (AGFI) > 0.90 . All of these criteria were adopted in constructing the measurement and structural models. Each of the indices is an assessment of fit between the hypothesised model and sample data.

If a hypothesised model did not fit the given data, the model could be modified to fit the data better through post hoc model testing. Post hoc analysis focuses on the detection and identification of the source of poor model fit in the originally hypothesized model, based on improvement information from LISREL (i.e. residual and modification index) . *Residual* is the discrepancy between the sample covariance matrix and the fitted covariance matrix. A well-fitted model should have standardised residuals that are less than 2.58 and symmetrically clustered around the zero point. The value of a *Modification Index* represents the expected drop in overall c^2 value if the parameter were to be freely estimated. This information provides a guideline about how a hypothesized model can be refined. However, parameters were only free to be estimated if there was substantive support from previous research or theory to indicate that there is justification for so doing.

In summary, this research adopted Anderson and Gerbing's (1988) two-step approach, where the measurement parameters for the latent variables were estimated and then fixed in the structural model. In this way, the reliability and validity of the measurement model are established before structural models are estimated.

RESULTS

Confirmatory factor analyses, both one and multifactor measurement models, were used to examine the theoretical constructs included in the structural equation analysis which are as follows:

- eight school-level environment constructs (Student Support, Affiliation, Mission Consensus, Professional Interest, Empowerment, Innovation, Resource Adequacy and Work Pressure)
- student socio-economic status
- student attitudes towards mathematics
- instructional practices
- success attribution

All the constructs were estimated using LISREL 8.30 with polychoric correlation and asymptotic covariance matrices and using a weighted least square (WLS) estimation procedure. The goodness of fit statistics, the maximised composite reliabilities (r_c), the estimated composite variable regression coefficients ($l^{\wedge}_c$), the measurement error variances ($q^{\wedge}_c$), and the unit weight reliability coefficients (a^b) for each of the composite variables estimated are summarised in Table 3 (SLEQ) and Table 4 (student-level scales).

Table 3

Fitted One-Factor Congeneric Models for School-Level Environment Scales: Goodness of Fit Summary and Composite Reliabilities

Composite Variable	c^2	df	p	RMSEA	NNFI	CFI	GFI	AGFI	Factor Loading l^	Error q^
Student Support	3.186	3	0.036	0.043	0.962	0.971	0.979	0.963	0.571	0.031
Affiliation	2.903	3	0.147	0.019	0.991	0.999	1.000	0.998	0.480	0.042
Professional Interest	3.266	4	0.089	0.040	0.903	0.923	0.945	0.932	0.539	0.020
Mission Consensus	6.201	3	0.066	0.006	1.000	1.000	1.000	0.989	0.483	0.025
Empowerment	3.600	4	0.172	0.042	0.953	0.972	0.982	0.973	0.532	0.040
Innovation	3.412	4	0.283	0.053	0.901	0.950	0.969	0.951	0.510	0.024
Resource Adequacy	9.801	4	0.063	0.009	1.000	1.000	1.000	0.999	0.489	0.041
Work Pressure	7.310	3	0.049	0.043	0.950	9.998	1.000	0.998	0.501	0.030

RMSEA Root Mean Square Error of Approximation NNFI Non-normed Fit Index

CFI Comparative Fit Index AGFI Adjusted Goodness of Fit Index

N[a] The number of teachers with complete data GFI Goodness of Fit Index

Each of the factors has adequate fit on one-factor congeneric measurement models. The goodness of fit statistics results for the eight SLEQ scales and the four student composite variables are reported in Tables 2 and 3 respectively. All fit statistics met acceptable levels of fit for all of the composite scales in this analysis. The maximised reliabilities for all composite variables, both the composite reliability (r_c) and the Cronbach's Alpha (a^b) in the SLEQ were high (r_c = 0.892 to 0.931 and a^b = 0.706 to 0.885). For all four of the student composite variables these reliability measures were also high (r_c = 0.873 to 0.965 and a^b =

0.706 to 0.891) suggesting that all observed variables, both SLEQ and student observed variables were reliable measures of each of the proposed latent constructs.

A fitted congeneric model allows large numbers of observed variables to be reduced to a single composite scale and subsequently reduces the number of variables to be included in the structural equation models and other subsequent analysis . The overall results of the analysis of one-factor congeneric measurement models were used to compute composite scores for each latent construct using factor score regression (FSR) as a proportional weight to determine the composite scale (maximised) reliability . The Statistical Package for the Social Sciences (SPSS) was used and a command file was created to compute composite factor scores and maximised reliability for each latent variable using a weighted factor loading which attributes only that proportion of each item to a latent variable that should be attributed to the variable. Some items load more heavily on a latent variable than other factors and using estimates obtained from

Table 4

Fitted One-Factor Congeneric Models for Student Scales: Goodness of Fit Summary and Composite Reliabilities

Composite Variable	c^2	df	p	RMSEA	NNFI	CFI	GFI	AGFI	Composite Reliability r_c	Factor Loading $\hat{l}$	Error $\hat{q}$
Socio-economic Status	8.960	3	0.111	0.029	0.997	0.998	1.000	0.999	0.895	0.895	0.031
Attitudes Towards Mathematics	9.103	3	0.098	0.041	0.990	0.992	0.998	0.995	0.965	0.905	0.025
Success Attribution	3.240	2	0.211	0.001	1.000	1.000	1.000	0.999	0.873	0.876	0.010
Instructional Practices	6.802	7	0.141	0.019	0.936	0.981	0.998	0.996	0.925	0.978	0.032

RMSEA Root Mean Square Error of Approximation NNFI Non-normed Fit Index

CFI Comparative Fit Index AGFI Adjusted Goodness of Fit Index

N[a] The number of students with complete data GFI Goodness of Fit Index

Multifactor School Environment Model

It seems intuitively obvious that school environment structure is best represented by a multidimensional model, however, many researchers have chosen to aggregate all scores on environment items to obtain an overall mean environment score. By doing so, they imply that school environment is a unidimensional concept (the one factor model). The inappropriateness of such assumption and practice is investigated in this section. The eight sets of school environment items (56 items) are linked together to further test if the eight-factor environment structure can be simplified to a one single factor model, i.e. general school-level environment. To achieve this goal, the following two hypotheses were tested.

Hypothesis A: School-level environment is a one-factor structure

The one-factor model, whereby all 56 items were loaded onto a single factor 'general school-level environment', was tested first. The c^2 for this one-factor model was 2579 (398) and was significant at the 0.001 level, however, this is not surprising considering the large sample size (620 teachers). More informative is the c^2/df ratio, which for this model is greater than 3. In addition, the value of the Non-Normed Fit Index (NNFI) is less than 0.90 and the value of the RMSEA is greater than 0.05. Overall, these results indicate that a one-factor 'general school-level environment' model does not adequately fit these data.

Table 5

Goodness of Fit Summary – School-Level Environment Models

Model	c^2	df	c^2/df	p	RMSEA	NNFI	DFI	GFI	AGFI
One-factor model	2579	398	6.479	0.000	.068	.895	.944	.956	.949
Eight-factor model	780	301	2.591	0.000	.039	.959	.944	.982	.979
Fit change between two models			c^2=1799		*p*=0.000				

Hypothesis B: School-level environment is an eight-factor model

The second model to be tested postulates *a priori* that school-level environment is an eight factor structure, composed of Student Support (stusup), Affiliation (affil), Professional Interest (profint), Mission Consensus (mission), Empowerment (empower), Innovation (innovat), Resource Adequacy (resource) and Work Pressure (workpress).

The results presented in Table 5 indicate a better fit than the on-factor model of school-level environment. The c^2 change (1799) is significant at the 0.001 level, suggesting that the fit of the eight-factor model is significantly better than that of the one-factor model. All other indices of fit suggest that, school-level environment is best represented by the hypothesised, eight-factor model. A schematic portrayal of this eight-factor model is available from the author on request.

Convergent and discriminant validity

Convergent validity refers to 'indicators specified to measure a common underlying factor all have relatively high loading on that factor' . All standardised parameter estimates (factor loadings) in the eight-factor model are greater than 0.05 and these were all significant at the 0.01 level of alpha. These results suggest that convergent validity was supported by the data in this study. Discriminant validity refers to the distinctiveness of the factors measured by different sets of indicators – estimated correlations between the factors are not excessively high [. The estimated correlations among the eight factors are moderate from that between Innovation and Affiliation (0.501) to that between Student Support and Work Pressure (-0.181). These correlations are all low enough to indicate that the dimensions measure quite different aspects of the school-level environment and suggest that discriminant validity was supported by the data in this study. Again, due to the size of the eight-factor model all parameter estimates are available on request form the author.

In summary, it is evident from these analyses that school-level environment is a multidimensional construct, which in this study comprised the eight factors, Student Support, Affiliation, Professional Interest, Mission consensus, Empowerment, Innovation, Resource Adequacy and Work Pressure.

Fitting full structural equation models

When fitting the full models, the estimated composite variable regression coefficients and measurement errors (see Tables 3 and 4) were used as fixed parameters, as recommended by other researchers in the measurement part of the structural equation models. These estimates are not shown in the path diagrams in this thesis in order to reduce complexity. The only parameters to be estimated here were the structural part of the model, which includes the regression coefficients betas (b s) and gammas (g s). The adequacy of the models was evaluated using the goodness of fit statistics, which are presented with the structural models in Figure 2 and Figure 3.

The models were tested first including only the hypothesised pathways (both gammas and betas). If the hypothesised model did not fit adequately, a post hoc data analysis was then adopted to explore the possible relationships revealed in these data. Re-specification of the model was guided as much as possible by substantive considerations which means that paths were included only if they are supported by theory.

Results of the structural equation models

Tests of the initially hypothesised model for student achievement generated poor goodness of fit statistics, which are reported in figure 2 with the structural equation model. An examination of the structural results revealed several non-significant paths.

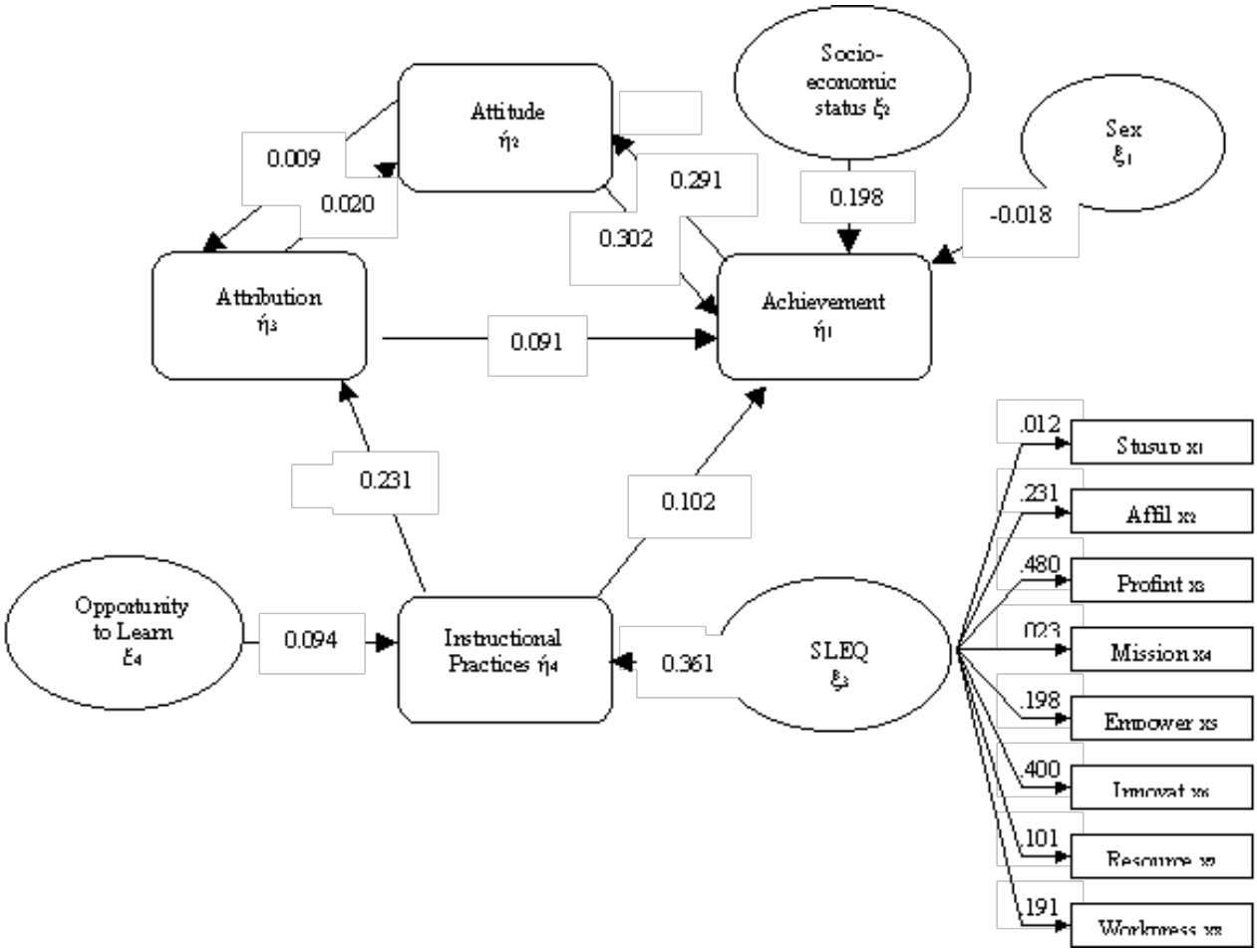

Figure 2. Hypothesised model of student achievement

The analysis identified that four of the relationships (depicted as paths on the models) between the dependent variables were not significant. The regression of attribution on achievement was positive but not significant (0.091), the regression of instructional practices on achievement was also positive but not significant (0.102), and the regression of both attitude on attribution (0.020) and attribution on attitude (0.009) were positive and not significant. The regression of the independent variable opportunity to learn on instructional practices was also not significant (0.094). Surprisingly, these data did not support the hypothesis that sex of the student had an effect on student achievement demonstrated by a non-significant regression (-0.018). As a consequence of the overall bad fit of the model a second hypothesised model was developed and tested on the basis of these results (see Figure 3).

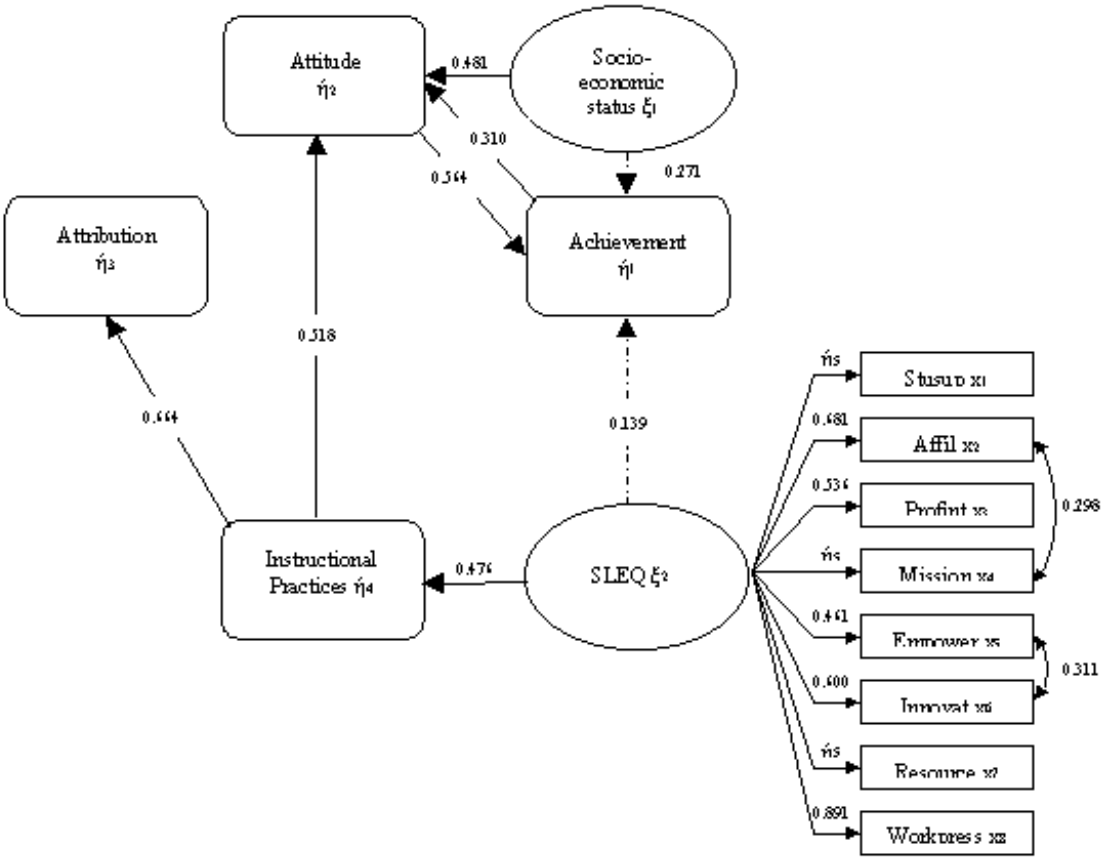

Figure 3. Final model of student achievement

The results of the modified model are presented in Figure 3 along with the goodness of fit statistics. This model is a good fit to the data indicated by the chi-square (c^2 = 176.34, *df* 51; *p*= 0.021) and supporting fit statistics (RMSEA = 0.54; GFI = 0.956). Although the estimates for Student Support (stusup), Mission Consensus (mission) and Resource Adequacy (resource) are not significant, these scales of the school-level environment were kept in the model on the basis of the results of the eight-factor and the post hoc analysis where to remove these three scales has a detrimental effect on the goodness of fit statistics.

The structural part of the model indicates several significant relationships some of which support previous research and those that add to the evidence of factors influencing student achievement. In this model, the relationship between attitude and achievement is recursive. Both effects are significant with the effect of student attitude towards mathematics strongly affecting student achievement (b = 0.564). The socio-economic status of students has a positive effect on student attitude (b= 0.481) resulting in an indirect effect on achievement (b = 0.271). Of particular interest in this study is the school-level environment and the effects on student achievement. These results show that there is an indirect effect of the school environment as perceived by teachers and student achievement (b = 0.139), a result of the school environment having a direct effect on the way teachers deliver the curriculum (b = 0.476). The better the environment for the teachers the more the instruction in classrooms is teacher-centred. The way teachers deliver the curriculum in their classrooms has a strong

and positive effect on student attitudes (b = 0.518), which has already been reported as having a significant effect on student achievement. The more teacher-centred the instruction the more positive the attitudes of students and the better the achievement is. These results support previous research that indicates students' respond to structured, directed, and clear instruction in classrooms. In this model, instructional practices positively affect success attribution (b = 0.664). There were no significant relationships between success attribution and other variables in the model. The model fit with these data did not allow for any paths showing the influence of success attribution.

Conclusion

The methodological process and outcomes of this research are both significant. The unique methodology used allowed for the variables used in the analysis and the investigation of influences on student mathematics achievement, to be developed in a manner that attributed only that proportion of the contribution to each variable from individual items. Previous studies have used a method of unit weighting which does not account for individual contribution. In addition, the development of the school-level environment scales is unique to this study and clearly demonstrates that the school-level environment is a multi-factor concept and not a unidimensional concept. In allowing the variables that contribute to the school-level environment to remain multi-factor in the analysis, it was possible to separate and distinguish exactly where the influences were and the strength and significance of those influences. The estimates resulting from the measurement models presented in this paper can be used in further research.

References