

PERCEPTION OF TEACHER EMOTIONAL SUPPORT AND
PARENTAL EDUCATION LEVEL: THE IMPACTS ON
STUDENTS' MATH PERFORMANCE

Thesis submitted for the degree of
Doctor of Education
at the University of Leicester

by

Kwong Yeung

School of Education

University of Leicester

August 2007

Acknowledgements

I would like to express deep gratitude to all the lecturers who guided me through this EdD program. Tremendous credit goes to my enthusiastic and competent supervisor, Dr. Glenn Fulcher. His constructive and swift feedback was not only critical to improving the content of my work, but also facilitated my progress on a quicker timetable.

My deep appreciation also goes to my department to provide me with Departmental Research Grant, GU-370, to consummate this dissertation.

Abstract

There is a paucity of research juxtaposing parental education level and teacher emotional support in a single study which examines their relative impacts on students' academic achievements. Therefore, the first objective of this dissertation is to study the influence of parental education level, in comparison to the influence of teacher emotional support, on students' math performance, by using more representative data and a rigorous statistical method. The second objective is to identify and examine how some important psychological traits (both affective and cognitive) mediate the effects of social factors on students' math performance. The third objective is to examine whether those relationships are moderated by gender.

Hong Kong's survey data is extracted from the Program of International Students Assessment (2003) as organized by Organization for Economic Co-operation and Development (OECD), on the math performances of 4,478 students at the age of fifteen. Measurement invariance was first tested, and then followed by Confirmatory Factor Analysis. Two structural models were tested by Structural Equation Modeling using Linear Structural Relations (LISREL) 8.5 which is computer software for SEM.

Results indicated that first, parental education level affects children's math scores by providing home education resources and enhancing children's math self-efficacy, and second the Self Determination Theory is applicable in supporting the hypothesis that teachers affects their students' math scores by providing a cooperative learning environment, which in turn, enhances students' affective and cognitive factors. Three important mediators, namely cooperative learning environment, math self-efficacy, and home education resources are concluded as significant mediating factors upon the effects of parents and teachers on students' math performance. The perceived support from parents and teachers are not significantly different across gender in Hong Kong. This is consistent with recent studies that differences favoring males in mathematics achievement are disappearing. Theoretical contributions and practical implications are discussed in the final part of the dissertation.

Key Words: *Social Economic Status, Confirmatory Factor Analysis, Measurement Invariance, Sobel Test, Variance Inflation Factor.*

TABLE OF CONTENTS

| | |
|---|----|
| List of Tables..... | 6 |
| Chapter 1: INTRODUCTION | |
| 1.1 Introduction..... | 8 |
| 1.2 Problem Statement..... | 9 |
| 1.3 Research Objectives..... | 12 |
| 1.4 Research Questions..... | 13 |
| 1.5 Limitations of the Study..... | 15 |
| 1.6 Significance of the Study..... | 17 |
| 1.7 Summary..... | 18 |
| Chapter 2: REVIEW OF RELATED LITERATURE | |
| 2.1 The Self-Determination Theory | 20 |
| 2.2 Introduction to the Hong Kong Context | 29 |
| 2.3 The Importance of Family Background | 33 |
| 2.4 The Importance of Teacher's Attributes | 38 |
| 2.5 Why Existing Data..... | 46 |
| 2.6 Structural Equation Modeling | 48 |
| 2.7 Measurement Invariance..... | 50 |
| 2.8 Development of Hypotheses..... | 55 |
| 2.9 Instrument Description..... | 60 |
| Chapter 3: RESEARCH METHODOLOGIES | |
| 3.1 Research Paradigms..... | 66 |
| 3.2 Sampling Design | 69 |
| 3.3 Questionnaire Design..... | 71 |
| 3.4 Method of Data Collection | 73 |
| 3.5 Data Analysis..... | 74 |
| 3.6 Reliability, Validity and Ethical Issues..... | 78 |

| | | |
|-----------------|------------------------------------|-----|
| Chapter 4: | FINDINGS OF THE STUDY | |
| 4.1 | Preliminary Analyses..... | 82 |
| 4.2 | Measurement Invariance | 89 |
| 4.3 | Confirmatory Factor Analysis | 94 |
| 4.4 | Structural Model Analyses..... | 98 |
| 4.5 | External Validity | 100 |
| 4.6 | Results | 102 |
| 4.7 | Discussions..... | 107 |
| Chapter 5: | CONCLUSIONS AND RECOMMENDATIONS | |
| 5.1 | Conclusions..... | 117 |
| 5.2 | Recommendations..... | 119 |
| 5.3 | Future Research..... | 121 |
| References..... | | 122 |

List of Tables

| | |
|---|-----|
| Table 1 – Participating schools for each sampling stratum in HKPISA 2003..... | 70 |
| Table 2 – Descriptive statistics for continuous variables..... | 83 |
| Table 3 – Exploratory factor analysis of seven constructs..... | 85 |
| Table 4 –Covariance matrix among items | 87 |
| Table 5 – Results of measurement validation..... | 88 |
| Table 6 – The Variance Inflation Factor for Testing Multivariate Multicollinearity.. | 90 |
| Table 7 –Common metric completely standardized solution for factor loadings and error variances under measurement invariance..... | 91 |
| Table 8 – The differences of latent means across gender in H.K..... | 92 |
| Table 9 – Goodness-of-fit indices for the measurement model | 96 |
| Table 10 – Discriminant validity: Chi-square differences between constrained and unconstrained models. | 97 |
| Table 11 – Goodness-of-fit indices of competing nested models..... | 99 |
| Table 12 – Goodness-of-fit indices for the different subjects..... | 101 |
| Table 13 – Relationships among constructs in the final model..... | 103 |

| | |
|--|-----|
| Table 14 – Test of hypothesized mediation effects..... | 106 |
|--|-----|

Chapter 1 INTRODUCTION

1.1 Introduction

Hong Kong students are ranked as one of the best performers in math, science and problem solving tests out of those living in forty countries or cities in the Program of International Students Assessment (PISA) 2003 organized by the Organization for Economic Co-operation and Development (OECD). OECD is internationally recognized as a reliable and comprehensive source of comparable economic and social data (OECD, 2003). Although these kinds of international studies provide us with valuable data, the factors causing these differences are generally not the focus of these studies and are simply unexplored. A detailed examination of the factors causing the differences in academic performances among Hong Kong students is therefore necessary.

Prior studies have stated that interactions of individuals with others in social contexts can affect the human development process (e.g. Bronfenbrenner, 1979). Wentzel (1998) suggested that students' learning processes are affected by their parents, teachers and significant others. Past studies have also identified families and teachers as major social factors affecting students' academic achievements (Ferguson, 1991; Sanders & Rivers, 1996). Hong Kong students who perceive their parents as supportive perform better than those students who perceive their parents as less supportive (Shek, Lee & Chan, 1998). Positive teacher emotional support mitigates students' anti-social orientation (Ma et al., 2000). If the students like their teachers, they will put in reasonable effort even if they do not enjoy their work (Close, 2001; Skinner & Belmont, 1993).

Nevertheless, many problems still remain unanswered with respect to the Hong Kong context. Firstly, how do these social factors affect students' academic achievements? Secondly, what are the relative impacts of these factors? Thirdly, will these effects hold true across gender?

1.2 Problem Statement

The equal education opportunity studies can be traced back to the status attainment literature from the United States in the 1960s. Blau and Duncan (1967) demonstrated that education and occupation are highly correlated and education can minimize the impact of Social Economic Status (hereafter as SES). SES is characterized as the economic, social and physical environments in which individuals live and work, as well as demographic characteristics they bear. ‘Socioeconomic status is probably the most widely used contextual variable in education research’ (Sirin, 2005:417). It describes an individual’s or a family’s ranking on a hierarchy according to his / her access to or control over some combination of valued commodities such as wealth, power, and social status. The tripartite nature of SES incorporates parental income, education and occupation (Sirin, 2005). Each domain should be regarded as a separate dimension from the others (Bollen, Glanville & Stecklov, 2001). Parental income reflects the potential for social and economic resources that are available to the student at home. Parental education tends to remain the same over time and is highly correlated with income. Parental occupation reflects the social and economic status, and the prestige and culture of a given socioeconomic stratum.

Published journal articles on SES from 59 independent samples ($k=59$) published between 1990 and 2000 showed a medium to strong SES-achievement relationship. For example, the average effective size (ES) is 0.35 for math achievement (Sirin, 2005). The strength of relationship is moderated if the students are: (1) high school students ($ES=0.26$, $k=22$); (2) with majority status ($ES=0.27$, $k=11$); and (3) from urban schools ($ES=0.24$, $k=13$). The strength of relationship is also moderated by the following methodological characteristics: (1) the type of SES components, like parental education ($ES=0.3$, $k=30$); (2) 3- to 7-point Likert scale ($ES=0.28$, $k=15$), (3) data source is derived from students ($ES=0.19$, $k=18$); and (4) achievement being measured is math ($ES=0.35$, $k=57$) where ES is the effective size and k =number of effective size. The effective size used in the meta-analysis review is the Pearson’s

correlation coefficient r (Sirin, 2005). Older students, students from two-parent households, and high-achieving students tend to report their SES more accurately than those in the opposite.

Prior studies found that the influence of SES is strong on students' academic achievement. Parents transfer their social, economic and cultural capitals to their children through education and therefore, make a difference in affecting students' academic performance (Bourdieu, 1986; Bowles & Gintis, 1976; Coleman et al., 1966). Besides, parents are found promoting the satisfaction of students' psychological needs in the school context which, in turn enhance students' ability to persist in the academic curriculum (e.g. Ratelle et al., 2005). They also influence the choice of school type, which, in turn influences the types of classroom environment, instructional arrangement, materials, teachers' teaching experience, class size, and parental-teacher relationship (e.g. Wenglinsky, 1998).

Educational inequality will persist due to the heritage of SES (Bourdieu, 1986; Coleman, 1988). Baker et al. (2002) analyzed the Third International Mathematics and Science Study (TIMSS, 2003) data and found the SES factor is still more important than the school factor regardless of the country's level of economic development. By the same token, 'the dynamic of parent-child and teacher-student relationships tend to change as adolescents seek autonomy from adults' (Chen, 2004:106) although subsequent studies still confirm that SES affects academic achievement significantly when the effect of school factors is controlled (Cohen & Rice, 1997; Schickendanz, 1995).

However, 'the magnitude of the relation between SES and school achievement is not as strong in the present review' (Sirin 2005:437) with average ES was reduced from 0.343 ($k=219$ studies) during 1918-1975 to 0.299 ($k=207$ studies) during 1990-2000. More importantly, the social environment changes over time. Parents become better-educated, family size is reduced and access to supplementary education such as TV and Internet becomes easier which, in turn, has accentuated or attenuated the impact of SES on academic achievement in recent years. 'Researchers must continue to

assess students' SES' (Sirin, 2005:443). Therefore, the family factor is worth re-examining.

Although the importance of SES is emphasized, teacher effects on student academic achievement should not be downplayed (Ferguson, 1991; Ferguson & Brown, 2000; Ferguson & Ladd, 1996). Indeed, many researchers argue that teachers play a more significant role (Reynolds & Teddlie, 2001). Teachers can not only provide emotional support, but also help develop a cooperative learning environment (Johnson & Johnson, 2004; Moriarty et al., 1995), which in turn, affects student's academic performance. According to Johnson and Johnson (2004), in cooperative learning situations students can perform 20 percentile better than competitive and individualistic learning students. They can activate their curiosity and intellect through discussion, learn from other perspectives, and give and receive support in small groups. Peer supports can reduce isolation and enhance students' own interest in learning. Leiken and Zaslavsky (1999) reported that if the teachers use cooperative learning activities, students' belief toward math will be increased.

Moriarty et al. (1995) indicated that cooperative learning leads not only to higher academic self-efficacy and achievements, but also to more appropriate learning behaviors. The impact of cooperative learning on cognitive and affective factors of students is well documented in the Self Determination Theory (SDT). Social contexts that promote autonomy, competence and relatedness will facilitate motivation (Deci & Ryan, 1991). Besides, Heyneman and Loxley (1983) argued that national Gross Domestic Product (GDP) would affect the relative weightings in importance of SES and school upon students' academic achievements. They had examined the data of the Second International Mathematics and Science Study (SIMSS) and found that (1) there is no linear relationship between maternal education level and students' academic achievements in three developing countries, namely, Thailand, Colombia and India; but (2) there is a significant positive relationship between maternal education level and students' academic achievements in three other countries, namely, Australia, England and Hungary. They concluded that school factors could better explain the variations in student academic achievements than SES could in some

countries. Maternal educational level as a prominent SES factor is directly correlated to students' academic performances, subject to the influence of GDP (Heyneman & Loxley, 1983).

All in all, schools are expected to lessen the impact of family background on student achievements so that students can be assessed fairly in terms of meritocracy. To support this goal, Hong Kong has been providing nine years of free compulsory education to all eligible students up to Form 3 (Grade 9) since 1978. The contradictory evidence on the relative effect of SES versus that of school in recent decades and its malleability justifies a re-examination. Having a better knowledge of the impacts of parental and teacher supports on student academic achievement can first answer how education levels off the playing field at entry for students and second, focus the direction of government policies on enhancing students' performances, especially in Hong Kong.

However, there is a paucity of research juxtaposing the parent and teacher in a single study in examining their relative impacts on students' academic achievements. Regarding the psychological factors, some researchers used self-concept as the only cognitive factor affecting the students and failed to include self-efficacy (Marsh, 1987; Marsh & Peart, 1988). Regarding the methodology, some research studies used small sample sizes (Chen, 2004; Kiamanesh, 2004) or multiple regression methods (Nonoyama, 2005) which may lead to unequivocal conclusions. To attenuate these biases, the development of a hypothetical model for studying the direct and indirect effects of the social factors on achievement using Structural Equation Modeling is necessary (Kiamanesh, 2004). This dissertation contributes to current literature by focusing on these gaps.

1.3 Research Objectives

The first objective of this dissertation is to study the influence of parents on student academic performances. The quality of educational opportunity is examined by

comparing the influence of teachers on student performances using more representative data and a rigorous statistical method.

Since math underlies every facet of science and technology (Hubbard, 1999), the second objective is to identify and examine how some important psychological traits (both affective and cognitive) mediate the effects of social factors on students' math performance. Teachers can affect students' learning motivations and behaviors and affect their performances as a result. Miserandino (1996) believed that when students enjoy learning, they will display interests, values and effort that gear them towards better performances, lower dropouts (Vallerand et al., 1989; 1993; 1997) and higher quality learning (Grolnick, Ryan & Deci, 1991).

The third objective is to examine whether those relationships are moderated by gender. Simply put, this study aims to investigate which social factors affect math scores in Hong Kong 15-year-old students, and in what way they are interrelated.

1.4 Research Questions

On the one hand parents have a strong influence on their children's academic achievements by transferring cultural resources to their children. Students whose parents were university-educated performed about two-thirds of a proficiency level higher than those whose parents had received no more than high school education. Students whose mothers had earned college degrees performed a full standard deviation in test scores better than those whose mothers had not (CME, 2004a). Parental education level is therefore used as a proxy to students' SES (Baker et al., 2002) and will be explained in more detail later.

On the other hand, teachers can help by developing a cooperative learning environment and the impact of a cooperative learning environment on cognitive and affective factors is well documented (Deci & Ryan, 1991). According to Self Determination Theory (SDT), if teachers can develop a cooperative learning

environment, they can arouse their students' academic self-concept and self-efficacy which can result in better academic achievement. Scaffolding occurs when learners are assisted by compeers in constructing knowledge according to the theory of proximal development (Vygotsky, 1978). Thus, the proposed model (Fig.1) represents the relationships to be examined in this study. Considering the objectives set forth in this dissertation and the time constraints, the research questions to be studied are as follows.

- 1) *Does parental education level affect student's math scores through the availability of home education resources and/or math self-efficacy?*
- 2) *Does teacher emotional support affect student's math scores through cooperative learning environment and/or math self-concept?*
- 3) *Does the effect of the support from parents and teachers on student's math scores vary by gender?*
- 4) *What are the relative impacts of parental education level and teacher emotional support on student's math scores?*

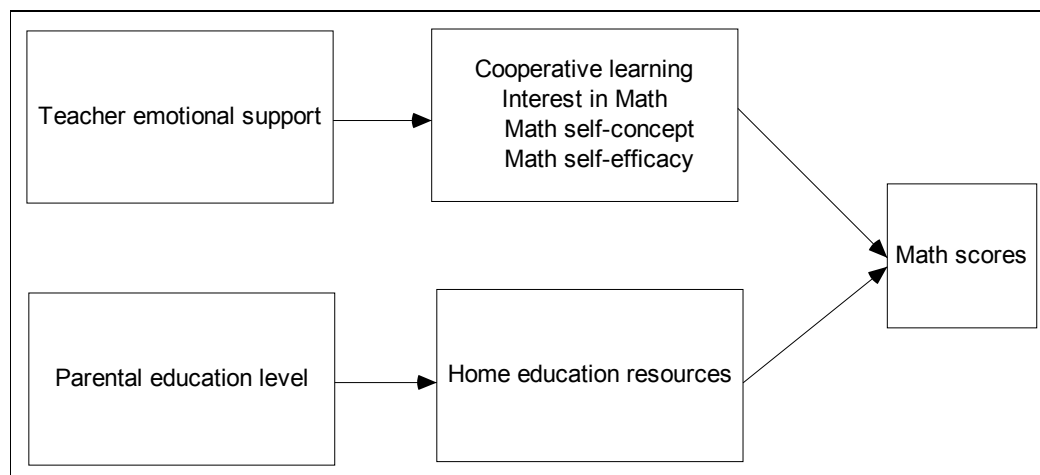


Figure 1. Proposed conceptual framework

1.5 Limitations of the Study

Five caveats need to be borne in mind in interpreting the results of this dissertation with respect to who, why, when, what and where.

First, constructs such as interest in math were based on 15-year-old students' self-reports, but not through direct measures such as in-depth interviews and observations (Boekaerts, 1999). These constructs, therefore, only offer a unidimensional perspective. Nevertheless, this approach is acceptable because we want to understand better how the social factors would affect students' math achievements from the perspective of students. There are multiple stakeholders in this field and the future research needs to consider the issue from multiple dimensions and collect information from various stakeholders such as parents, administrators and policy makers. The quantitative data examined in this dissertation can only reflect the perceptions of the students. Qualitative case study approach can be used to collect more in-depth and rich information in order to help understanding more about students' learning behaviors in the Hong Kong context.

Second, associations between variables do not equate to causal relations. For instance, a significant positive relationship between teacher emotional support and student performance can indicate the fact that better teacher emotional support enhances student performances. Equally, it can also indicate that high-achieving students demand more teacher support and prefer to have a better relationship with teachers. Besides, students may not do their best if the result does not count in formal scores of school records, which may then affect the validity of the math scores collected in the PISA 2003 study (Bandura, 1997).

Third, this dissertation has used the cross-sectional data in 2003 to test the association among variables. Such data only captured the perception of the Hong Kong students at a certain point in time. Although the results provide valuable insights in predicting the math performances, a longitudinal study is recommended to determine the

temporal dimension such as the reciprocal effect of self-concept and performances (Marsh, Martin & Hau, 2006).

Fourth, this study is limited to the testing of math scores because math does not only underlie every facet of science and technology (Hubbard, 1999) but also is one of the core subjects in focus in public examinations. Although judging students by their learning outcomes is only based on their academic performance, which is criticized for exerting stress and not being holistic in terms of assessments, it is less subjective and time-consuming, and an important indicator of students' school adjustment and future success (Chen, 2004).

Fifth, cultural factors will affect the respondents in terms of how they perceive and respond to the questionnaire items. Prior research has cautioned that self-reported characteristics are vulnerable to problems of comparability across different cultures (Schein, 1992; Van de Vijver & Leung, 1997). Cross-national comparison is hard to make for some variables like interest in math and instrumental motivation. As this dissertation generates a model based on the data collected from Hong Kong students, this may affect the generalizability of the results to other countries. Nevertheless, it is justifiable to focus on Hong Kong only because it is unique in history and affected by both British meritocracy and Chinese Wu Lun. The Hong Kong dataset consists of 4,218 valid cases which is representative enough for this complex model (Kline, 1998). Besides, the Hong Kong Education Bureau has been imposing many transformational changes which cause high pressure on schools, teachers and students since 2000 such as the mother-tongue education, new banding system, 3-3-4 academic structure, benchmarking test, as well as school evaluation and assessment, all of which are competitive but not cooperative in nature. It is important to know whether Hong Kong is moving on the right track.

1.6 Significance of the Study

The SDT proposes that autonomy, competence and relatedness lead to better motivation to learn (Deci & Ryan, 1991). Based on the SDT, theoretical constructs are developed to test the impacts of social factors, parents and teachers, on student achievements in Hong Kong. It is important to identify factors that contribute the most to the Hong Kong students' math performances. This can fill the existing gap of the research carried out in Hong Kong in this arena. After fitting a model to the Hong Kong context, we can validate this model across other subjects. With better understanding of their relative importance, it will shed light on how educational policies and school practices may address the differential resources that students bring from home so that educational administrators can spend money more wisely to enhance overall student performances. Simply put, this dissertation provides original contributions by juxtaposing important exogenous social factors, endogenous cognitive and affective factors and math performance in order to determine their relative weighting.

The SDT has been applied in many areas, such as treatment programs, religion, psychology, sports, physical exercise, political activity, health care, environmental activism and intimate relationships (Um, Corter & Tatsuoka, 2005). This dissertation is expected to provide significant theoretical contributions to the application of the SDT in educational management.

This dissertation is expected to provide practical implications to learning behaviors as well. Math and science have become more important as the economy grows but there is little research about the theoretical framework in relation to math learning specifically. Students perform better if they enjoy learning or can see that it helps them achieve their goals. Marsh, Hau and Kong (2002) recommended doing future research into individual student characteristics that mediate the effects of academic self-concept on subsequent achievements. Therefore, a rigorous statistical test on this theory can provide deeper understanding of how parents' education level and teacher emotional support can motivate students to learn and perform better. This underlies

the importance for school administrators and teachers to engage all stakeholders constructively with reference to the heterogeneity in students' psychological factors, gender and family background. This study can help policymakers to invest in approaches that address the aspects of attitudes and learning behaviors so as to enhance school-level performances.

This study has other strengths in relation to prior studies. Firstly, it focuses on Hong Kong, a non-western context, in contrast to the majority of research that has been conducted within the western context. Secondly, it extracts and analyses the dataset from an extremely large and nationally representative sample of over 260,000 adolescents. Thirdly, it focuses on reasoning abilities rather than curriculum-related scores. It is imperative to measure how well and how likely adolescents are to apply their math knowledge and skills in real-life situations rather than rote-learning. Fourthly, it juxtaposes and tests the SDT in a single model. This is important because families and teachers do not only exert a positive and significant influence on youth's perceptions and beliefs about their academic competency, but also on their actual academic achievements. Finally, it focuses on adolescents at the age of fifteen when the gender variations in academic achievement become evident (Entwisle, Alexander & Olson, 1994).

1.7 Summary

Children's learning processes are affected by their parents, teachers and significant others (Wentzel, 1998). Parents always play important roles by providing their children with resources which reproduce the social inequalities (Bourdieu, 1986; Coleman, 1988). Teachers can make a difference by providing their students a cooperative learning environment. Drawing on the SDT, parents and teachers also affect students' cognitive and affective factors to make a difference (Deci & Ryan, 1991; Ratelle et al., 2005). However, there is a paucity of research juxtaposing social factors in a single study to examine their relative impacts on students' academic achievements. Therefore, the objective of this dissertation is to develop and test a

model based on extrapolating the effects of social and motivational resources on math performance.

This dissertation used 4,478 Hong Kong cases as gathered by OECD in 2003. It is organized into five chapters. Chapter 1 presents problem statements, research objectives, research questions, limitations and the significance of the study. Chapter 2 presents the general education background of Hong Kong and reviews relevant literature about the impacts of the social factors upon students' academic achievements such as family background and teachers. Parental education level and teacher emotional support are analyzed in comparison to math scores. Chapter 2 also examines how these impacts take place by using the Self-Determination Theory. In chapter 3 the research methodologies of this study will be discussed thoroughly. Chapter 4 will present the findings and discussion. Conclusions and recommendations will be presented in chapter 5. Reference and appendices are attached at the end. This dissertation merges the themes emerging from the socio-cultural, educational and psychological contexts.

Chapter 2 REVIEW OF RELATED LITERATURE

2.1 The Self Determination Theory (SDT)

The SDT is a macro theory of human behavioral motivation in which choice or human autonomy are studied (Deci & Ryan, 1985; 2000). It proposes that human psychological needs are the basis for their motivations. In order to enhance motivations, three needs have to be fulfilled: (1) autonomy (DeCharms, 1968); (2) competence (Harter, 1978); and (3) relatedness (Baumeister and Leary, 1995; Levesque et al., 2004).

An autonomous choice is defined in terms of the experience of feeling of volition and thus the decision maker owns an internal perceived locus of causality (DeCharms, 1968) for the selected options. The behavior is self-directed. Competence is the need to experience a sense of accomplishment upon improvement of one's abilities (Deci & Ryan, 1985). Relatedness refers to the need to feel related to one's significant others (Levesque et al., 2004). People will act accordingly in order to meet these three needs (Deci & Ryan, 2000).

Prior research supports this reasoning. Researchers found that students' perceived autonomy (Grolnick, Ryan & Deci, 1991), competence (Harter, 1978) and relatedness would all affect their motivations. Levin and Long (1981) claimed that motivated students are more cooperative, open to the learning material, work longer, harder and with more rigor and intensity, and get more out of the instruction. Relatedness had received less attention in general and PISA did not include any questions about relatedness, and therefore, this element is excluded from this dissertation.

2.1.1 Autonomy

In order to effectuate a high degree of autonomy, the actor must perceive a true sense of choice which requires a decision to be accompanied by the experience of endorsement and willingness. Many researchers have studied the factors affecting

autonomy perception (Deci & Ryan, 2000). If the administrators provide a reasonable rationale to back up a praiseworthy behavior, show responsiveness to the actor's feelings, and allow freedom of choice, then autonomy is enhanced which leads to the better well-being of students (Ryan & Deci, 2000). True sense of choice must be clarified here in order to avoid misunderstanding. Providing one with many choices may not guarantee autonomy. If a person is given only one option but this option is endorsed by him or her, then this person still has a high degree of autonomy.

On the other hand, if the person is given many trivial options to sort through, he or she still cannot experience autonomy. Zuckerman et al. (1978) studied the impact of the availability of choices on the persistence of solving puzzles. One group is given the freedom to choose from three out of six puzzles to solve and the other group has no choice. As expected, they found that the group with choice is more intrinsically motivated than the group without choice. However, Baumeister et al. (1998) argued that choice making is ego-depleting or tiring. To screen through and to choose from the options will use up psychological resources. Iyengar and Lepper (1999) suggested that to make a purchase decision out of large number of options is disheartening. They tested two groups of participants with different options. One group is given two dozen options and the other is given six options. The result is that the group with two dozen options is less likely to buy any of the offered products. They concluded that more options do not imply greater autonomy, especially when the situation is overwhelming and the differences are trivial.

Similarly, coercing people to choose one out of many options is also de-motivating. Feeling forced to decide is antithetical to the experience of choice making – choose not to select. Iyengar and DeVoe (2003:130) stated thoroughly that ‘we believe that pseudochoice (i.e. misleading people to believe that they have choices while in fact they do not), excessive options (i.e. giving people too many options to sort through), and forced decision making (i.e. coercing people into decision making), all of which are considered as ‘choices’ by some writers but they are unlikely to result in the experience of choice, and are unlikely to enhance volition or autonomous motivation.’

Besides the above-mentioned issues, the degree to which choices reflect the person's needs, interests, or values are critical to determining whether it is autonomous versus controlled. People choose options which are consistent with their personal values. To sum up, in order to achieve autonomy and enhance motivation, the prerequisite is to attain a true sense of choice. To feel the true sense of choice, a person should be able to choose without any pressure in the decision making process and value the options that meet the personal goals.

Contextual support can also affect students' autonomy. Only when the learning environment is supportive and cooperative can autonomy be improved and student's interest in learning be aroused (Deci & Ryan, 1991). A positive attitude towards school and education is necessary for higher educational aspirations and success (Stodalsky, Salk & Glaessner, 1991). If the school administrators provide reasonable rationales for a praiseworthy behavior, show responsiveness to the actor's feelings, and allow freedom of choice, then autonomy is enhanced which in turn leads to the better well-being of students (Ryan & Deci, 2000). Moriarty et al. (1995) reported that cooperative learning environments are not only conducive to higher self-efficacy and achievement, but also to more appropriate learning behaviors.

Cheng (1996) studied different factors related to the effectiveness of Hong Kong schools like leadership, teacher effectiveness, environmental characteristics, organizational culture and school change. With respect to Hong Kong secondary schools, factors like teachers' instructional leadership, modes of instruction, teacher attitudes towards work, teacher efficacy in teaching and teacher stress in work were studied (Lo et al., 1997). They found that teachers and students had a general agreement on the importance of providing incentives for individual students in terms of instructional leadership only, but otherwise their views were conflicting. With respect to Hong Kong primary schools, Cheng, Cheung and Tam (2002) found the following:

- Students from more effective schools perform better in Chinese language, English language and math than those from less effective schools.

- The school climate of more effective schools has closer affiliation, better classroom order and organization, better teacher support and stronger teacher control than those of less effective schools.
- Teachers' teaching practices and the consistency of their behaviors affect the school effectiveness, which is aligned with the view of Creemers, Scheerens and Reynolds (2000).
- Teacher-centered methods may have more direct impact on students' academic achievements than creative thinking and motivation.

Scheerens and Creemers (1989) proposed a multi-level model of schooling which incorporates three organisational levels: the student, the classroom and the school, and the context of the school. Schools and classrooms in particular, can moderate the effects of students' SES on their academic achievements. Chan and Cheng (1993) studied 700 teachers to explore the relationship between principals' instructional leadership and teachers' affective and attitudinal outcomes in terms of teachers' sense of efficacy, sense of community, and professional interest. They found that promoting instructional improvement and professional development have significant and positive relationships with teachers' affective and attitudinal outcomes.

Experienced teachers are believed to present subject materials in smaller portions, ask students more questions, and display better classroom management skills. Novice teachers are believed to adopt a laissez-faire approach (Tam, 1997). The expert teachers used more pictures, word cards, songs, tape recordings and activities to help learning in elementary Chinese lessons. They encouraged students to ask questions, tell stories and make presentations. Effective teachers encourage their students to listen attentively and participate actively. They interact with their students in small groups, plan the lesson carefully, and use timed questioning and associated activities to help teaching (Biggs & Watkins, 2001; Lee, Lam & Li, 2003). On the other hand, less effective teachers: (1) adopt teacher-centered approach which results in poor teacher emotional support and student learning attitude; (2) are incapable of using different pedagogies to cope with the diversity of learner interests and abilities; and (3)

are incapable of using appropriate questioning technique to create lively debates and inspire high-order thinking (Education Department, 2001).

Johnson and Johnson (2004) believed that teachers can do five things to develop cooperative learning in math to make a difference: (1) show clearly the positive interdependence among each group in achieving the goal; (2) ensure that all group members finish their math assignments and promote the learning of their group mates; (3) ensure that students can handle interpersonal and small group skills; (4) ensure that students can interact face-to-face while completing math assignments; and (5) ensure that group members would review their own progress periodically. It is claimed that these suggestions require a lot of teacher-student interactions and close teacher-student relationship. More importantly, Chinese society has a fine tradition of respect for teachers. 'Teachers deserve our respect and recognition for their many toils nurturing talent for the community. We rely on our fine professional teachers to implement quality education in Hong Kong' (Policy address by Chief Executive, 2005).

2.1.2 Competence

The SDT argues that people must have the perception of a true sense of choice and perceived competence to be intrinsically motivated. Perceived competence has two components: self-concept and self-efficacy. Interest in math is translated into the self-concept, which has direct and indirect effects on academic performances (Williams, Freeman & Deci, 1998). Self-concept is perhaps the basis for all motivated behaviors. Kiamanesh (2004) found that self-concept would significantly affect students' math scores. Marsh (1987) believed that self-concept and achievement relationships are recursive. Koutsoulis and Campbell (2001) even recommended the improvement of students' academic self-concept as a basic educational outcome for academic achievement.

House (2003) believed that self-efficacy affects grade performance in science courses. Pajares and Kranzler (1995) found that general mental ability has a strong impact on self-efficacy and subsequently both elements have strong and direct effects on math

performance. Moriarty et al. (1995) reported that cooperative learning environments would lead to higher self-efficacy and achievements. In this section, these two beliefs are compared and contrasted.

Purkey (1988) referred self-concept as a complex system of learned beliefs, attitudes, and opinions that each person holds. Self-concept is the self-perception formed through interaction with the environment whereas self-efficacy is the belief in one's ability to make a difference (Bandura et al., 1999). Self-concept is a more global self-judgment which includes beliefs of self-worth associated with one's perceived competence. Self-efficacy acknowledges the diverse nature of human abilities in accomplishing tasks. These two beliefs can be differentiated by asking different questions. The self-concept belief is manifested by asking the 'being' and 'feeling' questions (e.g., "I get good marks in mathematics." "I learn mathematics quickly." "In my mathematics class, I understand even the most difficult work."). Pajares and Schunk (2002) stated that the self-efficacy belief is manifested by asking the 'can' or 'how confident' questions (e.g., "How confident do you feel in solving an equation like $3x + 5 = 17$?"). They also believed that self-efficacy is based on students' prior mastery experience and its effects are revealed through four major processes: cognitive, motivational, affective and selection. People with high self-efficacy are more willing to try harder in achieving more challenging goals (Bandura et al., 1999).

Academic self-efficacy focuses on one's perceived ability to attain academic achievements. It includes the student's belief in their ability to master the subject; to manage learning activities; and to achieve academic expected outcomes. It affects the student's commitment in academic and social activities (Bandura et al., 1999). Academic self-concept focuses on academic components of self-concept and can be further divided into English, science or math self-concepts (Marsh, 1987). Many researchers agree that academic achievement is also related to academic self-concept, though the causal sequence may not be in any specific order.

2.1.3 *Math self-concept and math scores*

Self-concept is ‘a description of one’s own perceived self accompanied by an evaluative judgment of self-worth’ (Pajares & Schunk, 2002:243). It is developed in early childhood and perceived as the cornerstone of both social and emotional developments of the children (Davis-Kean & Sandler, 2001). General self-concept at the top is divided into two components: academic and nonacademic. The academic component is divided into (i) self-concepts specific to general school subjects like math, science, English and (ii) non-academic self-concept which is sub-divided into the physical, social, and emotional components. These non-academic components can be further broken down into more specific components (Marsh & Craven, 2006). Academic achievement is only positively correlated with specific academic self-concept but not with nonacademic or general self-concept, e.g. math academic achievement is positively correlated with math self-concept. Academic self-concept is formed by comparing externally with other students’ performance and by comparing internally with one’s other academic self-concepts, e.g. compare students’ math self-perceived ability with their verbal self-perceived ability.

There are competing views about the causal relationship between academic self-concept and academic achievements. The self-enhancement model (Shavelson & Bolus, 1982) stated that academic self-concept affects academic achievements. The skill-development model (Skaalvik & Hagvet, 1990) claims the opposite. The reciprocal effects model (Marsh & Yeung, 1997) compromised the two conflicting authorities above by suggesting that a prior academic self-concept affects subsequent academic achievements, the latter of which would in turn affect self-concept as a result. Marsh and Craven (2006) believed that in the reciprocal-effects model (REM), self-concept and achievement are inter-related in a dynamic and reciprocal way. People believe that students can perform better, if they are efficient, confident and competent (“I believe, therefore, I am”). In REM, self-concept is hypothesized to have a multidimensional and hierarchical structure.

Skaalvik and Hagvet (1990) argued that the significant relationship between self-concept and achievement was found in the older-student cohort. Guay, Marsh and

Boivin (2003) rejected this argument by using three different age cohorts and found that stronger support lies within the self-enhancement model rather than skill-development model amongst all three age cohorts. The self-enhancement model is invariant across different ages. Marsh, Hau and Kong (2002) further proposed that individual student characteristics, e.g. gender, would mediate the effects of academic self-concept on achievement and that this finding has important theoretical implications.

Marsh and Peart (1988) studied how to improve high school students' physical fitness and physical self-concept. They formed three different groups, namely competitive, cooperative learning and non-treatment. Cooperative learning groups work in pairs and focus on individual improvements. Competitive groups work alone and focus on comparing with the best performers. They found that under cooperative learning interventions, physical fitness and physical self-concept had both increased. No significant gain was observed for academic self-concepts. Under competitive intervention, while physical fitness was increased, physical self-concept had decreased. More losers than winners are created.

Many researchers found that high self-concept can reduce anxiety, enhance motivation and mitigate disappointments or negative feedbacks after poor performances (Sommer & Baumeister, 2002). Low academic self-concepts coupled with failure cause poorer performance and slow recovery. After demonstrating the importance of academic self-concept, Marsh (1987) argued that prior self-concept effectuates upon academic achievements and affects subsequent self-concepts. The students are of the average age of fifteen years and eight months in PISA survey (OECD, 2003). They are assumed to be old enough to have strong self-concepts. Thus, math self-concept is hypothesized to have a positive association with interest in math and math scores.

2.1.4 Math self-efficacy and Math scores

Bandura defined self-efficacy as 'people's judgments of their capabilities in organizing and executing courses of actions that required attaining the designated

types of performances' (Bandura, 1986:391), that is, what people believe can make a difference in their capabilities. Pajares and Valiante (1999) found that even when the skills and knowledge are of the same standard, people with stronger self-efficacy within themselves can outperform those with weaker self-efficacy significantly, because a strong sense of self-efficacy induces individuals to attempt more challenging tasks, (Bandura & Schunk, 1981), persist longer in adverse situations (Bandura & Schunk, 1981), self-regulate better than others (Zimmerman & MartinezPons, 1990) and use more effective strategies (Pintrich & De Groot, 1990). Therefore, according to Bandura's Self Cognitive Theory, self-efficacy mediates the exogenous social factors, self-concept and concomitant performances.

Many researchers found that self-efficacy predicts and mediates various motivational variables on achievement. For instance, Randhawa, Beamer and Lundberg (1993) pointed out that self-efficacy ($\beta=0.32$) can predict math achievements. Pajares and Miller (1994) also concluded that self-efficacy significantly affects self-concept ($\beta=0.16$) and math achievement ($\beta=0.55$). Self-efficacy also mediates the effects of gender on math self-concept and achievement. Pajares and Graham (1999) believed that math self-efficacy is more important than self-concept in affecting math achievement. However, not all research displays this relationship. Norwich (1987) found that self-concept and prior math achievement affect math achievement but not self-efficacy. Marsh (1987) also believed that math self-concept has a positive and significant effect on math performance.

2.1.5 Motivation

Motivation is one of the mediators mediating the effects between self-concept, self-efficacy, and achievement. The SDT divides motivation into two broad categories: intrinsic and extrinsic. One is intrinsically motivated if one has a sense of choice and volition without any pressure upon decisions over any actions. One can do anything if one finds that interesting or important to him/ herself, or one can do something that is controlled by external forces. For instance, if a man cooks because he likes to do it, he is intrinsically motivated and so he has the autonomy to perform. If he is forced to do so because of the demands from his wife or neighbors, then he is motivated

extrinsically. The SDT argues that behaviors as motivated by autonomy have greater impacts and last longer than those as motivated by control.

Past studies support this reasoning. Mullis et al. (2000) found that students with more positive attitudes perform better in math. Papanastasiou (2002) concurred that the same phenomenon applied in Cyprus, Hong Kong and the U.S. where students like math, and hence they perform well. Ramserier (2001) found that self-determined motivation is positively related to math achievements in Switzerland. In regard to the antecedents, Vallerand (2001) and Vallerand and Losier (1999) proposed that social factors would affect students' motivations, and this will be discussed in the next section.

2.2 Introduction to the Hong Kong Context

It is imperative to study the Hong Kong context for three reasons: (1) in PISA 2000, Hong Kong students rank among the top ten best performers in four academic domains (ranked 1st in math, 2nd in problem solving, 3rd in science and 10th in reading). Single cases may be used to confirm or challenge a theory, or to represent a unique or extreme case (Yin, 1994); (2) Hong Kong has a unique history as marked by rapid political, cultural, economic, and social transformation. Hong Kong reverted to China in 1997; and (3) author's personal interest in Hong Kong education development. In the 156-year colonial era, the Hong Kong education system was affected by the British culture on the one hand and the traditional Chinese one on the other hand. This section will discuss the educational contexts in Hong Kong. The situation in Hong Kong seems competitive and oppressive rather than cooperative and collaborative.

Government

Education, health welfare and social welfare are the largest public expenditure items in Hong Kong. The Government spent HKD 33,873m on primary and secondary education in 2005/2006 which was around 16% of the total expenditures (Budget Speech, 2006). 'To help the economy power ahead, we must upgrade the quality of the local workforce to effectively cope with competition brought by globalization. A

quality workforce is more than a deciding factor in economic development. It also helps in creating social harmony. We place special emphasis on education, training and retraining. The resources put into education are higher than those in many advanced societies. For the new senior secondary education and higher education reform, we plan to introduce the '3-3-4' academic structure in September, 2009. We must upgrade our biliterate and trilingual proficiency... all teachers of English and Putonghua are required to reach the language benchmarks by September 2006' (Policy address by Chief Executive, 2005).

Parents

Parental support for children is also evident. Stevenson and Lee (1990) found that Hong Kong parents invest funds, space, and time to improve their children's academic performance. They do not only provide them with resources conducive to learning, but also devote time to supervise their children's schoolwork. Economic restructuring after 1997 reinforces education as a means to achieve upward social mobility (Blasko, 2003). Prior research has identified the relationship between education and occupation. There is no exception for Hong Kong especially during the times of economic recession. People with lower levels of educational attainment suffer more than people with higher levels of educational attainment. Societal culture affects parents' childrearing practices. Chinese culture is collective, rather than individualistic as in western cultures. Filial piety and parental expectation are the major motivators to higher achievement (Hofstede, 1991). Stevenson and Lee (1990) concurred that Chinese children need to succeed partly due to filial obligations and the duty of not disgracing their parents. Their successes are believed to be a direct reflection of their parents' wishes.

According to Bronfenbrenner (1979), micro-system (family) and meso-system (school) are affected by the macro system (society). Chinese societal culture is deeply affected by Confucian ethics which emphasize obedience along the lines of five cardinal relationships (Wu Lun): between ruler and the ruled, parents and children, husband and wife, elder brother and younger brother, and between two friends. Therefore, Chinese parents are characterized as 'restrictive', 'controlling' and 'authoritarian'

(Chao, 1994). However, on the positive side, they provide their children close supervision and care which in turn improves their academic achievements.

Students

The education system in Hong Kong puts tremendous pressure on students due to two characteristics: the banding system and meritocracy-based assessment. The Hong Kong education system comprises of three years of preschool learning (Kindergarten 1-3); six years of primary school (Primary 1-6); five years of secondary school (Form 1-5); two years of postsecondary school (Form 6-7); three or four years of undergraduate school; two years of graduate school and finally three years of doctoral study. All Hong Kong students between the ages of six and fifteen receive free compulsory education. Students are assigned to secondary schools according to their examination results. Starting from 2001, higher achievers are assigned to band-one schools; moderate achievers are assigned to band-two schools and low achievers are assigned to band-three schools. This allocation approach differentiates the Hong Kong education system from other education systems like the U.S. one. Besides, advancement to senior forms mainly depends on the public examination results in Hong Kong. Only a subset of students is promoted every year. 90% of the Form 3 students are promoted to Form four. They can continue their studies up to Form five.

After taking the Hong Kong Certificate of Education Examination (HKCEE, 2006), 30% of them can proceed to Form 6 and study for two years of matriculation program. After taking the Hong Kong Advanced Level Examination (HKALE), 18% of them can enroll in one of eight universities. That means only around 5% of Form 3 students can finally enroll in local universities. Limited space available in tertiary institutions exerts more pressure on Hong Kong students to compete and succeed academically than the U.S. students, who have many more universities to choose from. Higher anxiety associated with more pressure undoubtedly undermines academic self-concept, self-efficacy and interest in learning.

Teachers

In Hong Kong, on the one hand, teachers receive respect from their students due to Confucian Wu Lun ethics. They have high status in the society to serve as role models and knowledge dispensers. On the other hand, they face the burden of overloading and high stress. On 22 January, 2006, over 10,000 teachers - one-fifth of the entire teaching workforce of primary and secondary schools- protested against overwork and stress caused by the government rapid reforms. Teachers are working overtime by more than 31 hours per week and facing a 1% cut in education spending by the government. According to the Hong Kong Professional Teachers' Union, teachers' burnout, stress and anxiety are related to the policy of downsizing and school closures, putting schools and teachers under surveillance, introducing external and internal evaluations, and refusal for the implementation of small-class teaching (PTU News, 2006).

Professor Cheng Yin-cheong of the Hong Kong Institute of Education described the recent reform as 'the Great Leap Forward of Education'. The first wave of the Great Leap Forward was the school downsizing and closures, which triggered off destructive competitions among schools. These in turn changed teachers' lives tremendously, causing unbearably heavy workloads and feelings of alienation which had adversely affected the quality of education. The second wave was the sweeping curriculum reform and the introduction of school-based assessments, which had further aggravated teachers' plights and completely gone against the well-being of students (PTU News, 2006).

On the other hand, in the interim report submitted by the Committee on Teachers' Work (CTW) to the Secretary for Education Bureau, the chairman argued that there was no direct relationship between teachers' pressure and their workload. The pressure was attributed to administrative work. It was the school administrators who created unnecessary administrative work for the teachers not the education reforms. Incremental administrative work was evident on both reports and indubitably the supports provided by teachers to students would be undermined.

Schools

The fear of school closure, together with the stipulation of mother-tongue education, causes fierce competition in student recruitment among schools. For instance, more than 130 primary schools will be closed over the next three years due to the fact that their number of Primary One applicants is below the lower limit of 23. In order to survive, they may apply for Quality Assurance Inspection in order to validate their performance.

Besides, school principals also challenge that assessing schools merely on the basis of academic performances contradicts with the trends of education reforms. Some of these 31 schools have formed a “Joint Action Committee to Protect Schools’ Rights of School Governance and Recruiting Students” for the sake of fighting for their rights (PTU News, 2004). In response to these complaints, the Secretary for Education Bureau explained that the school closure policy was due to the declining birth rate and the choices made by the parents. The Bureau has allocated two times more teachers to band-three secondary schools. An additional HK\$1.76 billion will be spent annually. 1,400 teaching posts and 4,000 non-teaching posts for counseling will be created. It was claimed that to implement small-class teaching immediately without waiting for the result of the undergoing pilot scheme is not good for the students, but if that is successful, it may ensure full employment for teachers (Chong, 2006).

2.3 The Importance of Family Attribute

Out of the above-mentioned variables, parents play a more important role in affecting student academic performance by transforming various forms of capital into the children’s human capital (Bourdieu, 1986; Coleman, 1988). They provide educational resources and a learning environment at home, serve as a role model and hold particular values and beliefs towards education.

Bronfenbrenner (1979) believed that parents play a significant supportive role in a child’s social development. Whether construed as home-based support like monitoring homework, school-based support like attending school events, or teacher-

based support like parent-teacher meetings, parental involvement is found to affect children's academic achievement positively and significantly (Hill & Craft, 2003). By the same token parental involvement contributes to better psychological outcomes such as higher efficacy (Grolnick, Ryan & Deci, 1991; Hoover-Dempsey et al., 2005). Hoover-Dempsey and Sandler (2005) proposed four types of parental involvement: (1) more support like providing academic guidance and comfort; (2) more control like keeping children at home; (3) higher aspirations like future career; and (4) stricter regulation, like homework completion monitoring. Many researchers have specifically pointed out that parents provide more support to girls than to boys (Carter & Wojtkiewicz, 2000; Eccles, Jacobs & Harold, 1990). For example, girls discuss more about their academic-related matters with their parents than boys do. Hoover-Dempsey and Sandler (2005) also reported that parental involvement is affected by three factors, namely role construction, efficacy to make a difference and demand from others.

Parental role construction is defined as 'parents' beliefs about what they are supposed to do in relation to their children's education and the patterns of parental behavior that follows those need' (Hoover-Dempsey et al., 2005). In fact, there is an indigenous adage in Chinese culture that defines the parental role in children's learning process, "to rear but not to educate is the fault of fathers". Parents with stronger efficacy in helping their children to be successful in school are more likely to get involved in their children's learning activities (Bandura et al., 1999). Demand or invitations from schools, teachers and children enhance parental involvements. According to Cameron and Ettington (1988), school clan climates emphasize harmonious interpersonal relationships and consensus, humanistic, collaboration and corroboration which encourage parents to become involved. Teacher invitations enhance parents' sense of being welcome to participate in the teaching processes. Parents believe that teachers really care about their children and take their suggestions and ideas into considerations. Invitations from children activate many parents' wishes to be responsive to their children's school success, thereby also insulating themselves from the repercussions of failure.

Nevertheless, variations of resources limit parental involvement. Parents with lower educational levels work longer and harder. They have less time, energy, school-related skills and knowledge, but higher stress and depression. They tend to ask others to help particularly when their children's subject matter supersedes parents' knowledge or if they have multiple child-care, elder-care or related family responsibilities (Hoover-Dempsey et al., 2005). Weiss and Krappmann (1993) echoed the opinion that families develop students' personality, help the children to stay in contact with society directly and influence them through the domestic resources indirectly. Baker et al. (2002) used parents' education level and the number of books read to show the significant relationship between SES and children's educational achievements.

Blau and Duncan (1967) linked the prestige of father's job and years of father's education to his son's years of education and the status of first job for the latter. They concluded that the impact of father's job and education on his son's occupation was significant. Treiman (1977) found that prestige hierarchies were basically static temporally and spatially. All complex societies have similar occupational status hierarchies characterized by division of labor. Ganzeboom and Treiman (1996) used data from 16 countries to develop the International Socioeconomic Index (ISEI), which measures the attributes of occupations that explain the relationship between a person's education level and income. Occupation-related income may increase learning opportunities and resources. Different occupational skills may affect children's skill development and beliefs about the practicability of math in their future career. In fact, students whose parents' occupations require advanced math skills, like engineering and physics, performed almost one proficiency level better than those whose parents' occupations do not require advanced math skills, like law, after controlling for the education levels and income (CME, 2004a; 2004b).

In this study, although the tripartite nature of SES incorporates parental income, education and occupation (Sirin, 2005), only parental education was used as a proxy to SES in this study. Parental occupation was not considered because it is a different dimension from that of parental education (Bollen, Glanville & Stecklov, 2001;

Hauser & Huang, 1997) and can be excluded in this study. Blasko (2003) found that the impact of parental occupation becomes statistically insignificant if economic and cultural capitals are included in the model. Parental income is not considered either because it is difficult to obtain from students in real-life international studies, unless separate questionnaires are given to the parents to fill in, and it is unfair to make comparisons due to the differences in inflation rates among different countries or cities. Wealth data is also difficult to collect and calculate (Filmer & Pritchett, 1999; Wong, 1998).

Moreover, choosing the parental education level as a proxy to SES is consistent with prior studies (Baker et al., 2002). For example, out of 59 journal articles published during 1990-2000, 30 of them used parental education; followed by parental occupation ($k=15$) and parental income ($k=14$) as SES component (Sirin, 2005). Students whose mothers have earned college degrees perform better than those whose mothers have not by a full standard deviation in test scores; and teenage girls in poverty are more likely to become unwed mothers (Mayer, 1997).

While many researchers found that parental education level has a significant relationship with their children's educational achievements (Baker et al., 2002; CME, 2004a; Ferguson, 1991), others argue that one-quarter of the high achievers at the top 15% percentile had parents with at most a high school education standard only and one-quarter of those at the lowest 15% percentile had at least one parent with a university degree. Besides, the impact of family background tends to change as adolescents seek autonomy from parents (Chen, 2004). Therefore, the impact parental education level on students' math performance should be analyzed critically.

There are two sets of standard categories of educational attainments which are widely used internationally. The International Standard Classification of Education (ISCED) categories were developed by UNESCO. It is a general measure of educational attainment. The U.S. Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) was developed by the University of Mannheim to differentiate general and academic credentials from vocational ones. This classification is good for

countries with significant education-occupation associations, like the UK and the Netherlands. PISA has adopted the ISCED categories (OECD, 2003).

Gender disparity in parental involvement

According to Wang (2001:255), girls are valued lower than boys in Chinese culture. In a traditional Chinese marriage, the wife is supposed to move to her husband's domicile after the wedding. Females are regarded as a lost asset to their native families once they get married since she needs to serve her husband's family. Female inferiority is in line with the Confucian ethic "honorable males, inferior females". Girls in China are not doing well in traditionally male-dominated areas such as mathematics and science regardless of the location of provinces, namely Gansu, Guangxi, Hubel, InnerMongolia, and Jiangshu. Girls do not only do more household chores than boys, but they also do more gender-stereotyped housework like cooking and cleaning (Raley & Bianchi, 2006). This difference persists even when the parents' educational levels are high and they uphold egalitarian gender ideologies (Gager & Sanchez, 2004). Gender and housework association underscores the gender-role socialization processes which characterize adulthood. In adulthood, women rear children and do unpaid chores at home whereas men usually work outside in a breadwinning role.

Boys, on the other hand, were controlled less closely because they tended to spend more time outside home than girls do (Entwisle, Alexander & Olson, 1994). In the response to the question about the preference to single births, 19% of men prefer daughters but 48% prefer sons. 35% of women prefer daughters versus 30% preferring sons (Dahl & Moretti, 2004). Mothers are more likely to underestimate their daughters' but overestimate their sons' math capabilities. Families are more likely to support their sons than their daughters. Boys are more likely than girls to reap domestic resources in two-parent families (Lundberg et al., 2007). Parents with boys start reserving financial resources for their kids' college education earlier and more than those parents with girls do (Freese & Powell, 1999).

Nowadays, parents realize that females are more attached to parents than males after marriage because they are more relationally connected. Girls are more likely than boys to initiate and maintain kin contacts and to provide care to their elderly parents (Raley & Bianchi, 2006). Fathers spend almost the same amount of time with their children and face an increasing social pressure to become more egalitarian. Parental aspirations are actually higher for girls than for boys now because parents now realize that a higher level of educational achievement does not only benefit boys but also girls. Carter and Wojtkiewicz (2000) even argued that parents have higher expectations for their daughters and discuss educational matters with their daughters more often than with their sons. Academic differences between males and females in mathematics have become negligible (Nowell & Hedges, 1998; Willingham & Cole, 1997).

By and large, ‘the literature suggests that the gender of children has implications for the ways in which parents treat, spend time with, invest in, and ultimately receive care from their children later in life’ (Raley & Bianchi, 2006:417). ‘While informative, these studies do not provide an insight on whether these conditions may contribute to gender variations in the effect of parental support on student achievement.’ (Chen, 2004:77). Therefore, it is intriguing to investigate whether there are significant differences in math achievements between Hong Kong male and female students.

2.4 The Importance of Teacher’s Attributes

Teachers are another important social factor and are found to have a positive impact on student reading and math achievements after controlling for SES and prior achievements. Officials in western culture discern that higher educational effectiveness can only be achieved by improving teachers’ quality in terms of knowledge and skills. In 1992, the Office for Standards in Education (OFSTED) in the UK reemphasized the importance of teacher effectiveness and summarized that teachers and teaching qualities are required for improving learning outcomes: teachers should understand subject materials well; plan and prepare for subject teaching with

clear learning objectives and suitable pedagogies; encourage and expect the most from the students; make effective use of time, resources and methods to motivate students to learn and finally, use assignments and other monitoring practices to reinforce and facilitate students' progresses.

Besides, Ohio in the US now requires its student teachers to meet or even exceed the professional accreditation standards set by the National Council for Accreditation on Teacher Education (NCATE). New teachers need to pass the performance assessments in order to receive their professional licenses, plus to re-take the test every five years. Teacher quality is widely interpreted as the greatest influence on student learning and Section 207 of the Higher Education Act (1998) requires colleges to report information on teacher quality to U.S. Department of Education in 2001.

Teachers have a great impact on the quality of education in Hong Kong as well (Policy address by Chief Executive, 2005). Mujis and Reynolds (2001) believed that effective teachers can help students gain more in standardized achievement. An effective teacher improves student academic achievements and an ineffective teacher can cause irrecoverable damage to the entire educational success of a student (Sanders & Rivers, 1996; Wright, Horn & Sanders, 1997).

2.4.1 Teacher licensure test scores

U.S. teacher candidates are required to take the licensure test before they graduate and these test scores, which indirectly measure the content knowledge of the teacher candidates, are found to have impacts on student academic achievements (Ferguson & Brown, 2000). Higher-score teachers do not only help students from overcoming other obstacles like race, parents' income and education levels but also exert a cumulative effect on students so that those who score low in grade one will improve at higher grades. New findings corroborate that teacher effects on student academic achievements would last for the next several years (Rivkin, Hanushek & Kain, 1998; Sanders, Saxton & Horn, 1998). Greenwald, Hedges and Laine (1996) studied 24 independent coefficients measuring the relationship between teachers' scores and

student achievements and found that 21 coefficients are positive; and 12 out of these 21 positive coefficients are significant at the 0.05 confidence level.

2.4.2 Teaching experience

Teachers may have problems because of the new teaching environment and their inexperience but they can learn during their first years in work. However, the results are conflicting. Out of 109 previous studies, 33 studies show significant positive relationships between years of teaching and student academic achievement, but 7 are demonstrated to have negative relationships (Hanushek, 1986). More recent studies like Wiseman and Brown (2002) also found a positive significant correlation between years of experience and student scores (math: $r=0.095$ and science: $r=0.104$, $p<0.001$).

Besides teaching experience, field experience is not proven to have any relationship with student achievement. It only helps the professionals to learn and reduce the anxiety of new teachers.

2.4.3 Teaching attitude

Goldhaber and Brewer (2000) argued that teachers' attitude about professional development matters. Teachers with a higher degree in major teaching areas can improve student scores in math and science subjects, although not for English or history (Monk & King-Rice, 1994; Wiseman & Brown, 2002). Yet little research is available to support the argument that continuous professional training in non-subject-specific degrees is correlated with student outcomes. The author interprets the result as a phenomenon of which teachers' life-long learning attitude is conducive to students' learning only to the extent of subject-related training and subjects of a quantitative nature. Teachers' attitude to colleagues also matters. The more time staff members spend on collaborating with colleagues, the more effective the school is (Greenwald, Hedges & Laine, 1996).

2.4.4 Teaching practices

Teaching practices involve all kinds of activities: to keep track of student progress; to review pedagogical decisions; to help student learning such as questioning students during class, walking around the classroom when they are doing in-class exercises, grading homework, conducting periodic revision, examinations and review sessions.

Learning probe refers to the different ways to check whether the students understand the teaching materials. Questions to the class, quizzes or other methods are proven to improve student achievements if: (1) the questions can be answered by most of the students; (2) non-volunteers are not ignored; (3) students are asked to further comment on others' answers; (4) using information familiar to students in order to supplement the explanations (Fisher et al., 1981). The practice of monitoring student learning affects student achievements (Purkey & Smith, 1983). It comprises effective questioning/recitation skill strategies, learner's accountability and smooth transitions. Teacher performance is high if the teacher is well-prepared, empathic and consistent; has energy to help and a sense of humor; engages most of their students in the class work; and uses diversified pedagogies (Wilson, Floden & Ferrini-Mundy, 2001).

2.4.5 In-class exercises and homework

Teachers who monitor in-class exercises closely are more effective. They move around the classroom, are aware of students' progress and work with students on an individual basis if necessary. They demonstrate, supervise and motivate students when they are doing in-class exercises, leaving them to practice but interacting with other students by walking around, and giving extra time and attention to low achievers (Good & Grouws, 1979; Medley, 1977). Homework is found to be positively correlated with achievement if it is carefully designed to extend the learning process (Brophy, 1979; Good & Grouws, 1979). The homework should be related to the subject content; assigned regularly; commented on and returned quickly; within the students' abilities; and helped by the parents if designed to be so (Walberg, Paschal & Weinstein, 1985).

2.4.6 Periodic reviews and tests

Teachers who conduct periodic classroom reviews can improve students' achievements (Stiggins, 1985; Stiggins, Conklin & Bridgeford, 1986). Classroom testing is positively correlated with student achievement if it can be well-aligned with the subject matter; be administered regularly and returned quickly (Gronlund, 1982). Surprisingly, Peckham and Roe (1977) found that students have positive attitudes towards tests if the tests are held regularly and returned with comments promptly. Effective teachers use systematic procedures to trace students' progresses (Natriello, 1987). Using these monitoring methods in combination is more effective (Natriello, 1987; Rosenshine, 1983).

2.4.7 High expectations

Natriello (1987) emphasized that the performance standards must be set attainably high. Being challenged with new and complex materials, student efforts and achievements will increase. Worsham and Evertson (1980) found that establishing clear expectations and guidelines with rewards/sanctions can facilitate learning outcomes. Clarity, frequency and regularity of assessments and feedback are important. Given the strong research support on the relationship between monitoring and student achievements, it is imperative that the teacher can perform these activities effectively.

2.4.8 Commitment

According to Darling-Hammond (1997), fully prepared teachers are more loyal, highly rated and more effective with students than those who lack background training. They found that students perform better in national exams if their instructors are fully licensed in the subjects they teach. 80 percent of teachers who have undergone five-years of training would stay in the profession for three years or above in comparison to the figure of those under-qualified teachers, which amounts to less than 50 per cent (Darling-Hammond, 1997).

However, the above-mentioned literature focuses mainly on teachers' attributes which are individualistic and competitive. Bloom (1972) stated that what teachers have learnt influence what teachers do; what teachers do, in turn, influence what, and how

much, students learn. One can postulate that if teachers are assessed in individualistic and competitive ways, they normally adopt the individualistic and competitive methods of teaching rather than cooperative learning. In a highly competitive learning environment, there are few 'winners', a lot of 'losers', and a decline in self-concept (Covington, 2001). By de-emphasizing the competitive environments, educators can encourage students to pursue their own goals in comparing to criterion reference standards rather than to the performances of other students (Marsh & Craven, 2002). The author posits that teacher emotional support and cooperative learning are more effective than teachers' licensure test scores, teaching experience, teaching attitude, teaching practices, in-class exercises and homework, periodic reviews and tests, high expectations and commitment (i.e. 2.4.1- 2.4.8). Besides, teacher emotional support and cooperative learning can not only enhance students' academic achievements but also reduce teachers' and students' anxiety.

2.4.9 Cooperative learning

Cooperative learning is a relatively new trend in math learning. Leiken and Zaslavsky (1999) defined cooperative learning as the practice of which students learn in small groups with two to six members who mutually and positively depend on one another and on the group's work as a whole. They have an equal chance of communicating ideas in various ways and are held responsible for the outcomes. Students work together to accomplish shared goals. It not only helps students to understand math concepts but also helps to develop their social skills.

The relationship between cooperative learning and academic performance is well documented. Brown and Palincsar (1989) explained how students' learning motivations can be enhanced by group work. Berg (1999) found that cooperative learners achieve more than those students who are taught by conventional methods under a more supportive learning climate. Brancov (1994) and House (2003) found that cooperative learning activities for instructional situations can affect student's interest in learning math significantly. Schiefele and Csikszentmihalyi (1995) also recommended teachers to use small group activities and projects, computers and real life problems to arouse students' motivation. Johnson and Johnson (2004) believed

that teachers can make cooperative learning in math successful, but it requires a lot of teacher-student interactions and a close teacher-student relationship. Anderson, Manoogian and Reznick (1976) concurred that students are less intrinsically motivated when their teachers are detached from them.

Gender disparity in teacher support

The social reproduction theory focused on the differences in socialization process that lead to gender disparity within a school (Wong, Lam & Ho, 2002). Social reproduction theorists consider schools as agents that socialize students into the stereotyped men and women in society, which in turn perpetuates the inequality between genders. Social psychologists are interested in understanding more about how individuals form gender-differentiated self-concepts and behave in accordance to their gender-stereotyped roles.

The patterns of teacher emotional support are different and such a relationship tends to favor males (Lee, Marks & Byrd, 1994). Mael (1998) found that boys in co-educational schools are called out more often to answer questions, were allowed to talk longer, and were challenged to try difficult questions whilst girls are often less challenged and less engaged in questioning and discussion. Walkerdine (1998) reported that when girls do well teachers attributed the better academic performance to their hard work whereas when the boys do well, it is attributed to their abilities. However, the literature is full of speculative associations and the ways through which the gender of a child affects math performance are not yet well understood. Therefore, it is intriguing to investigate how true this belief is in the Hong Kong sample.

To recap, prior research found that support from parents and teachers affect adolescents' academic achievement but they have conflicting views on different levels of support from parents, teachers and their relative impacts (Deci & Ryan, 1991; Furman & Buhrmester, 1992).

Moreover, prior research also found that there are gender variations in adolescents' academic achievement (Wong, Lam & Ho, 2002). According to Hill and Lynch's

(1983) 'Gender Intensification Hypothesis', boys have been socialized to be rational, assertive and independent, whereas girls have been socialized to be emotional, compliant and interdependent. Gender differences in academic achievement may be more obvious in a non-western cultural context like Hong Kong. As it is influenced by the Confucian Wu Lun philosophy, Hong Kong is an adult-oriented and patriarchal society. It is virile in nature and males usually have higher social status than females. Traditionally, males have a higher priority to get the resources from their parents. Even teachers in Hong Kong have higher academic expectations for boys than girls (Zhang & Sternberg, 1998). Males are more advantageous in attaining higher educational and occupational status than females (Eccles et al., 1993; Hill & Lynch, 1983). Gilligan (1982) also found that women are more relationally oriented and interdependent than men. Based on this different emphasis on genders during the role socialization process throughout adolescence, it is logical to postulate that supportive relationships between parents and teachers may have more significant impact on girls' rather than boys' math achievements.

Nevertheless, British sovereignty upheld equal educational and employment opportunities for both men and women in the colonial era. All girls have had the same educational opportunities and have enjoyed a nine-year compulsory education since 1978. Women have had the same employment opportunities and enjoyed the equitable distribution of employment opportunities since the establishment of the Equal Opportunities Commission in 1996. Gradually, Hong Kong has been transformed from a traditional masculine society to a society which advocates egalitarianism.

Many researchers studied gender variations in academic achievements, ranging from difference in beliefs (Stipek & Gralinski, 1991), learning strategies (Wang & Staver, 1997) and learning preferences (Eccles, Jacobs & Harold, 1990). In this age of rapid globalization, Hong Kong's future depends on the next generation. How these children are educated will significantly affect the competitiveness of Hong Kong in the world arena. Knowing what factors contribute to adolescents' academic performances is important and their relative impact across gender cannot be overlooked (Chen, 2004). However, there is a paucity of research examining how

these gender differences in achievement are attributed to the support of parents and teachers. It is important to study whether parents and teachers exert differential support on male and female students and whether the support exerts differential influences on the academic achievements of male and female students.

Based on the previous studies, we posit that girls: (1) attract more parental and teacher support and resources; (2) benefit more in cooperative learning environment; (3) have lower math self-concept, math self-efficacy and interest in the male-dominated math subject; and (4) perform significantly worse than boys. Therefore, 2,152 female and 2,066 male students' responses to PISA 2003 questionnaires about their perceived support from parents, relationship with their teachers, their math self-concept, math self-efficacy, cooperative learning environment and interest in math were studied. Math achievement was gauged by a test. Major analyses were conducted by using Structural Equation Modeling.

2.5 Reasons for Use of Existing Data

This dissertation made use of data already existing. 'Existing data' refers to data acquired from secondary sources rather than from original data collection efforts (Hatry, 1994). They can include surveys or reports completed by outside sources, data banks, archives, or organizations' data sets publicly available, often at a reasonable or no cost. It follows that existing data was collected before the researcher's own project and, crucially, for other purposes. There is often reluctance amongst the research community to rely on existing data sources not only because there is often an assumption that research must rely on primary data but also there is a consensus that methodological training is geared to the collection and analysis of primary data.

Some of the concerns relating to the use of existing data are undoubtedly valid and any decision which relies on this data source requires a careful consideration of the latter's benefits and drawbacks. For example, existing data may not be appropriate or in an inappropriate format for the current research. Data may be outdated, or incomplete, with hidden errors which cannot be controlled by the researcher. Its

accuracy and reliability may be questionable. There is a fear of ‘dustbowl empiricism’ (Shultz, Hoffman & Palmon, 2001:35) and stagnation of theory in which theory development fails to ‘move on’ and progress because of an over-reliance on historic, and outdated data.

International comparisons can be criticized in many ways, but it would be wrong to go to the other extreme and suggest that there is no meaning in the numbers at all. Bearing in mind the limitations, it is possible to draw a board-brush picture of students’ educational performances. A detailed study can follow up the discrepancies within countries and between countries in order to validate the relationships and adopt the best practices commonly found in successful countries (Smithers, 2004). A strong case can be made for using existing data for research and there are a number of benefits derived from using it. All research limitations must be identified, understood and addressed; therefore the probability of misuse will be minimized.

In this dissertation, PISA data will be used after carefully weighing the advantages and disadvantages of this sort of data. The most compelling argument for using PISA data would be due to the scale and availability of the data collected. As it contains more than 260,000 valid cases which are more numerous than any single researcher can possibly collect, this makes newer and rigorous statistical analysis feasible. The sheer volume of this reliable data makes it possible to undertake a number of detailed statistical analyses that are quite distinct from the material and conclusions published in the original PISA papers. It is possible to identify a number of ways in which this data may be used. The data is international/ cross-cultural and allows cross-national comparisons. PISA material can also be used as pilot data and new data can be collected to supplement it if deemed necessary.

In addition, development of survey instruments was a collaborative effort of international educators. PISA receives strong government and media attention (Li, 2003) which significantly affects the respondents’ degree of participation and the quality of data. The PISA website is published by the OECD. If respondents believe OECD can protect their right of confidentiality, they are likely to provide accurate and

valid data (Topping et al., 2003). All data is gleaned under stringent procedures to ensure its reliability (PISA Technical Report, 2003). The intention of the proposal for using the PISA data is to test theories related to educational achievements. Using the PISA data source may in turn help strengthening them or induce the need for change in the next round of policy implementation (Shultz, Hoffman & Palmon, 2001).

Confidence in the quality of the data is important for such an accredited organization. If the quality of the data is in doubt, she will lose her goodwill and reputation as an objective source of trustworthy information. Of course the degree to which this data can be accepted as a trustworthy source of information is subjective and controversial. As a matter of fact, they are accepted in the form by many researchers, journals, universities and governments (Fuchs & Woessmann, 2004; Li, 2003; Nonoyama, 2005). For instance, Nash (2005) also used the PISA 2000 UK dataset to analyze the relationships between SES, wealth, number of books and reading performance and the findings were published in British Educational Research Journal.

2.6 Structural Equation Modeling (SEM)

This dissertation made use of SEM for data analyses. SEM is a second generation multivariate technique (Fornell, 1982) and has been widely applied in many disciplines like management sciences, management, marketing, organizational behavior and supply chain management recently. The idea of SEM is based on the observations that every theory comprises a set of covariances or correlations which exists in the observed data and, if the theory is valid, it explains and reproduces the patterns of covariance found in the empirical data (Kelloway, 1998). A structural equation model consists of two parts: the measurement model and the structural model, which should be tested separately (Hair et al., 1995). This is because a meaningful structural model cannot be established unless valid and reliable measurement models have been assured.

Bagozzi (1980) believed that the structural models developed concomitantly share four advantages. Firstly, all the assumptions, constructs and hypothesized

relationships are firmly based on the prior theories and concepts. Secondly, with the clear definitions of assumptions, constructs and relationships, these models can be tested against each other and the researcher's theory can be more credible. Thirdly, they provide a more holistic picture of complex theories with both direct and indirect effects. Finally, they provide a framework to validate the concepts and items together rigorously.

SEM is a well-developed data analysis method, which applies many traditional data analytical techniques to the special cases (Mueller, 1996). It serves two general purposes: evaluating the degree to which the hypothesized model fits in the observed data collected, and the estimation of factor loadings. A structural model with acceptable fit suggests that the empirical relationships among the variables are consistent with those implied by the model. The factor loadings show the magnitude and direction of the relationship between variables (Nasser, 2004). SEM allows researchers to frame increasingly precise questions about the phenomena in which they are interested (Kelloway, 1998). Put simply, the use of SEM entails a mode of thinking about theory construction, measurement of problems, and data analysis that is helpful in building and testing the theory more precisely (Harries & Schaubroeck, 1990).

However, the effective use of SEM is subject to the understanding of its limitations. First, SEM requires a precise and relatively narrow definition of research constructs. Operationalizing a construct by using too many indicators often results in a misfit in measurement models (Anderson & Gerbing, 1988). Second, SEM cannot be used in exploratory research (Hair et al., 1995; Harries & Schaubroeck, 1990; Kelloway, 1998; MacCallum, 1986; Silvia & MacCallum, 1988). It is essentially a confirmatory technique which requires a researcher to develop a few competing models for hypothesis testing. Using SEM in testing complex models will always lead to the rejection of the models regardless of their validity (Bentler & Chou, 1987) and therefore SEM is only suitable for small and medium structural models with few constructs (Yeung, Cheng & Chan, 2004).

2.7 Measurement Invariance

Assessing the applicability of conceptual frameworks in one population to other populations is contingent upon the establishment of the generalizability of theories. Verma (2002) defined comparability as data which can be legitimately (in statistical sense) aggregated, differenced and interpreted against each other on some common standards. Comparisons will be at best ambiguous and at worst erroneous if the measurement instruments are non-invariance. For example, cross-national difference in latent means (κ) might be due to the true differences between countries on the underlying latent factors. However, it could also be due to the systematic bias in the way respondents from different countries responded to certain items, the differences in factor reliability, or even nonequivalence of the configuration involved. Horn (1991: 119) stated a strong statement as ‘Without evidence of measurement invariance, the conclusions of a study must be weak’.

Measurement invariance can be applied to any situation in which data are garnered in at least two different populations, and the same set of questions is developed to manifest the latent constructs of interest. Most of the datasets consist of multiple subgroups namely countries, ethnicity or genders. Policy makers are also interested in the diverse characteristics of these subpopulations to fine tune the policies. Therefore, comparisons can be done on (1) the macro level in order to construct the whole population picture and (2) the micro level in order to contrast the subpopulations’ heterogeneity. Indubitably, measurement standardization is important to control biases of measurement in the comparison processes. This standardized yardstick is used to ensure that the same type of information is collected.

There are growing needs for data for national and international comparisons and calibration. Countries do not only need to know their places in comparison to other countries but they also want to share and learn the best practices from them. International agencies, e.g. OECD, have an interest in promoting statistical development and capability. National agencies, e.g. NCES, require similar data on different countries for their international programmes. The rise of super-national

bodies and organizations like the European Union requires a set of data at the aggregate level rather than at the individual member country's level. Information with more international essences would gain more attention and make it more easily funded and published. Recipients can get more financial and technical support on organization internationally and implementation locally (Verma, 2002).

Ironically, measurement invariance receives little attention in educational studies due to several reasons. Firstly, many researchers in education are idiosyncratic and emphasize collecting primary data for study. Inevitably, the sample size is small and cannot be tested against the invariance. Secondly, researchers are unfamiliar with testing measurement models which include latent and observed variables. Thirdly, the considerable methodological complexities involved in testing measurement invariance baffle many researchers. Therefore, this section will follow the practical and sequential procedures mentioned in Steenkamp and Baumgartner (1998) in order to assess the measurement invariance across gender before SEM. Horn and McArdle (1992:117) defined measurement invariance as 'whether or not, under different conditions of observing and studying phenomena, measurement operations yield measurements of the same attributes'

The latent constructs cannot be observed and measured directly. However, researchers can use CFA to hypothesize the relationships between observed indicators x_i and the underlying unobserved constructs ξ_j (Bollen, 1989). SEM assumes linear relationships between indicators and latent variables, and between latent constructs. Mathematically, this relationship is represented by a linear equation as:

$$x_i = \tau_i + \lambda_{ij}\xi_j + \delta_i \quad (1)$$

λ_{ij} is the slope or factor loading of the regression of x_i on ξ_j . It defines the metric of measurement and shows the rate of change of x_i in terms of the rate of change of ξ_j . The intercept τ_i , gives the expected value when ξ_j and $\delta_i = 0$. The stochastic error

term δ_i , indicates unexplained variance if any. For measurement invariance, one of the factor loadings of the construct has to be fixed to 1 as a marker (reference) item. The same item's factor loading should be used as marker in each country. 'The overall fit of the model is based on the discrepancy between the observed variance-covariance matrices S^g and the implied variance-covariance matrices $\hat{\Sigma}^g$, and the discrepancy between the observed vectors of means m^g and the implied vectors of means $\hat{\mu}^g$ ' (see Sorbom (1974) for details).

2.7.1 *Forms of measurement invariance*

The first stage consists of three concomitant forms which are configural invariance, metric invariance and scalar invariance. The second stage consists of another three forms of analyses, which are factor covariance invariance, factor variance invariance and error variance invariance. These six forms of invariance will be discussed here and then will be applied to PISA 2003 HK dataset for the purpose of measuring invariance across gender in chapter 4.

At the first stage, configural invariance checks whether the measurement instrument exhibits the same configuration of fix and free parameters across different groups (Horn & McArdle, 1992). The intercorrelations between items and corresponding factor should be large enough to demonstrate convergent validity while the intercorrelations between factors should be small enough to demonstrate discriminant validity. At this stage, only the fixed parameters are constrained across groups but the magnitudes of free parameters are not constrained until the next stage - metric invariance.

Metric invariance is a stronger test which is used subsequently to ensure respondents in different groups respond in the same way. Mathematically, this is done by constraining the factor loadings matrices, (Λ) to be the same across groups as follows:

$$\Lambda^1 = \Lambda^2 = \dots = \Lambda^G \quad (2)$$

If metric invariance is achieved, the different scores on factor loadings across groups can be compared for insights, and any difference will indicate cross-group differences but will not be due to different perceptions on the items. Scalar invariance implies that cross-group differences in the means of the observed items are due to differences in the means of the underlying constructs. Even if the factor loadings matrices are the same across groups, the means of the constructs can be biased upwards or downwards by bias added on and caused by taus (Meredith, 1993). Thus, this is done mathematically by constraining the tau matrices(τ) to be the same across groups as follows:

$$\tau^1 = \tau^2 = \dots = \tau^G \quad (3)$$

The second stage is to compare factor covariance (ϕ_{ij}) invariance, factor variance invariance (ϕ_{ii}) and error variance invariance (θ_{δ}^i). The covariances between factors across groups have to be constrained in the same magnitude. Similarly, variances of each factor have to be constrained in the same magnitude as well. If both the factor variances and covariances are invariant, that implies the groups are homogenous. Error variance invariance is tested by specifying mathematically as below:

$$\Theta^1 = \Theta^2 = \dots = \Theta^G \quad (4)$$

If all three invariances at the second stage are achieved, the items are then proved equally reliable across groups and are able to achieve full invariance.

2.7.2 Full versus partial invariance

Full measurement invariance is an ideal situation to strive for but it does not always hold true. To compromise between pursuing unrealistic full invariance and ignoring it, partial invariance was proposed. Byrne, Shavelson and Muthen (1989) proposed to

get at least one more item metrically invariant in addition to the marker item for each underlying construct. This can be done by referring to the Modification Indices (MI) and Expected Parameter Changes (EPC). MI in LISREL or the Lagrange multiplier tests in EQS software are actually one degree-of-freedom noncentrality parameters. The EPC is an estimate of the value that a fixed parameter will take if that parameter is freed (Kaplan, 1995).

Invariance constraints can only be relaxed if MIs and EPCs are substantial enough to avoid capitalization on chance, so that they only fit in this sample but not the others. Large MI associated with large EPC values imply that those corresponding parameters have a higher chance of being false. In other words, researchers may manipulate the data according to the MIs and change the originally hypothesized models resulting in poor consistency with different data. Thus, model parsimony and cross-validity of the model should be upheld to discourage data-driven model modifications (MacCallum et al., 1992).

After achieving partial metric invariance, the researcher can test partial scalar invariance. Some items with invariant factor loadings may still have different intercepts. Cross group noninvariant intercepts can also be located by MIs and EPCs and relaxed if substantial. Finally, factor covariances, factor variances and error variances can be checked. Error variances of those items with noninvariant factor loadings are allowed to be set free. Once full measurement invariance is achieved, the data can be treated homogeneously and aggregately for analysis.

2.7.3 *Methods*

The standard criterion for testing whether the invariance can be achieved is the chi-square difference test (Jöreskog, 1971) which has a well-known limitation - sensitivity to the sample size. A large sample size is needed to detect small effective size changes but this inevitably inflates chi-square and subsequently increases type I errors (Anderson & Gerbing, 1988; Kelloway, 1998). Thus, supplement fit indices are recommended, namely root mean square error of approximation (RMSEA); the

Comparative Fit Index (CFI); and the Non-Normed Fit Index (NNFI). Smaller RMSEA with larger CFI and NNFI indicate a better model fit. They are the most effective criteria because they consider both the model to be having the goodness of fit and parsimony by penalizing extra unnecessary free parameters (Williams and Holahan, 1994).

2.8 Development of Hypotheses

This research aims at providing additional insights by investigating specifically the relationships in a larger sample size from a single city. The author first develops a model based on the Vallerand and Losier (1999) framework and proposes one alternative nested model from different perspectives. Cooperative learning, math self-concept, math self-efficacy, intrinsic motivation or interest in math, and parental education level are discussed briefly below.

Johnson and Johnson (2004) believed that teachers can make cooperative learning in math successful, by requiring a lot of teacher-student interactions and a close teacher-student relationship. Students whose teachers are autonomy supportive rather than cold and uncaring have greater interest in math, curiosity and desire for challenge (Flink, Boggiano & Barrett, 1990; Ryan & Grolnick, 1986). Schiefele and Csikszentmihalyi (1995) recommended teachers to use small group activities and projects, computers and real life problems to arouse students' motivation. Berg (1999) found that cooperative learning under a more supportive learning climate achieves more than those who are taught by conventional methods. According to House (2003), cooperative learning activities for instructional situations can affect student's interest in learning math significantly. Using the foregoing as a basis, we formulate teacher emotional support is positively related to cooperative learning and math self-concept (e.g. Cornelius-White, 2007; Patrick, Ryan & Kaplan, 2007), which, in turn, is associated with math self-efficacy and interest in math (e.g. Blatchford et al., 2006; Pastorelli et al., 2001).

H1: Teacher emotional support is positively related to math self-concept.

H2: Teacher emotional support is positively related to cooperative learning.

H3: Cooperative learning is positively related to math self-concept.

H4: Cooperative learning is positively related to math self-efficacy.

H5: Cooperative learning is positively related to interest in math.

Self-concept has been defined as the cognitive appraisal one makes of the expectations, descriptions, and prescriptions that one holds about one's self (Pajares & Schunk, 2002). It has direct and indirect effects on academic performances as reflected by students' interest in math (Williams, Freeman & Deci, 1998). Sommer and Baumeister (2002) found that low academic self-concepts coupled with failure cause poorer performance and slow recovery. High self-concept can enhance motivations and mitigate disappointments or negative feedback after poor performances. Previous research has provided strong evidence for the causal relationships. For example, expectancy-value theory (Vroom, 1964) posited academic self-concept has causal effects on both academic interest and achievement, and academic interest to have an effect on academic achievement. Thus, we posit that there are positive relationships (1) between math self-concept and interest in math; and (2) between math self-concept and math scores (e.g. Deci & Ryan, 1991; Marsh et al., 2005) as follows:

H7: Math self-concept is positively related to interest in math.

H8: Math self-concept is positively related to math scores.

Self-efficacy has been defined as 'people's judgments of their capabilities in organizing and executing courses of actions required attaining the designated types of performances' (Bandura, 1986:391). Pajares and Graham (1999) believed that math self-efficacy is more important than self-concept in affecting math achievement. A strong sense of self-efficacy causes an individual to attempt more challenging tasks (Bandura & Schunk, 1981), put more effort in consummating the task (Salomon, 1984; Schunk, 1983), persist longer in adverse situations (Bandura & Schunk, 1981), self-regulate better than others (Zimmerman & MartinezPons, 1990) and use more effective strategies (Pintrich & De Groot, 1990). What people believe about their capabilities can make a difference. A low sense of self-efficacy leads to avoidance of potential learning opportunities and, in turn creates a self-fulfilling prophecy of

incompetence. Thus, math self-efficacy positively affects math scores. According to Pajares and Schunk (2001:244) there is positive relationship between math self-concept and math self-efficacy since they believed that ‘self-efficacy and self-concept beliefs are each related with and influence academic achievement’.

H6: Math self-concept is positively related to math self-efficacy.

H11: Math self-efficacy is positively related to math scores.

Motivation is found to be one of the mediators mediating the effects between self-concept, self-efficacy and achievement. Those with higher interest in math have more interest, excitement and confidence, resulting in not only better performance, persistence and creativity, but also higher vitality, self-esteem and well-being (Nix et al., 1999; Ryan, Deci & Grolnick, 1995). Intrinsic motivation or interest in math is defined as ‘the inherent tendency to seek out novelty and challenges, to extend and exercise one’s capabilities, to explore and to learn’ (Ryan & Deci, 2000:70). Vallerand (2001) also proposed motivation affects performances. In this study, we posit that the following:

H9: Interest in math is positively related to math self-efficacy.

H10: Interest in math is positively related to math scores.

Parental education level has a significant relationship with their children’s educational achievements (Baker et al., 2002; CME, 2004a; Ferguson, 1991; Heyneman & Loxley, 1983). Therefore, parental education should be selected as a proxy for SES and it is consistent with prior studies (e.g. Baker et al., 2002). This discussion forms the basis of the following hypotheses. Parental education level, which is available in the dataset, positively affects home educational resources which subsequently, positively affect math scores.

H12: Parental education level is positively related to home educational resources.

H13: Home educational resources are positively related to math scores.

Therefore, structural model A was built as in Fig. 2. Cudeck and Browne (1983) suggested comparing different plausible models that are nested in one another and can be justified theoretically. Nested models have the same constructs of a general model,

but with fewer estimated relations. Bentler and Chou (1987) pointed out that in an ideal situation a researcher should build and compare a series of sub-models that shed light on the key features of a general model. If the models are fit, then the simplest model is chosen with the smallest Parsimony Goodness of Fit index (PGFI) because it represents the theory better. Therefore, another structural model, model B, was built and compared in Fig. 3. Finally, a model would be selected which not only best fits the observed data, but also with the convergent and discriminant validity which will be discussed in the next chapter.

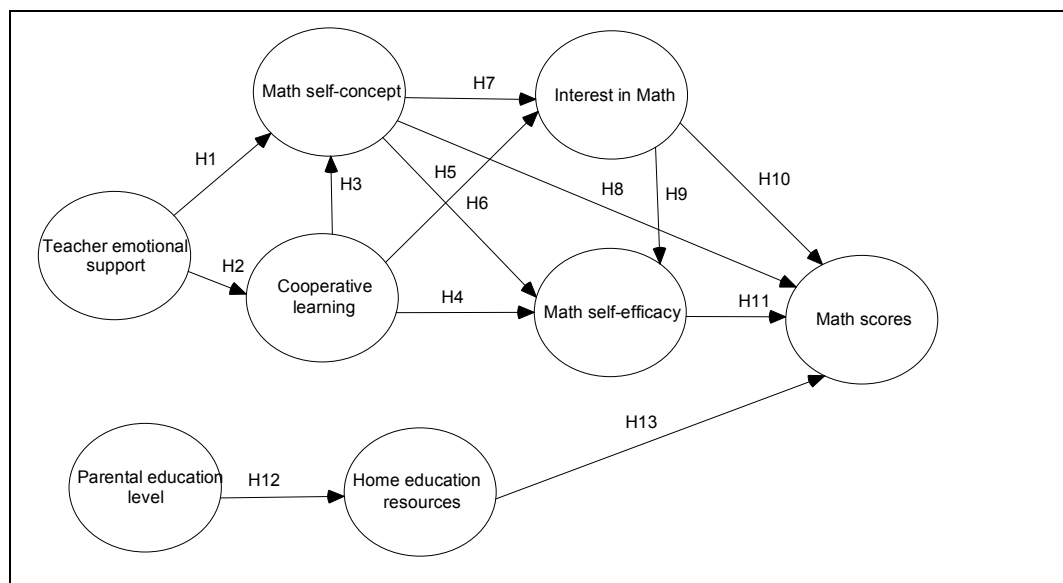


Figure 2. The proposed model A

Structural Model B:

Developed upon the previous models, the author further postulates that parents have specific effects on their children's academic performance, not simply transferring their home educational resources in the form of cultural and economic capital and in the form of social capital (Coleman, 1988). For instance, Weiss and Krappmann (1993) believed that families develop students' personality and help them to get in

contact with society. Kohn (1989) found that parents' occupations affect parents' interactions and communication styles with their children. Parents' communication styles, educational aspirations and social network are apparently important. Dika and Singh (2002:43) found nine out of fourteen studies in their meta-analysis had shown a positive link between social capital and psychological factors and stated 'the family is the first source of efficacy information for children... contribute to the build-up of a sense of personal efficacy' (Pastorelli et al., 2001:88) and that 'further research is needed to understand the interplay of factors and the access to and mobilization of social capital'. In line with their recommendation, the following hypothesis was set.

H14: Parental education level is positively related to math self-efficacy.

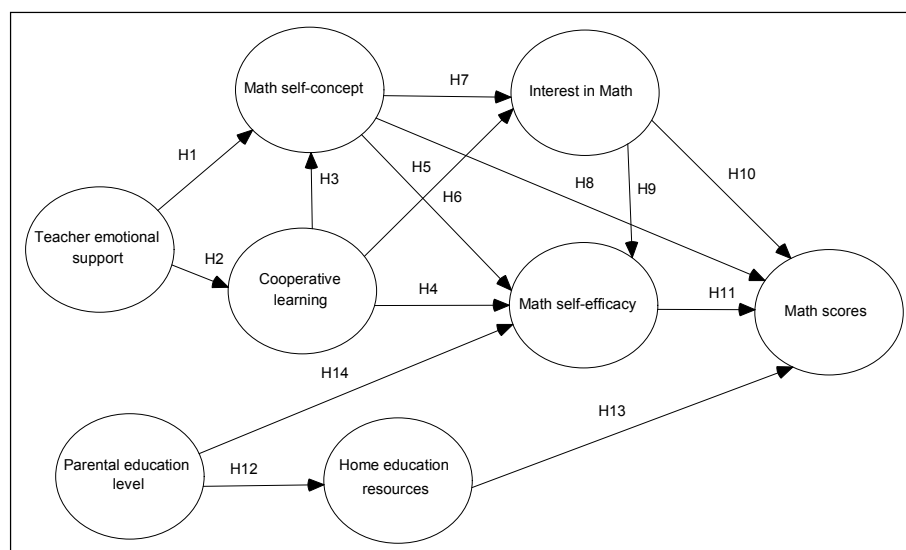


Figure 3. The proposed model B

Model A is nested within model B in this study because it satisfied two conditions. First, model A has one fewer parameter to estimate than model B and therefore has a larger number of degrees of freedom than does model B. Second, the parameter vector for model B does not include new parameter estimates that do not appear in the parameter vector for model A (Widaman & Thompson, 2003).

2.9 Instrument Description

This investigation is a theory-testing study to identify the significant mediators between social factors and academic performance. The research process for this study is outlined in Fig. 4.

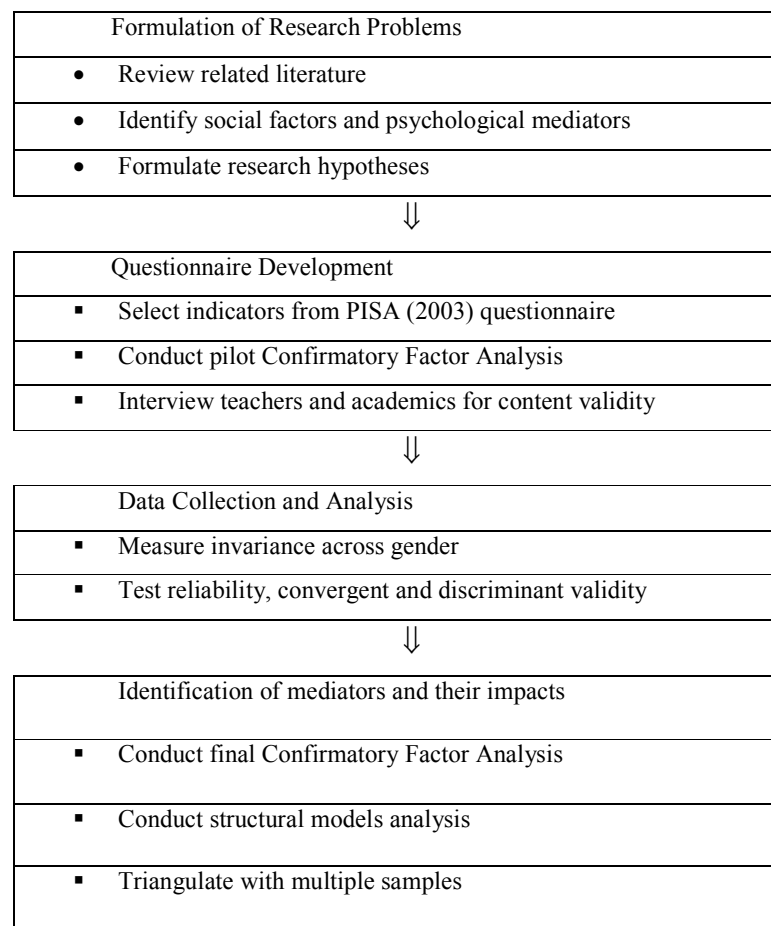


Figure 4. The research process

2.9.1 Teacher emotional support

This unidimensional scale assesses students' perceptions of the support and care provided by their teachers. This 3-item scale is similar to those used by Patrick, Ryan and Kaplan (2007) and included items such as 'Most teachers are interested in

students' well-being', 'Most of my teachers really listen to what I have to say', and 'Most of my teachers treat me fairly'. They are originally coded as ST26Q02, ST26Q03 and ST26Q05 respectively in PISA 2003 questionnaire. Subjects responded by using a 4-point Likert scale to indicate their levels of agreement with each statement, for example, from 1 (strongly agree) to 4 (strongly disagree). This scale was re-coded and applied to all variables except math scores. This measure was found to be reliable in the past (alphas=0.84; Patrick, Ryan & Kaplan, 2007:97) and in the present study, yielded a Cronbach alpha of 0.72.

2.9.2 *Cooperative learning*

This unidimensional scale indicates the essences of cooperative learning in math courses. This 3-item scale, similar to Shumway et al., (2001), included items such as 'In Mathematics I enjoy working with other students in groups', 'When we work on a project in Mathematics, I think that it is a good idea to combine the ideas of all students in a group', and 'I do my best in Mathematics when I work with other students.' They are originally coded as ST37Q02, ST37Q04 and ST37Q06 in the PISA 2003 questionnaire. This measure was found to be reliable in the past (alphas=0.83; Shumway et al., 2001:11) and in the present study, yielded a Cronbach alpha of 0.70.

2.9.3 *Math self-concept*

This unidimensional scale assesses students' beliefs of self-worth associated with their perceived math competence. Three items, which are similar to Academic Self-Description Questionnaire II (Marsh, 1992), included items such as 'I get good marks in Mathematics', 'I learn math quickly', and 'In my math class, I understand even the most difficult work'. They are originally coded as ST32Q04, ST32Q06 and ST32Q09 in the PISA 2003 questionnaire. This measure was found to be reliable in the past for English self-concept and Maths self-concept (alphas=0.88 and 0.92 respectively; Marsh, 1992) and in the present study, yielded a Cronbach alpha of 0.82.

2.9.4 *Math self-efficacy*

This unidimensional scale assesses students' judgments of their statistics capabilities. The self-efficacy belief is manifested by asking the 'can' or 'how confident' questions (e.g. 'How confident are you that you can solve $726-83 = \underline{\hspace{1cm}}$ ' (Pajares & Schunk, 2002; Phan & Walker, 2000). Therefore, three items as mentioned below, which are similar to college self-efficacy inventory (Solberg et al., 1998) and academic self-efficacy constructs (Pastorelli et al., 2001; Patrick, Ryan & Kaplan, 2007) are used. They are originally coded as ST31Q01, ST31Q03 and ST31Q06, in the PISA 2003 questionnaire. They have been proved with reliability ($\alpha=0.88$; Solberg et al., 1998) and, in the present study, yielded a Cronbach's alpha of 0.77.

How confident do you feel about having to do the following Mathematics tasks?

- Using a train timetable to work out how long it would take to get from one place to another
- Calculating how many square meters of tiles you need to cover a floor
- Finding the actual distance between two places on a map with a 1:10,000 scale

2.9.5 *Interest in math/ intrinsic motivation*

This unidimensional scale assesses students' interest in math courses. This 3-item scale, similar to Marsh et al. (2005) and Pintrich and DeGroot (1990), included items such as 'I enjoy reading about mathematics', 'I do mathematics because I enjoy it', and 'I am interested in the things I learn in mathematics'. They are originally coded as ST30Q01, ST30Q04 and ST30Q06 respectively in the PISA 2003 questionnaire. This measure was found to be reliable in the past ($\alpha= 0.75$; Pintrich and DeGroot, 1990) and, in the present study, yielded a Cronbach's alpha of 0.88.

2.9.6 *Parental Education Level*

Ferguson (1991) found that 25% of the total variance of the students' math performances was due to parental education levels. Thus, two normalized variables, namely mother's educational level (MISCED) and father's educational level (FISCED) were used to indicate SES in this study. Therefore, the specific four items: ST11 to ST14, in the PISA 2003 database are used to indicate parental education. Subjects responded by choosing their parents' education level, for example, a (ISCED level 3A), b (ISCED level 3B, 3C), c (ISCED level 2), d (ISCED level 1) and e (None of the above) as follows:

- Which of the following did your mother complete at school?
- Which of the following did your father complete at school?

Subjects responded by choosing their parents' qualifications, for example, a (ISCED level 5A, 6), b (ISCED level 5B), c (ISCED level 4) as follows:

- Does your mother have any of the following qualifications?
- Does your father have any of the following qualifications?

Even though this construct has only two indicators, father's and mother's educational levels, it is still acceptable according to Harries and Schaubroeck (1990) suggestion that each construct in a measurement model requires only two or three indicators. In the present study, the inter-item correlation is 0.75.

2.9.7 *Home Education Resources*

Filmer and Pritchett (1999) used a list of possessions such as television, refrigerator, motorcycle, type of toilet, and source of drinking water to measure domestic resources in 35 countries. Similarly, thirteen dichotomous items were asked to measure the availability of cultural and economic capital of Hong Kong students (Bourdieu, 1986). The availability of a dishwasher is used here as a proxy to family wealth since only rich families can afford to buy this type of luxury goods in Hong Kong. The final score was normalized to form a composite variable HEDRES, which, in turn, is used for the single indicator variable, Resources, to represent home educational resources.

Which of the following do you have in your home?

1 (Yes) 0 (No)

- A desk to study at
- A room of your own
- A quiet place to study
- A computer you can use for school work
- Educational software
- A link to Internet
- Your own calculator
- Classic literature
- Books of poetry
- Works of art
- Books to help with your school work
- A dictionary
- A dishwasher

‘Resources’ has one indicator, HEDRES, in this study. It is treated just like the other latent variables but the error term for the single indicator variable must be constrained low to ascertain the indicator is measured with high reliability (Garson, 2007). Therefore, the transposition from HEDRES to Resources is constrained by an arbitrary high factor loading, 0.95 in this study, and an error variance as low as 0.05. Failure to do so and letting the model estimate these parameters without constraints will cause the model to be underidentified and prevent it from getting a solution.

2.9.8 Math scores

Math score, was selected because first, it underlies every facet of science and technology, from computer games, cellular phones and the Internet to medical diagnostic tests, the design of new prescription drugs and minimally invasive surgeries (Hubbard, 1999). Secondly, it is one of the core subjects being focused on in the Hong Kong Certificate of Education Examination. Lastly, PISA data in 2003

focused on mathematics literacy and measured 15-year-old students' mathematics performances which provide the researchers with a large amount of information about math to analyze.

PISA (OECD, 2003) draws its math contents from four content areas. First, space and shape are tested in terms of spatial and geometric phenomena and relationships. Second, change and relationships are tested in terms of mathematical manifestations of changes, as well as functional relationships and dependency among variables. Third, quantity involving numeric phenomena as well as quantitative relationships and patterns; and fourth, uncertainty involving probabilistic and statistical phenomena and relationships. The score mean is normalized to 500 and the standard deviation is 100 marks. In the present study, the Cronbach's alpha is 0.98 among items.

Reliance on only one method of assessment (written examination) to evaluate students' academic performance is criticized by many researchers (Gal & Garfield, 1997; Schau & Mattern, 1997) for exerting extra stress on students and not being holistic in terms of assessments. Nevertheless it is less subjective and time-consuming to collect. Hidi (2006) stated that one of the most important unsolved problems in education is how to enhance students' academic scores. Therefore, this study focused on math scores rather than other performance measurement criteria like pro-social behaviour. Besides, a voluminous literature arguing about the factors most affecting students' academic achievements was already posed by western scholars. It is significant to test whether these factors can be applied to the non-western context like Hong Kong.

Chapter 3 RESEARCH METHODOLOGIES

3.1 Research Paradigms

Research is defined as “a focused and systematic enquiry that goes beyond generally available knowledge to acquire specialized and detailed information for analyzing the interested problem” (Johnson 1994: 3) or “a systematic and methodical process of inquiry and investigation that increases knowledge and/or solves a particular problem” (Sekaran, 1992: 4). A research paradigm is defined as “the entire constellation of beliefs, values, techniques, and so on, shared by the members of a community” (Kuhn, 1970: 175) or “frameworks that function as maps or guides... and define acceptable theories or explanations, methods and techniques” (Usher, 1996: 15). Research paradigms are ways of looking at the world. The paradigm provides a conceptual framework for the researcher to see and make sense of the world. It determines what kinds of questions may be asked.

To understand paradigms helps the researcher shape his/her world views. This is reinforced by the community of practitioners who work within the paradigm and also helps to identify the researcher’s role. The paradigm shapes and determines the research process. The philosophical belief, which underlies the paradigm, will be reflected in the research question, the research design, the data collection, the method of analysis, and the presentation format.

Paradigms are differentiated by different sets of basic beliefs derived from philosophical assumptions about the nature of reality. They differ at the ontological level, ontology being the study of being. Paradigms also differ at the epistemological level, with different views about what knowledge is, what makes it valid, and how it may be acquired. Methodology and research methods will also be determined by the paradigm, as well as the theory of value (axiology) that is utilized (Fitzgerald & Howcroft, 1998). Two major research paradigms are positivism and interpretivism.

The Positivist Paradigm

Positivism presents a “realist” answer. To the positivist, or realist, the social world is tangible, hard, and made up of one single, objective reality. This reality exists whether it is observed by human beings or not. In this view, the social world is independent and external to the individual. Individuals are the product of the social world, rather than the constructors. Burrell and Morgan (1979) describe positivist ontology. The reality that a positivist holds to exist is a true, objective reality that can be studied.

In positivist research, theory is established in advance; the hypotheses are set out beforehand, and then tested by deduction against the data collected. Burrell and Morgan (1979) show that in the positivist approach, knowledge is gained about the social world by proving or disproving, which leads to the discovery of laws that apply to all. Axiology asks what the role of value is. Positivism takes the approach that rigor is necessary. Rigor emphasizes the value-free and unbiased approach of the researcher. Deductive reasoning with emphasis on internal validity is required for rigor.

Methodology asks what the process of the research will be. Positivism requires a quantitative/explanatory/deductive/nomothetic process be used. Quantitative methods use statistical techniques to identify trends or relationships. Samples must be larger and more representative than for qualitative studies. Results are generalized to the population with a known degree of confidence. Quantitative research is context-free, well-structured, with a static (unchangeable) design. In-breadth/ surface data are collected to test the hypotheses or to verify theory. Theories can be tested using rigorous procedures. Accuracy and validity are achieved through internal reliability and strict specifications. At the rhetorical level, positivism reports its findings with formal language, an impersonal voice, and accepted quantitative words.

The Interpretivist Paradigm

To answer the ontological question about the nature of reality, the interpretivist paradigm takes a “relativist” approach. To the relativist/ nominalist the social world is intangible, soft, and made up of multiple subjective realities. Individuals create their social world. According to Burrell and Morgan (1979) interpretivist ontology has

multiple subjective realities. Truths are created by individual cognition and the social transmission of ideas. Rather than being found, or discovered, truth and reality are socially constructed.

The interpretivists believe there are no universal truth, that social science is subjective, and the researcher and researched are interdependent. Neutrality or objectivity is impossible. The social world can only be understood and interpreted from the researcher's own frame of reference. There is no barrier between researcher and researched. Their interactions generate output, and the researcher's values and beliefs mediate the findings. The interpretivist believes in an emic approach, the insider view. Furthermore, unlike positivist research which constructs a hypothesis beforehand, in interpretivist research the empirical data is collected first, and hypotheses are built by induction afterwards.

For the question of methodology, interpretivists are quite a bit more flexible than positivists and use qualitative/exploratory/ inductive/ideographic processes. Qualitative methods seek to determine what things exist, rather than how many there are. Qualitative research is context-bound with thick descriptions, less structured but more responsive because it is dynamically designed to be changed or modified as the study progresses and develops. In-depth/ substance data are collected to generate patterns or theories for explaining and/or understanding phenomenon being studied. Theories can be built, but new evidence may cause the conclusions to be amended. Accuracy and validity are achieved through verification or triangulation. When interpretivists report their findings, they use informal language, a personal voice, and accepted qualitative words.

Although Thrupp (1999:43) argued that 'quantitative research can find correlations but not investigate the causal processes that underlie them, while qualitative research can investigate those processes in more depth but not demonstrate causality either', this dissertation adopts the positivist paradigm rather than the interpretivist paradigm because the general process in this study is to test a theory (Cohen & Manion, 1994). Positivist paradigm is usually associated with the quantitative methodology. The

underlying assumption is that there is one single objective reality and it can be studied by rejecting or not rejecting the hypotheses through the deductive approach. Quantitative methods use statistical techniques to identify trends or relationships, large and representative samples. Results of the population can be generalized to a known degree of confidence which the qualitative methods cannot achieve. Moreover, questionnaire items in the PISA 2003 survey are similar to those of another reliable large scale study – Trends in International Mathematics and Science Study (TIMSS, 2003). The survey is a quantitative (non-experimental) method of collecting information on the target population at a certain point of time.

Participants and procedures

Models will be tested by Structural Equation Modeling using Hong Kong's survey data extracted from the Program of International Students Assessment (PISA) 2003 as organized by OECD, on the math performances of 4,478 students at the age of fifteen. PISA is an international assessment (began in 2000) that focuses on 15-year-olds' capabilities in reading, mathematics, problem-solving and scientific literacy. PISA is conducted once every three years and PISA 2000, PISA 2003 and PISA 2006 focused on reading, math and scientific literacy respectively. This study has used the data from PISA 2003. Over 260,000 students from more than 9,000 schools in 41 countries/cities took part in a two-hour math test at their schools. PISA defines mathematical literacy as 'an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen' (OECD, 2004:37).

3.2 Sampling Design

A sample or a census can be used to gather information about the characteristics of a population. A census covers all elements of a population but a sample is only a sub-group of the population selected for obtaining the required information. If (i) the budget is tight, (ii) the time available is short, (iii) population is large, (iv) the variance within the characteristics is small, (v) cost of sampling error is low, (vi) cost

of non-sampling error is high, and (vii) nature of measurement is destructive, then a sample instead of a census should be used. Sample design starts with identifying a target population for study. A sample frame is a representation of elements of the population (Malhotra, 2004). It is important to minimize any discrepancy between the sample frame and the population. A sampling technique will be selected subsequently. PISA adopted a two-stage stratified sampling design to deliberate a representative sample. Stratified sampling segregates the population into strata and selects the elements randomly to ensure that all the strata are represented in the sample (Malhotra, 2004).

Two-stage sampling with stratification is one of the most prevalent methods of survey sampling (Bayless, 1985). Other international educational studies have also used this sampling method (Puhakka, 1999). Firstly a random sample of schools from each stratum is selected with its probability proportional to the sample size. Schools are classified into three strata: (1) Government, (2) Aided and (3) Independent/ Private with three different student academic intake, namely high ability, medium ability and low ability.

Table 1. Participating schools for each sampling stratum in HKPISA 2003

| Explicit Stratum | Implicit Stratum | Total Number of schools | Number of schools selected |
|------------------|------------------|-------------------------|----------------------------|
| | High | 17 | 8 |
| Government | Medium | 9 | 3 |
| | Low | 10 | 4 |
| | High | 127 | 50 |
| Aided | Medium | 124 | 41 |
| | Low | 107 | 33 |
| Independent/ | Local/ DSS | 29 | 5 |
| Private | International | 20 | 1 |
| Total | | 443 | 145 |

Note: 1. There is no intake classification for Independent/ Private.

2. DSS refers to schools funded under the Direct Subsidy Scheme.

The next step is to determine the sample size. This can be done by the quantitative formula (see Malhotra, 2004) or 'sample size table' (Lipsey, 1990:137). Instead of using the quantitative formula or sample size table to determine the sample size, Kline (1998) recommended the usage of as many as 20 times the cases as parameters in order to ensure that the sample size is large enough for assessing significance. The sample size is 4478 cases which is more than 100 times the thirty-nine parameters in this study. According to the Kline's approach, the sample size of this study is large enough.

3.3 Questionnaire Design

The survey method of quantitative data collection has its own strengths and weaknesses. Surveys are relatively inexpensive (especially self-administered surveys) and they can describe the characteristics of large populations. No other observation method provides this general capability. The researcher should choose the appropriate types of questions. Closed-ended questions allow pre-set dichotomous answers, multiple-choice questions, or ranking scales. Open-ended questions encourage replies in the form of sentences, paragraphs, or pages.

Questionnaires are measuring instruments with formalized sets of questions given to respondents to elicit specific information. Respondents are asked various questions regarding their behaviors, intentions, attitudes, perceptions, demographic characteristics etc. either verbally, in writing or via computers. Data collection processes are well-structured and questions are asked in a direct, pre-arranged order. A good questionnaire can yield the required information; motivate respondents to answer, and reduce response errors. Research purpose(s) should be known by respondents. In typical questionnaires, most questions are closed with fixed alternatives. Respondents therefore must choose from a predetermined set of responses. There are several advantages of this mode: it is simple and fast; data obtained is reliable; and, coding, analysis and data interpretation are relatively simple. In reality, not only can an interviewer or organization collect customer feedbacks via

computers in real time; they can also use the data to make executive decisions (Malhotra, 2004).

Questionnaires are ‘fiendishly difficult to design’ (Oppenheim 1993:1). Questionnaire items can be measured by the Likert scale, ranking order of perception on a three, five, or seven-point range (e.g. strongly agree; somewhat agree; no opinion; somewhat disagree, strongly disagree). Intervals between each item need not be of the same magnitude (Cohen & Manion, 1994). Interval data can be compared against each other using terminologies like ‘higher or lower than’; ‘more or less than’, etc. Rose and Sullivan (1996) defined measurements in respect to some variables, where one case differs from another. Operationalization means linking concepts to observable indicators. In designing questions, ambiguity, imprecision and assumption should be avoided.

Converse and Presser (1986) suggested borrowing with acknowledgement, by the usage of relevant prior research questionnaires which had already been tested and used effectively, so that time and effort can be saved and thus, PISA (2003) questionnaire was screened and adapted for this study. ‘Measurement’ means assigning numbers or other symbols to object’s characteristics according to respondents’ perceptions, attitudes, preferences etc. Numbers permit statistical analysis. Scaling requires placing responses on a continuum, e.g. Likert scale 1 to 4.

Four primary scales of measurement are: nominal; ordinal; interval, and ratio. Nominal scales identify and classify objects only. Inferential statistics, e.g. Chi-square and Binominal test, are the examples. Ordinal scale is used when numbers indicate relative positions of objects, but not the magnitude of differences between them. Inferential statistics like the Spearman rank-order correlation coefficient (ρ) and Analysis of Variance (ANOVA) can be used for such analysis. Interval scale analysis indicates equal distances in measured characteristics, but the zero point is arbitrary. Inferential statistics like Pearson Product Moment Correlation Coefficient (r), independent samples t-test, regression analysis and Analysis of Variance (ANOVA), can be used for analysis. A ratio scale has a fixed zero point and is the highest scale

which allows the researchers to classify objects; rank orders of objects, and compare intervals and ratios of scale values, e.g. height; weight; age; money.

Two available scaling techniques are: comparative and non-comparative scales. Comparative scales or non-metric scales are used in this study to directly compare between two objects, and generate ordinal data. This requires fewer assumptions, and is easily understood. Itemized ratings such as Likert scales set the degree of agreement from 1 (strongly disagree) to 5 (strongly agree) scale; '3' is neutral. These types of questions are good on an item-by-item basis (profile analysis), or one may use the total scores to reflect favorable/unfavorable perceptions.

However, using an even number of items carries the advantage of avoiding the neutral central point and was adopted by PISA 2003. The indicators using an even number of items are presented in Table 2. The values given to the options of each item were re-coded in order to show positive relationships between independent and dependent variables (e.g. the higher the math self-efficacy, the higher the math scores are).

3.4 Method of Data Collection

PISA is coordinated by the OECD which is an intergovernmental organization of industrialized countries. Forty-one countries or cities participated in PISA in 2003. School assessments were done during regular school hours in April and May 2003. This assessment was a paper-and-pencil test which lasted for a total of two hours. In total, 85 questions were asked and each student had their abilities assessed in terms of five estimates, which were called the plausible values (PVMATH1 - PVMATH5). Tapping from these plausible values, estimates of the proficiency of each student and the Item Response Theory (IRT) scales for reporting student achievements were generated, which were subsequently scaled to a mean of 500 and a standard deviation of 100 like most international studies do. Besides, a 20-minute background questionnaire was administered and it gathered information about the students and their home environments. In addition, a 10-minute questionnaire on information

technology and communications was also given to the students. For school principals, they were asked to complete a 20-minute questionnaire about their schools.

3.5 Data Analysis

In this dissertation, the author uses SEM because it accounts for measurement errors, indirect effects and different items functioning across groups. A regression is applied to each variable in the model as a dependent on causes that the model has identified. In SEM, the raw data is not used; rather, the correlation or covariance matrix is used. The factor loadings predicted by the model (Σ) specified by the researcher will be compared with the observed correlation or covariance matrix (S) collected by questionnaires in order to calculate how well-fitted the predicted model is as related to the observed results. The best-fitted one will be selected.

3.5.1 *Methods for analysis*

Structural Equation Modeling (SEM) is usually depicted in a circle-and-arrow diagram. A circle represents a latent variable which cannot be measured. A rectangle represents an observed variable and single-lined arrows represent association or causation. SEM analyzes models that include latent variables manifested by multiple observed indicators. It is based on multiple regression analysis but it performs better than the multiple regression analysis since it estimates not just the direct effects, but also the indirect effects and total effects. This method is appropriate if theoretical, empirical, and commonsense knowledge of a problem provides good reasons for proposing the related links between the latent variables. It suits the purpose of this dissertation which aims to test relationships between variables and see whether they fit in the hypothesized models developed according to prior theories and thus, it is used to test the fitness of the four structural models. SEM using covariance analysis is also important in economics and behavioral sciences (Bollen, 1989).

It is imperative to include all relevant variables with correct specifications, otherwise direct and indirect effects on dependent variables will be miscalculated. A structural

model represents the hypothetical relationships between exogenous and endogenous variables. An exogenous variable is not dependent on other variables whereas an endogenous variable is dependent on other variables. Endogenous variables include mediator variables and dependent variables. The dependent variables only bear incoming arrows pointing towards them whereas the mediator variables bear both incoming and outgoing arrows. Single arrows are used to indicate the relationships between independent, mediators and dependent variables or to connect the error terms to their respective endogenous variables. Double arrows indicate the correlations between pairs of exogenous variables which are independent variables without explicit causes, i.e. having no arrows pointing towards them, other than the measurement error term.

Path coefficient or path weight or factor loading is a standardized regression coefficient (beta) showing the direct effect of an independent variable on a dependent variable in the path model. The advantage of the standardized estimate is that it produces a comparable metric for the variables. When the model has two or more incoming arrows, path coefficients then become partial regression coefficients which measure the extent of effect of one variable upon another in the path model controlling for other prior variables, by using the correlation or covariance matrix as an input. Thus, the beta for any path is a partial weight control for other given dependent variables. The residual error terms represent the unexplained variance plus measurement errors. The direct effect is that the partial coefficient for y on x controls all prior variables and all mediators in the models. The indirect effect is the total effect minus the direct effect, and measures the effect of the mediators. To ensure that the individual factor loading is significant, the t-value should be greater than 1.96. If a model is correctly specified, i.e. relevant variables and arrows are correctly indicated, then the sum of standardized path values should be equal to the correlation coefficient. Therefore, goodness-of-fit should be close to 1 (Garson, 2007).

The path coefficients are often estimated by the parametric Maximum Likelihood (ML) method. One limitation of ML is that the random variables should have multivariate normality. That means each indicator should be normally distributed for

each value of each and every other indicators; whilst each dependent latent variable should be normally distributed for each value of each and every other latent variables. However, violation of normality, especially high kurtosis, can result in poor estimates and incorrect standard errors and hypothesis tests, even if the sample size is large. According to Jöreskog and Sörbom (1993), ML method produces asymptotically incorrect standard errors of parameter estimates and incorrect chi-squares under non-normality situations.

Kline (1998) reported that under severe non-normality of data, path estimates are fairly accurate. However, the model is biased toward Type I errors and standard errors are deflated. These underestimated standard errors may exaggerate the t-values and the significance of the regression paths and factor/error covariances than they should be. It is better to use the Weighted Least Squares (WLS) estimation method which produces correct standard errors of parameter estimates and chi-squares. A correct weight matrix is the inverse of the estimated asymptotic covariance matrix of the estimated covariance matrix (Jöreskog & Sörbom, 1986a; 1986b). However, WLS requires sample sizes of larger than 1,000. The other remedy is to use Browne's asymptotic distribution-free (ADF) estimation method rather than WLS if the sample size is large enough (Browne, 1984; Browne & Cudeck, 1993).

According to Curran, West and Finch (1996), as Satorra & Bentler rescaled chi-square (χ^2_{SB}) works well in nearly every condition across sample sizes, this study will use asymptotic covariance matrix with adherence to the Satorra & Bentler rescaled chi-square (χ^2_{SB}) (Satorra & Bentler, 2001) even though the number of cases is larger than 1,000.

3.5.2 Measures for analysis

To decide whether the model fits in or not, the chi-square change ($\Delta\chi^2$) is used as a criterion measure. However, a well-known fact is that chi-square statistics are very sensitive to sample size, and the result of which may lead to the rejection of the null hypothesis that the model is a good fit to data even with a small discrepancy (Kline,

1998). Nevertheless, a large sample size can identify the relatively small effect sizes and increase the statistical power of the conclusions. The concept of statistical power in statistical theory is defined as the probability of rejecting the null hypothesis given that the null hypothesis is false, i.e. 1- type II error (Kaplan, 1995). Using a small sample size has the problems of convergence failures, insufficient power for model confirmation and lack of generalizability (Doest & Jonge, 2006). For example self-concept in this study was one of the sensitive variables which was hard to collect and required a large sample size to identify the effect sizes (Strein, 1993).

Thus, Jöreskog (1969) proposed using the normed chi-square (χ^2/df) statistics rather than the χ^2 statistic if the sample size is unusually large although there is no definition of what is meant by “large”. For reference, Cheung and Rensvold (2000:201) ‘disregarded the χ^2 statistic’ as a reliable measure with a sample size of 1,371 cases. Eventually, Normed chi-square (χ^2/df) ratio and Incremental Fitness Measures like CFI, NNFI, AGFI and PGFI are calculated for reference as well. Carmines and McIver (1981) and Kline (1998) recommended a Normed chi-square ratio of 3 or less to indicate an acceptable fit for large sample analysis.

Bollen (1989) also suggested reporting multiple goodness of fit tests to evaluate different aspect of these models, these being the chi-square goodness of fit test; the root mean squared error of approximation (RMSEA), standardized root mean square residual (RMSR), the adjusted goodness of fit index (AGFI; Jöreskog & Sörbom, 1993), comparative fit index (CFI); the non-normed fit index (NNFI; Bentler & Bonett, 1980).

3.5.3 *Criteria for analysis*

In next chapter, the analysis procedures will be: (1) to screen out multivariate normality through the data; to delete the missing data listwise and check reliabilities; (2) to use the Exploratory Factor Analysis (EFA) to assess measurement properties of the items; (3) to check the measurement invariance of the measurement model across gender; (4) to use the Confirmatory Factor Analysis (CFA) to assess the

dimensionality and construct validity; (5) to use the structural model analyses to test the fitness of two competing nested models; and finally, to discuss the relationships between independent and the dependent variables of the best fit model.

A good fit model should have Tucker Lewis Index or Non-Normed Fit Index [NNFI] > 0.95; Bentler Comparative Fit Index [CFI] > 0.95; Root Mean Square Error of Approximation [RMSEA] < 0.05; and Normed Chi-square < 3 (Byrne, 1989; Hu & Bentler, 1999). According to Jöreskog and Sörbom (1993), a goodness of fit index of 0.95 indicated that the sum of squares of the residuals of the estimated model was only 5 percent of the total sum of squares of the measured covariance. In other words, the implied model had reduced the total sum of squares of the measured covariance by 95 percent.

Some final remarks are stated here. First, this cross-sectional study deals with association but not causation of variables. It only tells which one of the competing theories can best fit the observed covariance or correlation matrix according to the hypothesized theories. Causation can only be analyzed by longitudinal study. Second, the researcher should test a priori model, but not to conclude findings which are solely based on the results capitalized on chance.

3.6 Reliability, Validity and Ethical Issues

The research instrument is good if the sample scores are stable and consistently reliable. Then, they are valid to draw meaningful and useful inferences. Reliability depends on the extent to which experiments or procedures yield identical outcomes repeatedly. Validity depends on the extent to which researchers can draw meaningful, justifiable inferences from scores of a sample or population.

Three types of reliability are important: (1) consistency; (2) alternate-forms and (3) test-retest (Malhotra, 2004). For consistency in reliability, if each item measures an aspect of a construct, there should be a consistent set of items in form of a scale. The simplest measure of internal consistency is split-half reliability. Items are scored as

continuous variables (from strongly agree to strongly disagree) and divided into halves; then a correlation coefficient can be found between the two halves. The average of all possible split-half coefficients resulting from scale items as split in different ways is defined as coefficient alpha or Cronbach alpha (Cronbach, 1951). It is between minus infinity and 1, with 0.6 considered acceptable for exploratory purposes, 0.7 considered adequate for confirmatory purposes, and 0.8 considered good for confirmatory purpose (Nunnally, 1978). However, Cronbach alpha has some disadvantages. It tends to increase with the number of scale items. It is not suitable for measuring multi-dimensional constructs, and reliability should be calculated for each dimension separately. Therefore, Cronbach alphas will be calculated for each factor separately to measure its internal consistency.

To further ensure the reliability and consistency, PISA (2003) recruited fifty-one test administrators to administer the assessments at schools and 16 markers for marking. Open-ended response questions of 900 booklets were selected for multiple marking by four markers in order to ensure inter-marker's reliability. Detailed guidelines and training were provided to ensure the reliability and international comparability. In total, 85 questions were asked in PISA 2003 and the final score was scaled to a mean of 500 and a standard deviation of 100 (OECD, 2003). In order to ensure reliability, participating countries were required to sample 160 schools and, within those schools, 5500 students. To ensure comparability of samples across countries, PISA set the minimum acceptable response rates of not less than 85% at the school level and 80% at the student level. It fitted in the purpose of triangulation by interviewers and observers (Denzin, 1970). The measurement instruments were obviously the collaboration effort of international experts and thus assumed to be highly reliable. In addition, Cronbach alpha and lambdas will also be tested in the next chapter.

For alternative-forms reliability, two equivalent forms of the scale are designed, and answered by the same group of respondents two to four weeks after the first interview. Results are correlated to the assessment of reliability. However, it is time-consuming, expensive and difficult to design. For test-retest reliability, respondents are assessed at two different points of time using the same research instrument. However, the longer

the time interval, the lower the reliability would be. Also, the first measurement may alter the characteristics being measured, and affect the result of the second test. Moreover, carry-over effects may occur as respondents try to remember previous answers. These two forms of reliability were not used in PISA 2003 study.

Validity is the extent to which researchers can draw meaningful, justifiable inferences. Traditional validity discussion feature the ‘three C’s’: Content, Criterion, and Construct validity. Content validity is a subjective but systematic evaluation of research instruments by researchers, colleagues, or other professionals based on established theories for support. They evaluate whether scaled items adequately cover the entire domain of a construct being measured. Criterion Validity measures whether scores from an instrument are good predictors of some outcomes, e.g. GRE scores are expected to predict graduate school performances. Construct validity is a determination of the significance, meaning, purpose and use of scores generated from an instrument. It requires a sound measure of the theory and nature of the construct, how it relates to other constructs and its convergent, discriminant and nomological validity (Malhotra, 2004).

Convergent validity measures the extent to which the scale correlates positively with other measures of the same construct (Churchill, 1987). Correlation of a quantitative test with the final grade in a math course would be an example. Discriminant validity measures the extent to which a measure does not correlate with other constructs from which it is supposed to differ (O’Leary-Kelly & Vokurka, 1998). This correlation should be close to zero. Nomological validity is the extent to which the scale correlates, in theoretically predictable ways, with measures of different but related constructs. Perfect validity implies perfect reliability but not vice versa, since systematic errors may still exist. Generalization cannot be done without the backup of reliability and validity.

Convergent validity is also reflected by comparative fit index (CFI) and normed fit index (NFI) values of 0.95 or above (Ahire & Golhar, 1996). Both indexes equates to 1.00 in this model which reconfirmed the model has convergent validity. Discriminant

validity can be judged by fixing the correlations between various constructs to 1.0, then re-estimating the fixed model. A significant difference of the chi-square statistics between the fixed and unconstrained models ($\Delta\chi^2 > 6.635$) indicates high discriminant validity (Chau, 1997; Yeung, Cheng & Chan, 2004). These will be demonstrated in chapter 4 in detail.

Regarding the ethical issues, respondents have three rights: confidentiality, anonymity and safety. Confidentiality is an interviewer's promise that the respondent will not be identified, or his or her responses presented in an identifiable form. Anonymity is the assurance that even the interviewer cannot trace the sources of responses from the respondents (Sapsford & Abbot, 1996). Safety is the precaution against outspoken respondents from being harmed. The primary risk on using PISA data is the violation of confidentiality. However, since the data are recorded without any identifiers and it is publicly accessible from the website, the risk is reduced because the data are anonymous (PISA Technical Report, 2003).

Chapter 4 FINDINGS OF THE STUDY

4.1 Preliminary Analyses

4.1.1 Skewness and kurtosis indexes

Many classical statistical tests and intervals depend on normality assumptions. Significant skewness and kurtosis clearly indicate that data are not normal. Skewness is defined as how unevenly the data is distributed, in the manner of score clusters piling up on one side of the distribution and a few taper at either tail of the distribution. Preliminary analyses were calculated by using SPSS v.17, in which 0.0 indicates a perfect normal distribution. The scores may also be negatively skewed to the right or positively skewed to the left (Creswell, 2002). Fox (1997) suggested using power transformation (x to x^2) for correcting negatively skewed variables, and log transformation (x to $\log x$) for correcting positively skewed variables.

Kurtosis is defined as how protruded or flat a distribution looks. If too many scores pile up around the mean, then the distribution has a peak or peaks and thus it is not normal with positive kurtosis. The same logic applies to a distribution which is too flat with negative kurtosis. One well-known fact is that the larger the sample size, the easier it is to violate these skewness and/or kurtosis requirements with just small deviations. In this study, the sample size was large. However, preliminary analyses of all measures revealed no outliers; and skewness and kurtosis indexes were within acceptable limits as shown in Table 2 (Meehan, Hughes & Cavell, 2003). Descriptive statistics for continuous variables are also presented.

Table 2. Descriptive statistics for continuous variables

| | Mean | Std. Dev. | Skewness | Kurtosis |
|---|--------|-----------|----------|----------|
| Q26b. Most teachers are interested in students' well-being. | 2.33 | .68 | .43 | .16 |
| Q26c. Most of my teachers really listen to what I have to say. | 2.31 | .69 | .44 | .19 |
| Q26e. Most of my teachers treat me fairly. | 2.20 | .68 | .74 | .94 |
| Q30a. I enjoy reading about mathematics. | 2.73 | .78 | -.16 | -.40 |
| Q30d. I do mathematics because I enjoy it. | 2.48 | .83 | .07 | -.55 |
| Q30f. I am interested in the things I learn in mathematics. | 2.49 | .82 | .04 | -.51 |
| <i>How confident do you feel about having to do the following Mathematics tasks? (Q31)</i> | | | | |
| Q31a. Using a train timetable to work out how long it would take to get from one place to another | 2.01 | .74 | .38 | -.14 |
| Q31c. Calculating how many square meters of tiles you need to cover a floor | 1.85 | .82 | .62 | -.38 |
| Q31f. Finding the actual distance between two places on a map with a 1:10,000 scale | 2.10 | .90 | .32 | -.83 |
| Q32d. I get good marks in Mathematics. | 2.94 | .77 | -.38 | -.20 |
| Q32f. I learn math quickly. | 2.59 | .73 | .00 | -.32 |
| Q32i. In my math class, I understand even the most difficult work. | 2.83 | .74 | -.24 | -.22 |
| Q37b. In Mathematics I enjoy working with other students in groups. | 2.20 | .71 | .44 | .29 |
| Q37d. When we work on a project in Mathematics, I think that it is a good idea to combine the ideas of all students in a group. | 2.28 | .69 | .43 | .23 |
| Q37f. I do my best in Mathematics when I work with other students. | 2.29 | .70 | .24 | -.05 |
| Does your mother have any of the following qualifications? (ISCED) | 2.05 | 1.39 | .85 | .51 |
| Does your father have any of the following qualifications? (ISCED) | 2.30 | 1.49 | .86 | .25 |
| Which of the following do you have in your home?1 (Yes) 0 (No) (WLE) | -.22 | 1.05 | -1.00 | .74 |
| Plausible value in math | 555.02 | 98.06 | -.46 | .31 |
| Plausible value in math | 555.28 | 96.83 | -.41 | .14 |
| Plausible value in math | 556.83 | 97.46 | -.39 | .12 |
| Plausible value in math | 556.00 | 97.01 | -.42 | .19 |
| Plausible value in math | 556.18 | 97.90 | -.39 | .18 |

4.1.2 *Exploratory Factor Analysis*

A rigorous process was used to develop and validate the survey instrument. First, we use exploratory factor analysis (EFA) to ensure unidimensionality for such an exploratory study in Hong Kong. The EFA was conducted at the cross-factor level. We use principal component analysis for all measurement items, which determined the main constructs and their related measurement items. Varimax rotation with Kaiser normalization was used to clarify the factors (Loehlin, 1998). A seven factor solution was deemed to be the most interpretable. The seven factors were labeled as F1= Teacher emotional support; F2= Math self-concept; F3= Math self-efficacy; F4= Interest in math; F5= Cooperative learning; F6= Parental education level; and F7= Math scores as in Table 3. The last factor F8 is Resource which is a single indicator variable.

The covariances among the items in the model are shown in Table 4. The covariance was used rather than the correlation matrix in this study. The covariance between two variables is the product of the two variables' standard deviation and their Pearson correlation coefficient. The flexibility associated with covariance matrix was shown to outweigh the small biases caused by ordinality and nonnormality problems. Besides, the correlation matrix is known to be problematic (Cudeck, 1989; Hoyle & Panter, 1995). Cudeck (1989) demonstrated the latter's adverse effects on test statistics and standard errors of estimates.

After recoding the questions, most of the covariances are positively varied with the math scores. Since indicators for the construct Resource included thirteen dichotomous questions with 1 for 'yes' and 0 for 'no', a composite variable HEDRES is computed in the PISA database to represent how many home educational resources are available to the student. It has a positive covariance ranging from 1.01 to 41.46 with math scores, meaning that, the higher the cumulative scores for these thirteen questions, the higher the math performance of students would be.

Table 3. Exploratory factor analysis of seven constructs

| | Factor loading | | | | | | |
|---------------------------------|----------------|------------------|-------------------|--------------------|---------------------------|----------------------|--------------------------|
| Items* | Math score | Interest in math | Math self-concept | Math self-efficacy | Teacher emotional support | Cooperative learning | Parental education level |
| Interested in Students Q26b | .027 | .110 | .019 | .047 | .798 | .074 | -.023 |
| Listen to me Q26c | .010 | .041 | .048 | .034 | .827 | .096 | .002 |
| Treat me fairly Q26e | -.083 | .032 | .026 | .009 | .752 | .079 | .028 |
| Attitude enjoy reading Q30a | -.130 | .778 | .311 | .144 | .064 | .153 | .015 |
| Attitude enjoy Maths Q30d | -.151 | .793 | .290 | .155 | .102 | .191 | .016 |
| Attitude interested Q30f | -.142 | .804 | .311 | .139 | .107 | .214 | -.010 |
| Confident timetable Q31a | -.268 | .105 | .149 | .761 | .044 | .109 | -.020 |
| Confident area Q31c | -.319 | .156 | .146 | .737 | .040 | .107 | -.056 |
| Confident distance Q31f | -.313 | .153 | .190 | .713 | .035 | .086 | -.012 |
| Feel study good <marks> Q32d | -.194 | .307 | .737 | .148 | .033 | .092 | -.011 |
| Feel study quickly Q32f | -.164 | .282 | .777 | .174 | .038 | .122 | -.015 |
| Feel study underst. diffc. Q32i | -.111 | .245 | .801 | .155 | .045 | .120 | -.020 |
| Attitudes group work Q37b | -.075 | .159 | .054 | .073 | .066 | .776 | .006 |
| Attitudes project Q37d | -.068 | .133 | .095 | .053 | .120 | .727 | -.033 |
| Attitudes work with other Q37f | -.019 | .094 | .114 | .106 | .076 | .776 | .023 |
| Educational level of mother | .078 | .011 | -.010 | -.027 | -.007 | -.008 | .890 |
| Educational level of father | .092 | .002 | -.021 | -.034 | .016 | .005 | .887 |
| Plausible value in math | .929 | -.094 | -.114 | -.202 | -.023 | -.051 | .055 |
| Plausible value in math | .931 | -.100 | -.115 | -.191 | -.015 | -.050 | .044 |
| Plausible value in math | .930 | -.098 | -.115 | -.190 | -.013 | -.057 | .053 |
| Plausible value in math | .933 | -.094 | -.105 | -.191 | -.015 | -.053 | .062 |
| Plausible value in math | .928 | -.104 | -.115 | -.197 | -.025 | -.053 | .062 |

- See Table 2 for the survey questions of the measurement items.

If an item was loaded on more than one factor and the difference between factor loadings was less than 0.10 across factors, the item is considered as cross-loaded (e.g. Jambulingam et al., 2005). Table 3 shows that each item was well loaded on a single construct (i.e. no item was cross-loaded). The cross-factor level EFA resulted in seven eigenvalues that were greater than one as in Table 5. The total variance explained by the seven factors was 74.22%. Therefore, all seven factors extracted from EFA represent their corresponding items well.

4.1.3 Scale Reliabilities

We use Cronbach's alpha which is a widely accepted reliability measure. As shown in Table 5, all the measures of our instrument were found to be highly reliable with Cronbach alphas ranged from 0.70 to 0.98. All values are not less than the cut-off criterion 0.7 ((Nunnally 1978; Nunnally & Bernstein, 1994). In addition, the Corrected Item-Total Correlation (CITC) reliability test is also used (Kerlinger, 1986). Table 5 shows that all CITC values were larger than the minimum acceptable value of 0.30.

We further assess the reliability of an indicator by checking the proportion of variance (R-square) in the observed variables, accounted for by the theoretical constructs influencing them. All the R-square values, ranged from 0.34 to 0.91, were above 0.30 which were considered acceptable (e.g. Carr & Pearson, 1999) and confirmed that all indicators are significantly related to their underlying theoretical constructs. Based on the Cronbach's alpha values, CITC values and R-square values, we triangulated and concluded that the scales were reliable.

Table 5. Results of measurement validation

| Scale name | Variable name | CITC. | Scale statistics |
|---------------------------|---------------|-------|---|
| Teacher emotional support | Q26b | .546 | Cronbach's alpha: 0.72, |
| | Q26c | .585 | Largest eigenvalue (variance explained): 1.92 (64%), |
| | Q26e | .484 | Second largest eigenvalue (variance explained): 0.62 (21%) |
| Interest in math | Q30a | .724 | Cronbach's alpha: 0.88, |
| | Q30d | .767 | Largest eigenvalue (variance explained): 2.41 (80%), |
| | Q30f | .802 | Second largest eigenvalue (variance explained): 0.35 (12%) |
| Math self-efficacy | Q31a | .595 | Cronbach's alpha: 0.77, |
| | Q31c | .630 | Largest eigenvalue (variance explained): 2.07 (69%), |
| | Q31f | .609 | Second largest eigenvalue (variance explained): 0.49 (16%) |
| Math self-concept | Q32d | .655 | Cronbach's alpha: 0.82, |
| | Q32f | .702 | Largest eigenvalue (variance explained): 2.21 (74%), |
| | Q32i | .671 | Second largest eigenvalue (variance explained): 0.43 (14%) |
| Cooperative Learning | Q37b | .527 | Cronbach's alpha: 0.70, |
| | Q37d | .478 | Largest eigenvalue (variance explained): 1.86 (62%), |
| | Q37f | .515 | Second largest eigenvalue (variance explained): 0.61 (20%) |
| Math scores | PV1MATH | .947 | Cronbach's alpha: 0.98, |
| | PV2MATH | .947 | Largest eigenvalue (variance explained): 4.17 (93%), |
| | PV3MATH | .946 | Second largest eigenvalue (variance explained): 0.09 (1.7%) |
| | PV4MATH | .949 | |
| | PV5MATH | .948 | |
| Parental education level | MISCED | .60 | Cronbach's alpha: 0.75, |
| | FISCED | .60 | Largest eigenvalue (variance explained): 1.60 (80%), Second largest eigenvalue (variance explained): 0.4 (20%) |

4.1.4 *Test for Multivariate Multicollinearity*

To re-assure that multicollinearity did not exist and cause ‘attenuated coefficients’, i.e. negative betas despite positive bivariate correlations (Patrick, Ryan & Kaplan, 2007:91), variance inflation factor values were calculated and are presented in Table 6. Inspection of the correlation matrix shows only bivariate multicollinearity, whereas inspection of the variance inflation factor (VIF) reveals multivariate multicollinearity. When VIF is 4 or above, there is high multivariate multicollinearity and the beta coefficients will not be stable (see Kline, 1998). In this study, all the VIF values were found not to exceed 4, below the common cut-off criterion and indicating no multicollinearity problem or serious threat to the power of this study. The only concern was the construct, interest in math, which has relatively higher VIF than the others.

4.2 Measurement Invariance

Evidence of measurement invariance underpins scientific inference and lacking this evidence equivocates conclusions and cast doubt upon the theory (Horn & McArdle, 1992). However, it is usually not presented in many studies due to its complicated and time-consuming requirements. The PISA 2003 Hong Kong dataset used in this study has 4478 cases. Sample sizes after deleting missing values listwise were 2,066 and 2,152 for male and female respectively. It is necessary to test the measurement invariance of the Hong Kong dataset.

Latent means and covariances were calculated by using LISREL 8.5 which is computer software used for SEM (Jöreskog & Sörbom, 2001). Due to the collective nature of Hong Kong society (Hofstede, 2004), the author started with ‘a test of the equality of covariance matrices and mean vectors, both separately and jointly’ (Steenkamp & Baumgartner, 1998:82). The covariance matrix and Robust Maximum Likelihood (RML) were used to test the model. RML was used instead of ML because it takes multivariate nonnormality into consideration (Satorra & Bentler, 2001).

Table 6. The Variance Inflation Factor for Testing Multivariate Multicollinearity

| Coefficients(a) | | | | | | | | |
|---|-------------------------------------|-----------------------------|------------|---------------------------|--------|------|-------------------------|-------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | Collinearity Statistics | |
| | | B | Std. Error | Beta | | | Tolerance | VIF |
| 1 | (Constant) | 326.243 | 9.032 | | 36.122 | .000 | | |
| | Interested in Students Q26b | -6.570 | 2.124 | -.047 | -3.093 | .002 | .670 | 1.491 |
| | Listen to me Q26c | -4.908 | 2.142 | -.035 | -2.292 | .022 | .641 | 1.560 |
| | Treat me fairly Q26e | 8.163 | 2.020 | .057 | 4.040 | .000 | .754 | 1.326 |
| | Attitude enjoy reading Q30a | .571 | 2.291 | .005 | .249 | .803 | .438 | 2.285 |
| | Attitude enjoy Maths Q30d | 4.483 | 2.348 | .039 | 1.909 | .056 | .370 | 2.700 |
| | Attitude interested Q30f | .252 | 2.546 | .002 | .099 | .921 | .327 | 3.062 |
| | Confident timetable Q31a | 19.699 | 2.034 | .151 | 9.683 | .000 | .622 | 1.608 |
| | Confident area Q31c | 25.403 | 1.924 | .214 | 13.201 | .000 | .576 | 1.735 |
| | Confident distance Q31f | 21.918 | 1.702 | .205 | 12.881 | .000 | .596 | 1.678 |
| | Feel study_good <marks> Q32d | 10.496 | 2.173 | .083 | 4.830 | .000 | .510 | 1.961 |
| | Feel study_quickly Q32f | 6.389 | 2.378 | .048 | 2.687 | .007 | .466 | 2.144 |
| | Feel study_underst. diffc. Q32i | -3.465 | 2.240 | -.027 | -1.547 | .122 | .512 | 1.954 |
| | Attitudes group work Q37b | 3.013 | 2.024 | .022 | 1.488 | .137 | .697 | 1.435 |
| | Attitudes project Q37d | -.728 | 2.011 | -.005 | -.362 | .718 | .736 | 1.359 |
| | Attitudes work with other Q37f | -7.594 | 2.008 | -.055 | -3.783 | .000 | .709 | 1.410 |
| | Educational level of mother (ISCED) | 1.173 | 1.076 | .017 | 1.090 | .276 | .632 | 1.583 |
| | Educational level of father (ISCED) | 3.501 | 1.009 | .054 | 3.472 | .001 | .633 | 1.580 |
| | Home educational resources (WLE) | 14.105 | 1.215 | .150 | 11.613 | .000 | .906 | 1.104 |
| a Dependent Variable: Plausible value in math | | | | | | | | |

Encouragingly, the fit for full invariance was satisfactory. Although the chi-square was significant ($\chi^2_{SB}(496) = 897.43$, $p < 0.001$), the RMSEA of 0.02 indicated an acceptable fit, and the two other fit indices were also above the commonly recommended level (CFI= 0.99, NNFI = 0.99). ΔCFI was examined to determine the invariance and a ΔCFI value less than or equal to -0.01 indicates that invariance should not be rejected (Cheung & Rensvold, 2000).

All twenty-three factor loadings were significant across gender. All standardized factor loadings, ranged from 0.58 to 0.98 as in Table 7, exceeded the minimum

Table 8. The differences of latent means across gender in H.K.

| | Teacher emotional support | Interest in math | Math self- efficacy | Math self- concept | Cooperative learning | Parental education level | Home education resources | Math scores |
|-----------------------------------|---------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|--------------------------------|--------------------------------|------------------------------|
| Latent means for females | 0.03 (2.82) [0.01] | -0.26 (3.34) [-0.08] | -0.32 (3.43) [-0.09] | -0.25 (2.96) [-0.08] | -0.03 (2.75) [-0.01] | -0.07 (5.38) [-0.01] | ----- | -18.5 (514.52) [-0.04] |

Note. Standard errors are given in (.). T-values are given in [].

Then the author cross validated the model with separate samples. The importance of cross-validation of modified models to prevent post-hoc capitalization on chance has been repeatedly emphasized in the literature but is usually omitted in those less stringent studies (MacCallum et al., 1992). By comparing different observers, theories, methods and data, researchers can attenuate the weakness or biases and the problems coming from single method, interviewer, source and theory. Cross validation in survey is a triangulation process to increase the external validity or generalizability. Denzin (1970) suggested several types of triangulation like investigator triangulation, theory triangulation and source of data triangulation.

Triangulation by different investigators was done by PISA in 2003. PISA recruited fifty-one test administrators to administer the assessments at schools and 16 markers for marking. Open-ended response questions of 900 booklets were selected for multiple marking by four markers in order to ensure inter-marker's reliability. Detailed guidelines and training were provided to ensure its reliability and international comparability. Theory triangulation will be done by comparing two competing nested models in section 4.4.

Data triangulation was done here. To make this measurement invariance more stringent, cross-validation against the PISA 2003 all countries data ($N = 220,617$) was done to verify that this measurement model does not capitalize on chance and could be generalized to other samples. Since PISA data is international and allows

cross-national comparisons, all data from 41 countries are divided into two genders: 108,220 cases and 112,397 cases for male and female respectively after listwise deletion.

The fit for full invariance of this measurement model was found satisfactory. Although the chi-square is significant ($\chi^2_{SB}(496) = 35142$, $p < 0.001$), the RMSEA of 0.025 indicates an acceptable fit, and the two other fit indices are also above the commonly recommended level (CFI = 0.99, NNFI = 0.99). All twenty-three factor loadings are significant across gender. All standardized factor loadings, ranging from 0.57 to 0.98, exceed the minimum loading 0.48 (Steenkamp & Baumgartner, 1998) and no more relaxation is necessary. Compared to the validation mode, the chi-square increase was significant ($\Delta\chi^2 = 34047$, $p < 0.001$) which was possibly inflated by sample size as expected ($N = 220,617$) (Bentler & Chou, 1987; Kelloway, 1998) because the fit indices did not deteriorate at all. Although almost all of the latent means for females are lower than those of males, they are not significant since all t-values are smaller than 1.96. This finding is consistent with the research findings of a lot of the research cited in Kiamanesh (2004) as follows:

Pintrich and DeGroot (1990), Randhawa, Beamer and Lundberg (1993), Pajares and Miller (1994) and Pajares (1997) reported that boys have higher math self-efficacy than girls do. Pajares and Miller (1994) and Skaalvik and Rankin (1992) reported that boys have higher math self-concept than girls do. Pajares and Miller (1994) reported that boys have higher math performance than girls do. Davis-Kean, Eccles and Linver (2002) found that boys have higher interest in math than girls do, and the difference in math performance between the genders is not significant.

4.3 Confirmatory Factor Analysis (CFA)

Since measurement invariance between genders had been confirmed in the last section, male and female samples could be aggregated together for CFA. Anderson and Gerbing (1988) and Hair et al. (1995) proposed a two-step approach for SEM. The first step is to assess the measurement model independently and then evaluate the structural model. Otherwise, it is impossible to discriminate how much of the poor fit is due to the measurement model. They also believed that improving the measurement model could first improve estimates of structural parameters and then the overall goodness of fit. The measurement model was tested by LISREL 8.5. Confirmatory Factor Analysis (CFA) seeks to determine if the number of constructs and the loadings of indicators conform to the expectations of prior theories. Two latent exogenous variables (ξ), parental education level and teacher emotional support, are hypothesized and indicators are loaded on each construct according to previous research. Domestic education resources, math self-efficacy, math self-concept, interest in math, cooperative learning and score are endogenous latent constructs (η).

However, these constructs cannot be verified by CFA individually. This is because a single measurement model consisting of a single construct with two indicators has a negative degree of freedom, resulting in an under-identified model. Identification is the concept that a unique solution can be calibrated for the model. Overidentification associated with positive degree of freedom is necessary for a solution to converge. Degree of freedom equals the difference between the number of elements of the covariance matrix and the number of parameters to be estimated. The number of elements of the covariance matrix equals to $n(n+1)/2$ where n is the number of variables in the model.

Similarly, a single construct with three indicators has a zero degree of freedom. That is, the number of elements of covariance matrix equals the number of parameters to be estimated, resulting in a just-identified model. In that case, chi-square is also zero and p-value cannot be computed. Thus, these latent constructs are tested together as in

Fig. 5. The author would like to sort out whether measures created to represent a latent variable really align with each other. A curve arrow between each pair of latent exogenous variables in Fig. 5 represents their covariance. The underlying latent variable is hypothesized to interact with the observed variable. A straight arrow from the latent variable to the observed variable represents the standardized factor loading, λ . Another straight arrow directing to their respective indicators represents the error and disturbance term, δ .

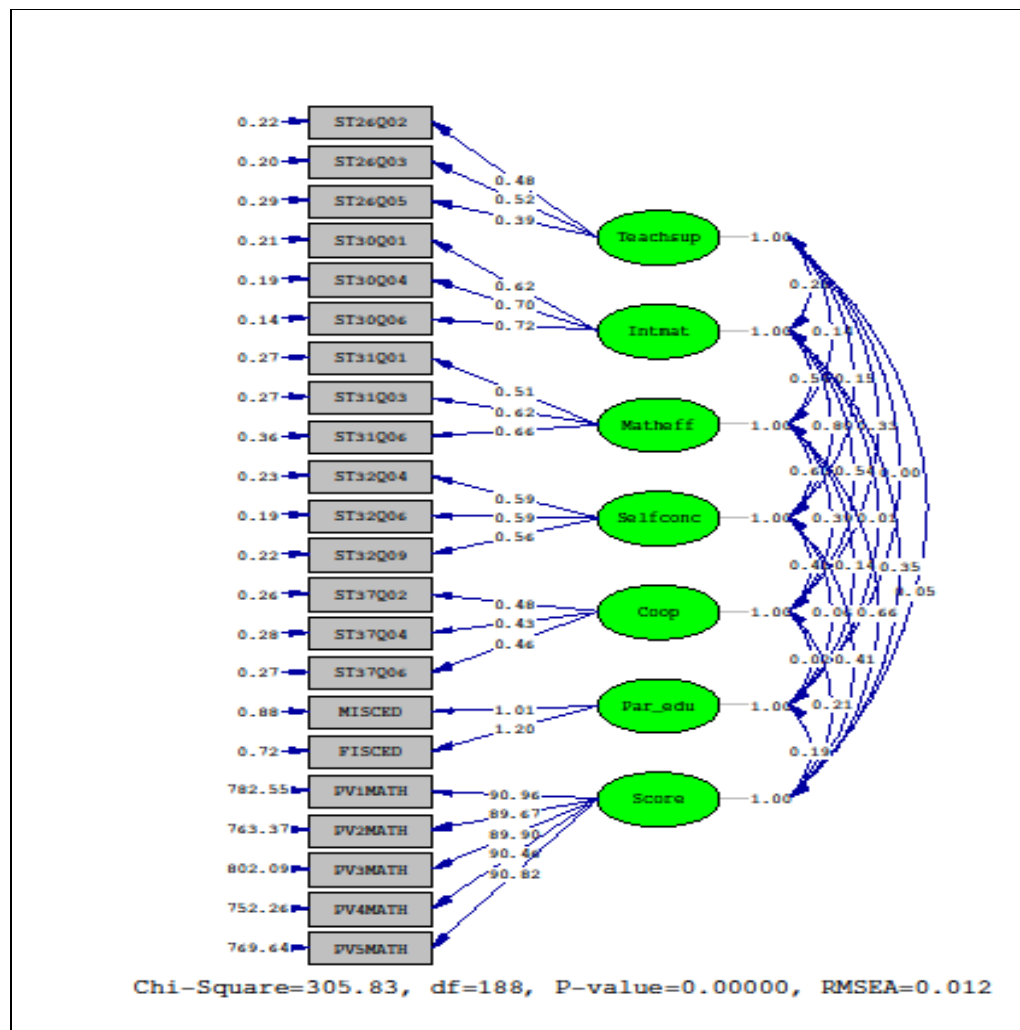


Figure 5. Confirmatory Factor Analysis

To restate, the model is fit if RMSEA < 0.05, AGFI > 0.95, NNFI > 0.95, CFI > 0.97 and $(\chi^2/df) < 3$ (Byrne, 1989; Carmines & McIver, 1981; Hoyle & Panter, 1995; Hu & Bentler, 1999). In this study, CFA found $\chi^2_{SB}(188) = 305.83$ ($p < 0.001$), NNFI = 1.00, CFI = 1.00, AGFI = 0.98, RMSEA = 0.012 and $\chi^2/df = 1.62 < 3$ as depicted in Table 9. The absolute goodness of fit values are well above 0.95 and the RMSEA value is below 0.05, suggesting a good fit between the implied covariance in the model and the observed covariance from the data. Comparative fit measures are also well above the general criteria, providing evidence against the hypothesis of nullity. All these measures suggested that the measurement model had a good model fit. The significant chi-square statistics of this measurement model is possibly inflated by the large sample size ($N = 4,218$).

Table 9. Goodness-of-fit indices for the measurement model

| <i>Goodness of fit measures</i> | Criteria | Result |
|---|-------------|--------|
| Satorra-Bentler Scaled Chi-square (χ^2_{SB}) | Lowest | 305.83 |
| <i>Absolute Fit Measures</i> | | |
| Probability of χ^2 (p) | ≥ 0.05 | 0.00 |
| Normed Chi-square (χ^2/df) | [1, 3] | 1.62 |
| Root Mean Square Error of Approximation (RMSEA) | ≤ 0.05 | 0.012 |
| Standardized Root Mean Square Residual (RMSR) | ≤ 0.08 | 0.018 |
| <i>Comparative Fit Measures</i> | | |
| Non-Normed Fit Index (NNFI) | ≥ 0.95 | 1.00 |
| Comparative Fit Index (CFI) | ≥ 0.95 | 1.00 |

4.3.1 Construct Validity

All factor loadings in Fig. 5 had t-values of larger than 2.0 and significant at $p < 0.001$, which confirmed the convergent validity (Anderson & Gerbing, 1988; Chau, 1997; Fornell & Larcker, 1981). Besides, all constructs have eigenvalues exceeding 1.0 and all factor loadings exceed the minimum value of 0.30 which provide further evidence of convergent validity for the constructs (Hair et al., 1995; Reines-Eudy, 2000). To

determine the discriminant validity, a stringent method is fixing the correlations between various constructs to 1.0 pairwise. If the chi-square change is significant, then the constructs shows discriminant validity (see Chau, 1997). The results for baseline CFA is $\chi^2_{SB}(188) = 305.83$ ($p < 0.001$). After fixing the covariance, for example, between math self-concept and interest in math from 0.8 to 1.0, the new results are $\chi^2_{SB}(189) = 627.63$ ($p < 0.001$) as shown in Table 10. Similarly, fixing the other phi (ϕ) between the constructs pairwise, the chi-square values have increased significantly, with a range from 321.8 to 1768.62, as well as which they are much larger than 6.635 with an increase in one degree of freedom and therefore, the discriminant validity is re-confirmed (see Yeung, Cheng & Chan, 2004 for detail).

Table 10. Discriminant validity: Chi-square differences between constrained and unconstrained models. All chi-square differences are different at 0.05 level

| | Teacher emotional support | Interest in math | Math self-efficacy | Math self-concept | Cooperative learning | Parental education level | Math scores |
|---------------------------|---------------------------|------------------|--------------------|-------------------|----------------------|--------------------------|-------------|
| Teacher emotional support | - | | | | | | |
| Interest in math | 1082.14 | - | | | | | |
| Math self-efficacy | 1137.24 | 1164.43 | - | | | | |
| Math self-concept | 1134.01 | 627.63 | 925.24 | - | | | |
| Cooperative learning | 792.43 | 762.90 | 821.29 | 854.72 | - | | |
| Parental education level | 1093.46 | 1097.15 | 1073.06 | 1098.01 | 921.50 | - | |
| Math scores | 1132.85 | 2074.45 | 1104.59 | 1486.46 | 905.20 | 1052.92 | - |

Content validity was assured by an extensive review of the literature and the judgment of educational experts for the original PISA study. The criterion validity was achieved by examining the covariance between constructs. The covariance between the criterion variable, math score, and independent variables were all positive (see Table 4), providing evidence of the criterion-related validity of the instrument (Yeung, Lee & Chan, 2003). The dependent variable, math scores, exhibits generally the same directions and magnitudes of correlation with other independent variables as accepted within the education community. Under construct validity, convergent and discriminant validity are achieved. All the lambdas are significantly larger than 0.48 and significant chi-square change after fixing ϕ to 1.0 pairwise.

4.4 Structural Model Analyses

According to Anderson and Gerbing (1988), the second stage was to test the structural models. With measurement model fit, structural models could then be analyzed using the LISREL 8.5 program (Jöreskog & Sörbom, 2001). Structural model testing is better done by comparing different plausible models with different numbers of free parameters. These practices are developed according to prior theories. This approach is consistent with theory triangulation (Denzin, 1970). If the final model is achieved by post hoc modification, it may capitalize on chance and be incomparable with other samples. In other words, it may have a good model fit but be invalid in representing a true theory (Yeung, Cheng & Chan, 2004).

According to Bentler and Chou (1987), a researcher should build and compare several nested models that shed light on the key features of a baseline model. With the number of constructs unchanged, a model with fewer free parameters is called 'nested' within a model with more free parameters (Schmidt & Finnigan, 1992). If a simpler model with fewer free parameters is not a significantly worse fit than the baseline model, it would be chosen as it better represents the theory. The model with the lowest PGFI is preferred as it is more parsimonious (Yeung, Cheng & Chan, 2004).

The literature recommended comparing competing models using rigorous methodologies (Doest & Jonge, 2006). The characteristics of each model tested have been shown in Fig. 2 and 3. There are theoretical arguments and empirical evidence corroborating these two competing models which have been thoroughly discussed in Chapter 2. Model A is developed as the baseline model. Since Model A has one fewer free parameters to estimate than Model B does, therefore, Model A is nested within Model B. Whether the model is a baseline model depends on the literature than on the number of free parameters. The results of structural models analysis are shown in Table 11. A Robust Maximum Likelihoods estimation method is adopted to generate (χ^2_{SB}) as previous sections do. For all models, independence of error terms was specified and factors were allowed to be correlated. Multiple goodness-of-fit tests were again used to evaluate these models. A statistically significant χ^2 can often be produced as an artifact of sample size and small variations in the data (Hu & Bentler, 1995), and therefore ignored.

Table 11. Goodness-of-fit indices of competing nested models

| Goodness of Fit Measures | Criteria | Model A | Model B |
|--|-------------|---------|---------|
| <i>Absolute Fit Measures</i> | | | |
| Chi-square (χ^2_{SB}) of estimated model | Lowest | 563.36 | 512.61 |
| Degrees of Freedom | | 217 | 216 |
| Significant Level of Chi-square (p) | ≥ 0.05 | 0.00 | 0.00 |
| Chi-square/ Degrees of Freedom (χ^2/df) | ≤ 3.00 | 2.60 | 2.37 |
| Standardized Root Mean Square Residual ($RMSR$) | ≤ 0.08 | 0.056 | 0.046 |
| Root Mean Squared Error of Approximation ($RMSEA$) | ≤ 0.05 | 0.019 | 0.018 |
| <i>Comparative Fit Measures</i> | | | |
| Chi-square of Null or independent Model: | | | |
| Non-Normed Fit Index ($NNFI$) | ≥ 0.95 | 1.00 | 1.00 |
| Comparative Fit Index (CFI) | ≥ 0.95 | 1.00 | 1.00 |
| Adjusted Goodness of Fit Index ($AGFI$) | ≥ 0.95 | 0.97 | 0.98 |

The author started the nested models analysis by checking the fit of Model A. The fit of model A was satisfactory by standards of these criteria: CFI=1.00, NNFI= 1.00, AGFI= 0.97, RMSEA= 0.019 and normed chi-square ($\chi^2/df = 2.60$) below 3. The RMSR values are well below 0.08, indicating a low discrepancy between the implied covariance in the model and observed covariance in the data (Yeung, Cheng & Chan, 2004). The nested model B was also checked. Model B was better than model A because of two reasons. First, model B was equally fit by same standards of criteria. Second, the chi-square statistics decreased with one more degree of freedom gained ($\Delta\chi^2 = 541.08 - 563.36 = -22.28$) and χ^2/df ratio was reduced from 2.60 to 2.38. According to Yeung, Cheng and Chan (2004), the chi-square is significant at the 0.01 level if the chi-square statistics decreases by 6.635 or more with a decrease in one degree of freedom. Therefore, model B was preferred since the chi-square statistics decreased was statistically significant.

This hypothesized model as a whole accounts for 43% of the total variance in math scores (i.e. $R^2=0.43$). This is consistent with prior findings. Ferguson (1991) found that domestic factors accounted for 49% of variance in math test score gains in Texas from Grade 3 to 5 students. The R^2 value ‘represents the strength of the joint linear relationships in a model and is a summary measure of the squared multiple correlations across all structural equations in a model’ (Medsker, Williams & Holahan, 1994:442).

4.5 External Validity

Results should be triangulated by different sources of data (Denzin, 1970). Shah and Goldstein (2006: 160) stated that ‘multiple alternate models should be evaluated and cross validated using split or multiple samples’. Thus, assessing the generalizability of model B developed in Hong Kong to other subjects is an important step in establishing external validity of this model. According to Marsh and Peart (1988), self-concept should be subject-specific. Following this logic, math self-concept can only be relevant to math performance or related subjects but not reading. Therefore,

the author cross-validated this model B (Table 12) to problem-solving and reading in order to test the subject specificity.

Table 12. Goodness-of-fit indices for the different subjects

| Subject | χ^2 | df | p | χ^2/df | AGFI | NNFI | CFI | RMSEA |
|-----------------|----------|------|------|-------------|------|------|------|-------|
| Math | 541 | 216 | 0.00 | 2.50 | 0.97 | 1.00 | 1.00 | 0.019 |
| Problem-solving | 536 | 216 | 0.00 | 2.48 | 0.98 | 1.00 | 1.00 | 0.019 |
| Reading | 559 | 216 | 0.00 | 2.59 | 0.97 | 1.00 | 1.00 | 0.019 |

The results in general supported the view that math self-concept is specific to math and problem-solving subjects but not reading since chi-square increased by 18. The possible reason is that math underlies every facet of science and technology (Hubbard, 1999), and thus the changes of chi-square in problem-solving subjects is expected to be not as large as those in reading which is more qualitative in nature. When it is gender-typed, the math subject is perceived as a masculine domain, whereas reading is stereotyped to be belonging to the feminine domains (Ho & Willms, 1996; Schreiber & Chambers, 2002). Surprisingly, the model was a better fit for problem-solving than math, which is a recommended subject for future research.

4.6 Results

The results of the chosen model, Model B, were reported in this section as in Fig. 6 and Table 13. In this study, only two constructs, teacher emotional support and parental education level, were exogenous variables and the rest were endogenous variables.

Results indicated that teacher emotional support is positively and significantly related to cooperative learning ($\beta = 0.32, p < 0.01$), which in turn, is positively related to math self-concept ($\beta = 0.39, p < 0.01$); interest in math ($\beta = 0.28, p < 0.01$); and math self-efficacy ($\beta = 0.10, p < 0.01$). Math self-concept is positively and significantly related to math self-efficacy ($\beta = 0.48, p < 0.01$); interest in math ($\beta = 0.70, p < 0.01$); and math scores ($\beta = 0.08, p < 0.05$). Math self-efficacy is positively and significantly related to math scores ($\beta = 0.63, p < 0.01$). Interest in math is positively and significantly related to math scores ($\beta = 0.09, p < 0.05$). Math scores are positively and significantly related to home education resources ($\beta = 0.17, p < 0.01$). Home education resources are positively and significantly related to math self-efficacy ($\beta = 0.13, p < 0.01$). Parental education level is positively and significantly related to home education resources ($\beta = 0.28, p < 0.01$).

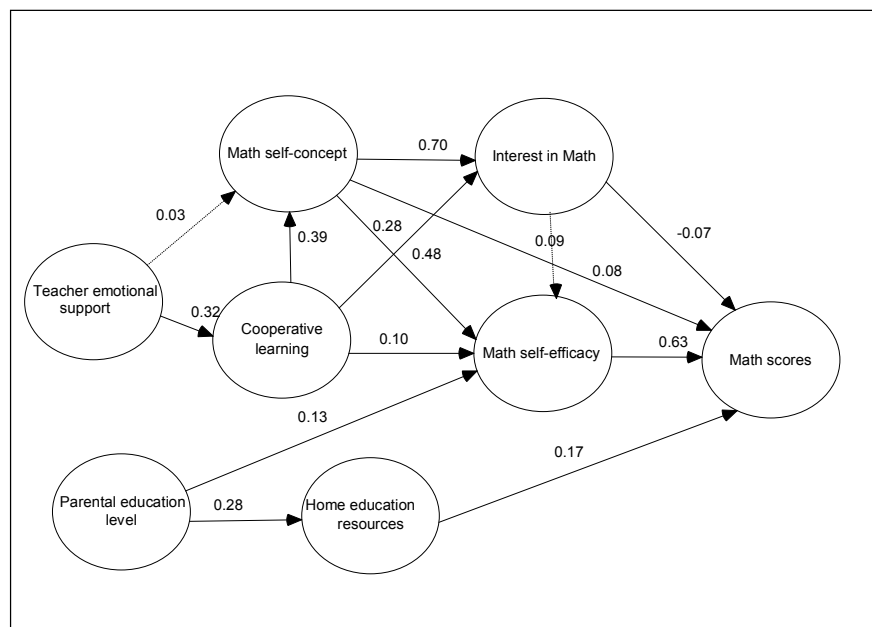


Figure 6. Structural model analyses results

Parental education level is positively and significantly related to home education resources ($\beta = 0.28, p < 0.01$) and math self-efficacy ($\beta = 0.13, p < 0.01$), which in turn,

are positively and significantly related to math scores ($\beta = 0.17, p < 0.01$) and ($\beta = 0.63, p < 0.01$) respectively. However, there are two hypotheses, H1 and H9, which are not supported. The positive associations between teacher emotional support and math self-concept; and between interest in math and math self-efficacy were found insignificant.

Table 13. Relationships among constructs in the final model

| Effect | R^2 | Regression Weights | Standard Error | t- value |
|------------------------------|-------|--------------------|----------------|----------|
| On Math scores | 0.45 | | | |
| of Math self-efficacy | | 0.63 | 0.024 | 26.19 |
| of Home education resources | | 0.17 | 0.014 | 12.26 |
| of Math self-concept | | 0.08 | 0.032 | 2.36 |
| of Interest in math | | -0.07 | 0.029 | -2.34 |
| On Home education resources | 0.08 | | | |
| of Parental education level | | 0.28 | 0.018 | 15.72 |
| On Cooperative learning | 0.10 | | | |
| of Teacher emotional support | | 0.32 | 0.027 | 11.74 |
| On Math self-concept | 0.16 | | | |
| of Cooperative learning | | 0.39 | 0.028 | 13.77 |
| of Teacher emotional support | | 0.03 | 0.023 | insig. |
| On Math self-efficacy | 0.38 | | | |
| of Interest in math | | 0.09 | 0.047 | insig. |
| of Math self-concept | | 0.48 | 0.043 | 11.06 |
| of Cooperative learning | | 0.10 | 0.029 | 3.47 |
| of Parental education level | | 0.13 | 0.017 | 7.28 |
| On Interest in math | 0.72 | | | |
| of Math self-concept | | 0.70 | 0.021 | 34.14 |
| of Cooperative learning | | 0.28 | 0.020 | 13.62 |

Surprisingly, interest in math was found to have a negative relationship with math scores ($\beta = -0.07, p < 0.05$). Interpretation of this finding should be made with caution, keeping in mind the issue of high correlation between interest in math and math self-concept ($r = 0.81$). One possible explanation is that math self-concept acted as a suppressor variable in the model. Suppressing is a statistical phenomenon that can

occur in regression analyses when two predictors are strongly correlated. In such cases, the relation between two variables will be artificially inversed. In SEM, suppressing occurs frequently in models that have latent variables (Massen & Bakker, 2001). The negative path from interest in math to math scores is suggested to be a spurious relationship.

The mediation effects can be further confirmed by Sobel's (1982) asymptotic z test. Baron and Kenny (1986) pointed out that correlations are not sufficient to determine mediation. Mediation occurs when there is a reduction of the effect of the initial variable on the outcome after controlling for the mediator (indirect effect). An indirect effect of a variable x on a variable z through a variable y can be expressed as ab , with a being the effect of x on y and b being the effect of y on z . The Sobel test provides a z statistic for this indirect effect by dividing the product of the regression weights a and b by its estimated standard error, the square root of $(b^2SE_a^2 + a^2SE_b^2)$. The Sobel test revealed significant indirect effects if p value is smaller than 0.05. The Sobel (1982) test statistics was calculated for each path to determine which variables in Fig. 3 mediated the relationships between teacher emotional support and students' math scores; and the relationships between perceived parental education level and students' math scores. The results of the mediation analyses are presented in Table 14.

Comparing Fig. 6 and Table 14, it is apparent that teacher emotional support affected math self-concept and that math self-concept affected interest in math, math scores, and math self-efficacy. However, math self-concept was only a weak mediator of these relationships. Their p values are 0.19, 0.25 and 0.20 respectively. Nevertheless, it is important to note that math self-concept had a significant direct and indirect effect on math scores.

The first major mediation process involved cooperative learning environment as a strong mediator of teacher emotional support on affective and cognitive psychological factors. All p values are smaller than 0.01. In other words, teacher emotional support enhanced cooperative learning environment, which subsequently had impacts on students' math self-concept, math self-efficacy and interest in math.

A second major mediation process involved students' math self-concept as a strong mediator of cooperative learning environment on: (1) interest in math ($p<0.01$); (2) math self-efficacy ($p<0.01$); and (3) math scores ($p<0.05$). A third major mediation process involved students' math self-efficacy as a strong mediator: (1) of cooperative learning environment on math scores ($p<0.01$); (2) of math self-concept on math scores ($p<0.01$); and (3) of parental education level on math scores ($p<0.01$). A final mediation process involved the effects of parental education level on students' home educational resources and the subsequent effect on math scores. Home educational resources is a significant mediator of parental education level on math scores ($p<0.01$).

Results of this study reveal that perceived teacher emotional support affected students' math scores through the mediating role of students' feeling of autonomy and competence; that is, students' competence and autonomy fully mediated the relationship between their perceptions of teacher emotional support and their math scores. These findings also suggest that perceived teacher emotional support, while predicting self-processes such as autonomy and competence feelings, is not directly related to students' math scores. These findings are in line with the SDT.

Results of this study also reveal that perceived parental education level affected students' math scores through the mediating role of students' feeling of competence and home education resources; that is, students' feeling of competence and home education resources fully mediated the relationship between their perceptions of parental education level and their math scores. These findings also suggest that perceived parental education level, while predicting students' feeling of competence and home education resources, is not directly related to students' math scores.

Table 14. Test of hypothesized mediation effects

| Path-a | Beta | Path-b | Beta | Indirect effect | SE_a | SE_b | SE_{a*b} | Z | p value |
|---|------|---|-------|-----------------|--------|--------|------------|-------|---------|
| Teacher emotional support → Math self-concept | 0.03 | Math self-concept → Interest in math | 0.70 | 0.021 | 0.023 | 0.021 | 0.016 | 1.30 | 0.19 |
| Teacher emotional support → Math self-concept | 0.03 | Math self-concept → Math scores | 0.08 | 0.002 | 0.023 | 0.032 | 0.002 | 1.16 | 0.25 |
| Teacher emotional support → Math self-concept | 0.03 | Math self-concept → Math self-efficacy | 0.48 | 0.014 | 0.023 | 0.043 | 0.011 | 1.30 | 0.20 |
| Teacher emotional support → Cooperative learning | 0.32 | Cooperative learning → Math self-concept | 0.39 | 0.125 | 0.027 | 0.028 | 0.014 | 9.03 | 0.00* |
| Teacher emotional support → Cooperative learning | 0.32 | Cooperative learning → Math self-efficacy | 0.10 | 0.032 | 0.027 | 0.029 | 0.010 | 3.31 | 0.00* |
| Teacher emotional support → Cooperative learning | 0.32 | Cooperative learning → Interest in math | 0.28 | 0.090 | 0.027 | 0.02 | 0.010 | 9.05 | 0.00* |
| Cooperative learning → Math self-concept | 0.39 | Math self-concept → Interest in math | 0.70 | 0.273 | 0.028 | 0.021 | 0.021 | 12.85 | 0.00* |
| Cooperative learning → Math self-concept | 0.39 | Math self-concept → Math scores | 0.08 | 0.031 | 0.028 | 0.032 | 0.013 | 2.46 | 0.01* |
| Cooperative learning → Math self-concept | 0.39 | Math self-concept → Math self-efficacy | 0.48 | 0.187 | 0.028 | 0.043 | 0.021 | 8.71 | 0.00* |
| Cooperative learning → Interest in math | 0.28 | Interest in math → Math self-efficacy | 0.09 | 0.025 | 0.02 | 0.047 | 0.013 | 1.90 | 0.06 |
| Cooperative learning → Interest in math | 0.28 | Interest in math → Math scores | -0.07 | -0.020 | 0.02 | 0.029 | 0.008 | -2.38 | 0.02* |
| Math self-concept → Interest in math | 0.70 | Interest in math → Math self-efficacy | 0.09 | 0.063 | 0.021 | 0.047 | 0.033 | 1.91 | 0.06 |
| Math self-concept → Interest in math | 0.70 | Interest in math → Math scores | -0.07 | -0.049 | 0.021 | 0.029 | 0.020 | -2.41 | 0.02* |
| Cooperative learning → Math self-efficacy | 0.10 | Math self-efficacy → Math scores | 0.63 | 0.063 | 0.029 | 0.024 | 0.018 | 3.42 | 0.00* |
| Math self-concept → Math self-efficacy | 0.48 | Math self-efficacy → Math scores | 0.63 | 0.302 | 0.043 | 0.024 | 0.029 | 10.27 | 0.00* |
| Interest in math → Math self-efficacy | 0.09 | Math self-efficacy → Math scores | 0.63 | 0.057 | 0.047 | 0.024 | 0.030 | 1.91 | 0.06 |
| Parental education level → Math self-efficacy | 0.13 | Math self-efficacy → Math scores | 0.63 | 0.082 | 0.017 | 0.024 | 0.011 | 7.34 | 0.00* |
| Parental education level → Home education resources | 0.28 | Home education resources → Math scores | 0.17 | 0.048 | 0.018 | 0.014 | 0.005 | 9.57 | 0.00* |

4.7 Discussions

After testing the models rigorously, we can now answer those research questions with certain level of confidence. These equations as LISREL 8.5 outputs calculate the total effect, i.e., the summation of all direct and indirect effects from the independent variables to the dependent variable. To restate, β represents how many standard deviations increase in dependent variable for each 1 standard deviation increase in independent variable. It is significant if the t-value is above 1.96 at 95% level of confidence.

- 1) *Does parental education level affect student's math scores through the availability of home education resources and/or math self-efficacy?*

Firstly, the latent exogenous variable, parental education level, positively and significantly affects latent endogenous variables, home educational resources ($\beta = 0.28, t = 15.72$); and math self-efficacy ($\beta = 0.13, t = 7.28$). This finding does not only support hypotheses H12 and H14, but it also explicates the interplay between psychological factors and SES (Dika & Singh, 2002). Only students' math self-efficacy, but not math self-concept and interest in math, has a significant relationship with parental education level.

This finding is consistent with prior studies, e.g. in information technology. Hoyles, Healy and Pozzi (1994) believed that the availability of computer educational resources can increase social interactions. Powerful interactions between students and involvement in computer-based tasks can lead to higher-order thinking, hypothesis formation and reflection. Light (1993) concluded that the computers can increase the value added of group tasks and collaborations which subsequently improve the frequencies of learners' dialogues, interchange of ideas and negotiation of solutions. Isolated learners can share their computer resources and connections through networking in order to enhance learning. Oliver, Omari and Knibb (1997) stated that the World Wide Web has powerful communication components which connect

individual learners to virtual communities and those who share a common learning goal. Besides enabling communications via the process of learning, the virtual environment enables the learners to share the resources and materials, the products of which are the outcomes of learning.

In sum, not only the relationship between parental education level and home educational resources provided by parents (H12), and the relationship between parental education level and students' math self-efficacy (H14), are found to be positive and significant, but also the relationship between home educational resources provided by parents and the students' math scores (H13), and the relationship between students' math self-efficacy and students' math scores (H11), are found to be positive and significant. Parental education level was found to be affecting student's math scores through the availability of home educational resources and students' math self-efficacy.

Parents positively affect their children's math performances by transferring their own cultural and economic capital to their offspring (Bourdieu, 1986). Those whose families have more cultural possessions and educational resources such as a dictionary, a desk for study, textbooks, calculators, and a quiet place to study tend to perform better in math. Secondary school students with more computer facilities such as educational software at home perform better in math than others (OECD, 2003). PISA (2003) reported that students, who communicate with their parents more often about their daily topics, dine together, and discuss more about schoolwork and school life, tend to perform better in schools. This finding supports the view that equality of education regardless of student's family background is unrealistic.

Although this dissertation finds that parental education level has a significant and positive relationship with home educational resources, which, in turn, affects math scores, it does not confirm with the view that parental education level is the only factor. Teachers are found to play significant roles in the Hong Kong context as well.

2) *Does teacher emotional support affect student's math scores through cooperative learning environment and/ or math self-concept?*

Firstly, teacher emotional support has a positive relationship with cooperative learning ($\beta = 0.32$, $t = 11.74$). This is supported by prior studies. Teachers help develop a cooperative learning environment (Johnson & Johnson, 2004; Moriarty et al., 1995). They can activate students' curiosity and intellect through discussion, help them learn from other perspectives, and encourage them to give and receive support in small groups. However, teacher emotional support has no significant relationship with math self-concept ($\beta = 0.03$, $t = 1.29$).

Subsequently, cooperative learning directly affects interest in math, math self-concept and math self-efficacy. Their factor loadings are $\beta = 0.28$, $t = 13.62$; $\beta = 0.39$, $t = 13.77$; and $\beta = 0.10$, $t = 3.47$ respectively. Student's math self-concept, in turn, affects interest in math and math self-efficacy ($\beta = 0.70$, $t = 34.14$ and $\beta = 0.48$, $t = 11.06$ respectively). These findings are consistent with the point made by prior studies that self-concept has a significant association on interest in math (Sommer & Baumeister, 2002) and math self-efficacy (Pajares & Schunk, 2001).

Math self-efficacy and math self-concept have a direct effect on math scores ($\beta = 0.63$, $t = 26.19$; and $\beta = 0.08$, $t = 2.36$) which is consistent with prior studies (Williams, Freeman & Deci, 1998). However, interest in math was found to have a negative relationship with math scores ($\beta = -0.07$, $t = -2.34$) which is inconsistent with prior studies (Ryan & Deci, 2000). One possible explanation was that since interest in math has a very high correlation with math self-concept, suppression effect occurs and the sign was changed from positive to negative.

In cooperative learning situations students can perform 20 percentile better than competitive and individualistic learning students (Johnson & Johnson, 2004). Leiken and Zaslavsky (1999) reported that if the teachers use cooperative learning activities, students' belief toward math will be increased. Peer support can reduce isolation and

enhance students' own interest in learning. Therefore, all the other hypotheses were supported except H1 and H9. Teacher emotional support was found to affect student's math scores through fostering a cooperative learning environment. The finding is consistent with prior studies. Students with supportive relationship with their teachers are motivated to participate actively and appropriately in the classroom (e.g. Birch & Ladd, 1998). They will work harder, accept teacher direction and criticism, cope better with stress, and attend more to the teacher (e.g. Hughes & Kwok, 2007). This study contributes to fill in this gap based on a large sample size and rigorous statistical analysis with proven reliability and validity. The finding supports the view that parents and teachers significantly influence their children's academic achievements through psychological and affective mediators.

3) *Does the effect of the support from parents and teachers on students' math scores vary by gender?*

Based on the measurement invariance analyses, boys and girls received similar levels of support from their parents in Hong Kong. Nowadays, parents neither provide more support to boys due to a male-dominated culture, nor provide more support to girls due to the fact that the two are relationally oriented already. It is consistent with prior studies that there is no significant difference in math performance between the genders in recent years (Davis-Kean, Eccles & Linver, 2002), and differences favoring males in mathematics achievement are declining or disappearing (Marsh & Yeung, 1997).

The finding is consistent with Younger, Warrington and Williams (1999) study that many boys and girls perceived that teachers provide support to girls more than boys in the U.K. Birch and Ladd (1998) in the U.S. and Zhang and Sternberg (1998) in Hong Kong also found similar results. Younger, Warrington and Williams (1999) further suggested four reasons: (1) girls took initiative to ask if necessary and more frequently; (2) teachers perceived girls to be more self-disciplined and less disruptive; (3) girls are underrated as weaker than boys due to gender-bias; and (4) girls maintain a closer and less conflictive relationship with teachers.

This study supports the arguments that teachers significantly and positively influence their students' academic achievements through psychological and affective mediators. The SDT is applicable in supporting the claim that teachers can make a difference by providing a cooperative learning environment; but invalid in hypothesizing how it works through the mediators. This study provides an insight that with the same level of perceived support from parents and teachers, gender moderates the impact of support on adolescents' math performances.

This study supports the claim that cooperative learning among peers significantly and positively influences adolescents' math achievements through enhancing their interests in math, math self-concept, and math self-efficacy. Both genders benefit from cooperative learning environment but healthy peer cooperative learning relationships seem to be more essential for boys' self-concept than that of girls. It is consistent with prior evidence.

Prior evidence suggests that negative schooling experiences induce negative psychological feelings in adolescents like guilt, shame, anxiety and failure in Chinese culture (Cheung, 1997). Instead of seeking assistance from their parents and teachers, these students at risk may seek emotional comfort from their deviant peers who are facing the same predicament (Chen, 2004). Many researchers found that peer pressure has a differential and positive impact on adolescents' behavioral and academic outcomes. For instance, Fuligni et al. (2001) found that boys turned to their peers at the expense of academic achievement. However, girls tended to get support in return for positive influences more than boys. Ma et al. (2000) reported that boys suffered negative peer influence more than girls do, whereas girls gained positive peer influence more than boys. However, little research has examined how the varying types of peer support may contribute to differences in academic achievement under different cooperative learning contexts.

This study provides an insight that both genders have similar levels of interest in math, math self-concept, math self-efficacy and math scores according to the analysis of

their latent means. This study corroborates that neither girls perform better in math due to being more self-disciplined and putting in more effort, nor do boys perform better in math due to the subject's perceived masculine nature and more interest being aroused.

Prior research found that girls study harder, are more conscientious and more disciplined in general. They put more time and effort on completing their schoolwork. Conversely, boys are less intrinsically motivated to learn and they are less focused on their schoolwork in the U.K. (Younger, Warrington & Williams, 1999). Fuligni and Stevenson (1995) also found that boys spend more effort in extra-curricular activities and less effort on homework in the U.S., Taiwan and Japan. Rao, Moely and Sachs (2000) found that effort is positively associated with student academic achievements. However, Stipek and Gralinski (1991) reported that boys perform better in masculine subjects like math than girls do. Little is known about whether stereotype-bias may arouse interest in math and effort differently for boys and girls.

Therefore, previous research which reported that boys and girls perceive different levels of support from their parents and teachers are not supported by this study. This section contributes to the existing theories about (1) whether parents and teachers exert different levels of support to their boys and girls and (2) whether the same level of support exerts different impacts on math achievement for boys and girls and how it happens.

This study contributes to the understanding that; (1) a cooperative learning environment improves boys' math performances more than that of girls. It is found to affect math self-efficacy and interest in math in different magnitudes across gender. Math self-efficacy, parental education level and home educational resources play more important roles than self-concept and interest in math do on predicting adolescents' math scores in Hong Kong. Even though the SDT is not supported as a whole in this Hong Kong sample, the findings still support the claim that teacher emotional support can make a difference. This is consistent with prior study.

Deci and Ryan (1985) hypothesis that autonomy-supportive climates would enhance interest in math is supported. The relationship between interest in math being positively associated with self-concept is also supported (Williams, Freeman & Deci, 1998) but there is not a significant relationship with math scores when interest in math and math self-concept are juxtaposed with math self-efficacy. The claim by the SDT that students who are interested in math will value the experience, put in more effort, enhance math self-efficacy and perform better, is not supported. This study suggested that math self-concept acted as a suppressor variable in the statistical model (Ratelle et al., 2005).

Vallerand and Losier (1999:145) suggested a motivational sequence: 'social factors → psychological mediators → types of motivation → behavioral consequences'. This motivational sequence suggests that social factors influence individuals' perceptions of psychological mediators, which in turn determine their interest or motivation. Therefore, this sequence is also not supported. This study provides an insight that the relationship between the precursor, psychological mediators, and types of motivation is moderated by genders though it is not significant.

The above discussion sheds light on both the effectiveness and equality of the education system of Hong Kong. In the next section, the author will discuss the research findings in relations to the broader educational and social-cultural context in Hong Kong. Considering the fact that 16% of the annual total government expenditure is invested in education, effectiveness and equality are the two important issues on top of policy makers' agenda. Equality of educational opportunity means having the same chance of receiving a good education regardless of students' backgrounds (OECD, 2003). Effectiveness is defined as the extent to which the education system enhances students' academic achievements.

4) *What are the relative impacts of parental education level and teacher emotional support on student's math scores?*

One of the goals of education systems is to provide students with high quality and equal access to education regardless of their socio-economic status. Concerning the equality of education in Hong Kong, the differences between the high (95th percentile) and low (5th percentile) achievers in all domains are smaller than those of the OECD averages. 30.7% of Hong Kong students achieve high proficiency levels (OECD, 2003). In comparison to students in other areas who share similar SES backgrounds, Hong Kong students' scores are higher. This finding implies that Hong Kong students, no matter what their backgrounds and genders are, basically receive a similar standard of education.

This study finds that parents have a larger indirect effect (0.12) upon students' math performance than teacher emotional support effects does (0.08) in the Hong Kong context. The ratio of teacher emotional support effects to parental education effect on student academic performance is 0.66. This is consistent with prior research. Nonoyama (2005:183) used the Sheaf coefficients in determining the relative magnitudes of school factors versus family effects for PISA 2003 dataset and found that the ratios ranged from '0.18 (the United States) to 0.94 (Switzerland)'. These ratios showed that the effect of school is smaller than that of family in most of the countries. Parental education level still plays a significant and positive role in affecting adolescents' math scores in Hong Kong. Possible explanations are that the societal culture of Hong Kong is deeply affected by Confucian ethics, which emphasize obedience among five cardinal relationships (Wu Lun). Chinese parents are characterized as 'restrictive', 'controlling' and 'authoritarian' (Chao, 1994) but provided that their children are given close supervision and care, their academic achievements can be improved.

Nevertheless, the ratio for Hong Kong, 0.66, is above-average and implies Hong Kong is doing well in terms of providing equal access to education. The higher the teacher emotional support, the more capable a cooperative learning environment is in

enhancing math self-efficacy, which, in turn, increases students' math scores. Possible explanations are that teachers receive respect from parents and students. They have high status in the society since they serve as the role models and a knowledge dispenser. In addition, the government promises to ensure that more resources will be allocated to the education sector.

Undoubtedly, Hong Kong students have been performing consistently well in international assessments. Their scores on the four math sub-scales are 558, 558, 545, and 540 in 'space and shape', 'uncertainty', 'quantity' and 'change and relationship' respectively, which are significantly higher than the corresponding OECD averages of 496, 502, 501 and 499 (OECD, 2003). Statistically speaking, Hong Kong students' math mean scores are higher than those of other participant countries / cities except for Finland, Japan, Korea, the Netherlands and Liechtenstein. The same happens similarly for other literacy domains. According to Bronfenbrenner's ecological model, learning outcomes are affected by the micro-, meso- and macro-systems. This study supports the model that high quality education in Hong Kong is achieved by home educational resources support (micro-system); cooperative learning and a supportive teaching environment and close student-teacher relationship (meso-system). The impact on academic achievement of the government's promise to allocate more resources to education (macro-system) is left for future study.

Wentzel's (1998) propositions that students' learning processes are affected by their parents and teachers are supported. The effectiveness of education in Hong Kong in terms of international assessment scores is good but this does not answer the questions as to what factors contribute to the effectiveness. This study finds that teacher emotional support is important in establishing a cooperative learning environment so as to improve the learning effectiveness. It is consistent with prior findings (Berg, 1999; Brown & Palincsar, 1989; Cameron & Ettington, 1988; Ferguson, 1991; Ryan & Deci, 2000). Cooperative learning, in turn, has positive and significant impacts on math self-efficacy, math self-concept and interest in math, which is consistent with prior findings (Brancov, 1994; House, 2003; Moriarty et al.,

1995). The claim that teachers can arouse students' motivations is supported (Schiefele & Csikszentmihalyi, 1995).

This study contributes to previous literature that the associations between math self-concept and cooperative learning, interest in math, and math performance are moderated by gender. Prior research found conflicting conclusions because they usually overlook the gender differences. Therefore, prevention and intervention programs are designed to enhance students' math performance are more effective if they consider gender disparity which affects academic needs. For instance, programs aimed at enhancing math self-efficacy can directly improve boys' and girls' math performance though at different magnitudes. Programs aimed at enhancing math self-concept can directly improve girls' interest in math and math performance more than that of boys.

Chapter 5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The measurement model achieved measurement invariance across gender in Hong Kong. Males and females perceived items similarly and the model was cross validated with the whole PISA dataset to confirm it was not resulted from capitalization on chance.

The structural model has used PISA 2003 data and juxtaposed two social factors: parental education level and teacher emotional support. The model is guided by the SDT and finds that they fit well into this measurement model and they explain altogether 45% of the total variance in math scores in Hong Kong. When examining the overall parental effect, which refers to the total inequality in educational outcome as attributed to family background within a country, the result shows that the total effect from parents upon math scores is significant. Their role in affecting student math scores is partially mediated by home educational resources (including electronic resources).

When examining the overall teachers' effect, the model concludes that teachers do make a difference on students' math scores in Hong Kong. The magnitudes and directions of the model parameters are mostly in line with expectations of SDT. However, the SDT can only partially explain why teachers can make a difference on affecting students' math performances. This dissertation contributes to fuller understanding of their interrelationships. Math self-efficacy plays a more important role than math self-concept and interest in math do in Hong Kong in affecting math scores. This supports Bandura's but contradicts Marsh's findings.

The above-mentioned findings are mostly ignored in the PISA reports. In the official reports there is a brief comment about the social factors but PISA does not provide

answers as to which policies would lead to success in any given country. ‘They have already posed some very interesting questions which, in their present form, they themselves do not seem to be capable of answering’ (Smithers, 2004:35). This paper fills in this gap by focusing attention on math self-efficacy in Hong Kong.

It is only by understanding whether parents and teachers exert differential levels of support to boys and girls and whether the support has differential impacts on math achievement to boys and girls, that educators will better utilize the support to effectively help improve the math performance of both genders. The findings of this study provide clear insight for parental and teacher support. This study finds that both parental and teacher support are a powerful predictor of math performance for both genders. The measurement invariance analyses find that girls in Hong Kong perceive that they have: (1) less interest in math; (2) less math self-efficacy; (3) less math self-concept; (4) less cooperative learning environment; (5) less math scores; and (6) their parents are at lower education level than those of the boys. On the other hand, girls perceive that they get higher teacher emotional support than boys do. However, this gender disparity perception was found insignificant.

From the macro perspective, Hong Kong’s education system is concluded to have achieved both the effectiveness and equality goals. Hong Kong students have been performing significantly better than most of the other countries. From the micro perspective, parents exert significant impacts on students’ math performance but have been minimized already because the teacher to family effect ratio is not too low in comparison to those of other countries (Nonoyama, 2005; OECD, 2003). Parents invest more and are involved more in their children’s study nowadays. Empirical findings support the view that students perform better if they perceive their parents as supportive. From the meso perspective, Hong Kong math teachers are doing well in increasing students’ math self-efficacy via cooperative learning. Equality and effectiveness are not mutually exclusive in the Hong Kong context.

5.2 Recommendations

Parents still have a significant and positive direct effect on their children's math self-efficacy and home education resources. Suitable training programs and workshops on improving the quality of parent-child communication should be provided to the parents for the sake of improving their communication skills, attitudes towards their children and selecting suitable home educational resources. In line with the findings, parents are recommended to improve their education levels through life-long learning so as to enhance their children's math self-efficacy. Although this study supports the significance of parental involvement, the author is also aware that those parents who involve themselves overwhelmingly may suffocate their children's personal development and debilitate teachers' pedagogical practices.

The first recommendation is that educational resources must be given to students either by families, or if not possible, by schools or public libraries instead, to reduce the achievement gap between students of the 'haves' and the 'have-nots'. In line with this logic, schools with low SES students need more resources than those schools with high SES students. Policy makers are suggested to focus on adequacy rather than equity to allocate extra funds that are adequate enough for those band-two or three schools to attract and keep qualified teaching staff to level the students' playing field.

Secondly, although Hong Kong's education system is performing well, competition and pressure are keen due to the examination-oriented and meritocratic design. Diverse teaching and learning pedagogical training are recommended in order to create a cooperative learning environment so as to cater for the diverse needs of students. An autonomy supportive teaching style is found to be teachable (Reeve, 1998). This recommendation answers the call for more teacher preparation and teacher professional development in building supportive and warm relationships with students (Hughes & Kwok, 2007).

This study provides an insight that cooperative learning can enhance affective and cognitive factors. This helps building healthy relationships with peers at school rather than making friends with deviant gangs outside and this is particularly essential for boys. The peer-peer and teacher-student relationships should be healthier and more constructive. The fear of school closure, together with the stipulation of mother-tongue education, cause fierce competition in student recruitment among schools. According to the Hong Kong Professional Teachers' Union, teachers' burnout, stress and anxiety are related to the policy of downsizing and school closures, putting schools and teachers under surveillance, introducing external and internal evaluations, and refusal for the implementation of small-class teaching (PTU News, 2006). Based on these findings, schools and teachers can benefit by convincing the EB to provide more support and to impose less competitive assessments on the schools and teachers in order to reduce their burden and stress. Not only can these benefits be extended to the parents, teachers and policy makers, school administrators can also benefit from spending more wisely by ensuring that educational policies, methodologies and activities would enhance students' math self-efficacy if score is the assessment criterion.

Thirdly, based on the lower latent mean of math self-concept for females, to infer a bit further, teachers are recommended to enhance females' math self-concept which is found to be closely related to interest in math. This study corroborates the need for: (1) increasing females' awareness of the importance of math; (2) increasing their awareness of the careers in math; (3) relating the female students to significant women figures working in those careers; (4) developing cooperative learning groups in order to encourage their interest; and (5) teaching math from a feminist perspective so as to explore more humanized aspects of the subject rather than the purely algorithmic procedures (Hoover-Dempsey et al., 2005).

Fourthly, math self-efficacy is found to be an important mediator in affecting math performance. Parents, school administrators, teachers and policy makers should ensure that educational policies, methodologies and activities would enhance

students' confidence in math. Students should be exposed to a high-level of mathematics to enhance their efficacy with a high degree of autonomy.

Finally, the current high demands for the proficiency in basic knowledge and higher-order mathematical skills have reduced students' interest in math and math self-concept. Pong and Chow (2002) reported that academic pressure is one of the major causes of adolescent suicide in Hong Kong. Academic success contributes to a better life in the future but at the expense of present well-being for some students. All stakeholders are therefore recommended to think together whether it is time to sacrifice temporary high ranking for long-term enthusiasm in math. There are two reasons: (1) this can nurture their other capabilities like ethical, physical, social and aesthetical as well as intellect to become well-rounded individuals; and (2) this can prevent academic pressure from becoming a source of destruction for adolescents. Outcome and process are equally important. Students should be able to learn efficiently, effectively and happily. Prior research supports this reasoning. 'A tradeoff between excellences in achievement rankings by clinging to traditional pedagogy and a more balanced development in students' competencies towards the desirable goals of education may be inevitable' (OECD, 2003). Policy makers are advised to provide more support and less assessment to the schools and teachers to reduce their burden and stress.

5.3 Future Research

It is imperative to acknowledge opportunities for future research. The author used an existing dataset because one of the reasons mentioned was to improve the instrument in return. One construct, parental education level has only two items. Future research should refine it and consider adding new indicators, e.g. elder brother's / sister's education level, that fully fit in the construct. Another construct, cooperative learning, was unequivocal in terms of its Cronbach alpha, 0.70 and requires more and better indicators to operationalize it in future studies.

References

- Ahire, S. L., Golhar, D. Y., & Waller, M. A. (1996). Development and validation of TQM implementation constructs. *Decision Science*, 27 (1), 23–56.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two step approach. *Psychology Bulletin*, 103, 411–423.
- Anderson, R., Manoogian, S. T., & Reznick, J. S. (1976). The undermining and enhancing of intrinsic motivation in preschool children. *Journal of Personality and Social Psychology*, 34(5), 915–922.
- Bagozzi, R. P. (1980). *Causal models in marketing*. Wiley, NY.
- Baker, D., Goesling, B., & Letendre, G. (2002). Socioeconomic status, school quality, and national economic development: A cross-national analysis of the ‘Heyneman-Loxley Effects’ on mathematics and science achievement. *Comparative Education Review*, 46(3), 291-312.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1999). Self-efficacy Pathways to Childhood Depression. *Journal of Personality and Social Psychology*, 76(2), 258-269.
- Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*, 41, 586-598.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117, 497-529.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). 'Ego Depletion: Is the Active Self a Limited Resource?' *Journal of Personality and Social Psychology*, 74 (5), 1252-65.

Bayless, D. L., (1985). Sampling in Education and the Social Sciences. *Journal of Educational Measurement*, 22 (4), 318-321.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.

Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods and Research*, 16, 78-117.

Berg, G. A. (1999). Community in distance learning through virtual teams. *Educational Technology Review*, 12, 23-29.

Biggs, J. B., & Watkins, D. A. (2001). Insights into teaching the Chinese learner. In D.A. Watkins & J.B. Biggs (Eds.). *Teaching the Chinese learner: Psychological and pedagogical perspectives* (pp.277-300). Hong Kong: Comparative Education Research Centre, The University of Hong Kong.

Birch, S. H., & Ladd, G. W. (1998). Children's interpersonal behaviors and the teacher-child relationship. *Developmental Psychology*, 34, 934 - 946.

- Blasko, Z. (2003). Cultural reproduction or cultural mobility? *Review of Sociology*, 9(1), 5-26.
- Blatchford, P., Baines, E., Rubie-Davies, C., Basset, P., & Chowne, A. (2006). The effect of a new approach to group work on pupil-pupil and teacher-pupil interactions. *Journal of Educational Psychology*, 98 (4), 750-765.
- Blau, P., & Duncan, O. (1967). *The American Occupational Structure*. New York: Wiley.
- Bloom, B. S. (1972). Innocence in education. *School Review*, 80, 333-352.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, 31, 445-457.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K., Glanville, J. A., & Stecklov, G. (2001). Socioeconomic status and class in studies of fertility and health in developing countries. *Annual Review of Sociology*, 27, 153-185.
- Bourdieu, P. (1986). Forms of capital. In J. G. Richardson (ed.). *Handbook of Theory and Research in the Sociology of Education*, 242-258.
- Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Educational reform and contradictions of economic life*. New York: Basic Books.
- Brancoy, T. (1994). *Cooperative learning in mathematics with middle school Indian students: A focus on achievement and on-task behaviour*. University of South Dakota, AAT 9506443.

Bronfenbrenner, U. (1979). *The Ecology of Human Development: Experiments by Nature and Design*. Cambridge, MA: Harvard University Press.

Brophy, J. E. (1979). Teacher Behaviour and Its Effects. *Journal of Educational Leadership*, 71, 733-750.

Brown, A. L., & Palincsar, A. S. (1989). Guided, cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*, 393-451. Hillsdale, NJ: Erlbaum.

Browne, M. W. (1984). Asymptotically distribution-free methods for analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models*, 136-162. Newbury Park, CA: Sage.

Budget Speech (2006). (ON-LINE <http://www.budget.gov.hk/2006/eng/speech.htm> March 2007).

Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis: Elements of the sociology of corporate life*, London, Heinemann.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: SpringerVerlag.

Byrne, B. M., Shavelson R. J., & Muthen, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance. *Psychological Bulletin*, 105, 456-466.

Cameron, K. S., & Ettington, D. R. (1988). The conceptual foundations of organizational culture. In Smart, J.C. (eds). *Higher education: Handbook of theory and research*, 4, 356-396. New York: Agathon.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G.W. Nohrnstedt & E.F. Borgatta (Eds.), *Social measurement: current issues*. Beverly Hills, CA: Sage, 63-115.

Carr, A. S., & Pearson, J. N. (1999). Strategically managed buyer-supplier relationships and performance outcomes. *Journal of Operations Management*, 17, 497-519.

Carter, R. S., & Wojtkiewicz, R. A. (2000). Parental involvement with adolescents' education: Do Daughters or sons get more help? *Adolescence* 35(137), 29-44.

Chan, Y. C., & Cheng, Y. C. (1993). A study of principal's instructional leadership in Hong Kong secondary schools. *Educational Research Journal*, 8, 55-67.

Chao, R. K. (1994). Beyond parental control and authoritarian parenting style: Understanding Chinese parenting through the cultural notion of training. *Child Development*, 65, 1111-1119.

Chau, P. Y. K. (1997). Reexamining a model for evaluating information center success using a structural equation modeling approach. *Decision Science*, 28 (2), 309-334.

Chen, J. J. L. (2004). *Academic support from parents, teachers, and peers: relation to Hong Kong adolescents' academic behavior and achievement*. Thesis (Ed. D.), Harvard Graduate School of Education, Ann Arbor: University Microfilms International.

Cheng, Y. C. (1996). *The pursuit of school effectiveness: Research, management and policy*. Hong Kong: Hong Kong Institute of Educational Research.

Cheng, Y. C., Cheung, W. M., & Tam, W. M. (2002). The Pacific Rim and Australia – Hong Kong. In D. Reynolds, B. Creemers, S. Stringfield, C. Teddlie & E. Schaffer. *World class schools: International perspectives on school effectiveness*. 138-155. London: Routledge/Falmer.

Cheung, Y. W. (1997). Family, school, peer and media predictors of adolescent deviant behavior in Hong Kong. *Journal of Youth and Adolescence*, 26(5), 569-596.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187-212.

Chong, W. (2006). *Teachers' fears eased on closures*. The standard, Metro. H.K.

Churchill, G. A. (1987). *Marketing Research: Methodological Foundations*, 4th ed. Chicago, IL: Dryden.

Close, M. (2001). *Integrating Self-determination Theory and Social Cognitive Theory to Predict Urban High School Students' Distress, Achievement, and Retention*. The University of Wisconsin- Milwaukee.

CME (2004a). *Measuring Up: Canadian Results of the OECD PISA Study: The Performance of Canada's Youth in Mathematics, Reading, Science and Problem Solving*. Canada and Statistics Canada, 2. Council of Ministers and Education.

CME (2004b). *Aspiration of Canadian youth for higher education*. Council of Ministers of Education. (ON-LINE

<http://www.hrsdc.gc.ca/asp/gateway.asp?hr=/en/cs/sp/hrsdclp/publications/2004-002631/page00.shtml&hs=1xa> February 2007).

Cohen, D. A., & Rice, J. (1997). Parenting Styles, Adolescent Substance Use and Academic Achievement. *Journal of Drug Education*, 27, 199-211.

Cohen, L., & Manion, L. (1994). *Research methods in education*, 4th edn. Routledge, London.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L., (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, 95-120.

Converse, J., & Presser, S. (1986). *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, CA: Sage.

Cornelius-White, J. (2007). Learner-Centered Teacher-Student Relationships Are Effective: A Meta-Analysis. *Review of Educational Research*, 77(1), 113 - 143.

Covington, M. V. (2001). Goal theory, motivation, and school achievement: An integrative review. *Annual Review of Psychology*, 51, 171-200.

Creemers, B. M., Scheerens, J., & Reynolds, D. (2000). Theory development in school effectiveness research. In C. Teddlie & D. Reynolds (Eds.). *The international handbook of school effectiveness research*, 283-298. London: Falmer press.

Creswell, J. W. (2002). *Educational Research – Planning, Conducting, and Evaluating Quantitative and Qualitative research*. Merrill Prentice Hall. Upper Saddle River, New Jersey.

- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 343-360.
- Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, 105, 317-327.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavior Research*, 18, 147-167.
- Curran, P. J., West, S. G., & Finch, J. (1996). The robustness of test statistics to non-normality and specification error in confirmatory factor analysis. *Psychological Methods*, 1, 16-29.
- Dahl, G. B., & Moretti, E. (2004). *The Demand for Sons: Evidence from Divorce, Fertility, and Shotgun Marriage*. National Bureau of Economic Research, Inc.
- Darling-Hammond, L. (1997). *Doing What Matters Most: Investing in Quality teaching*. National Commission on Teaching and American Future.
- Davis-Kean, P., Eccles, J., & Linver, M. (2002). *Influences of Gender on Academic Achievement*. Biennial meeting of the Society for Research on Adolescence.
- Davis-Kean, P., & Sandler, H. M. (2001). A meta-analysis of measures of self-esteem for younger children: A framework for future measures. *Child Development*, 72, 887-906.
- DeCharms, R. (1968). *Personal causation: The internal affective determinants of behavior*. New York: Academic Press.
- Deci, E. L., & Ryan, R. M. (1985). The General Causality Orientations Scale: Self-determination in personality. *Journal of Research in Personality*, 19, 109-134.

Deci, E. L., & Ryan, R. M. (1991). A motivational approach to self: Integration in personality. In R. Dienstbier (Ed.), *Nebraska Symposium on Motivation: Perspectives on motivation*. Lincoln: University of Nebraska Press, 38, 237-288.

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 227-268.

Denzin, N. (1970). Strategies of multiple triangulations. In Denzin, N. (ed.) *The research act in sociology: a theoretical introduction to sociological method*. New York, McGraw-Hill, 297-313.

Dika, S. L., & Singh, K. (2002). Applications of Social Capital in Educational Literature: A Critical Synthesis. *Review of Educational Research*, 72(1), 31-60.

Doest, L. T., & Jonge, J. D. (2006). Testing causal models of job characteristics and employee well-being: A replication study using cross-lagged structural equation modeling. *Journal of Occupational and Organizational Psychology*, 79, 499-507.

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, 46(2), 183-201.

Eccles, J. S., Wigfield, A., Midgley, C., Reuman, D., MacIver, D., & Feldlaufer, H. (1993). Negative effects of traditional middle schools on students' motivation. *Elementary School Journal*, 93(5), 553-574.

Education Department (2001). *Quality assurance inspection annual report 2000/2001*. Hong Kong: Printing Department.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review*, 59(6), 822-838.

Ferguson, R. F. (1991). Paying for Public Education: New Evidence on How and Why Money Matters. *Harvard Journal on Legislation*, 28(2), 465–498.

Ferguson, R. F., & Brown, J. (2000). Certification test scores, teacher quality, and student achievement. In D. W. Grissmer & J. M. Ross (Eds.), *Analytic issues in the assessment of student achievement*. Washington, DC: National Center for Education Statistics

Ferguson, R. F., & Ladd, H. F. (1996). How and Why Money Matters: An Analysis of Alabama Schools. In H. F. Ladd (ed.), *Holding Schools Accountable: Performance Based Reform in Education*. Washington, DC: Brookings Institution Press.

Filmer, D., & Pritchett, L. (1999). The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries. *Population and Development Review*, 25(1), 85-120.

Fisher, C. W., Berliner, C. D., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1981). Teaching Behaviors, Academic Learning Time, and Student Achievement: An Overview. *Journal of Classroom Interaction*, 17, 2-15.

Fitzgerald, B., & Howcroft, D. (1998). Competing dichotomies in IS research and possible strategies for resolution. In R. Hirschheim., M. Newman & J. DeGross (Eds) *Proceedings of the 19th International Conference in Information Systems*, Helsinki, December.

Flink, C., Boggiano, A. K., & Barrett, M. (1990). Controlling teaching strategies: Undermining children's self-determination and performance. *Journal of Personality and Social Psychology*, 59, 916-924.

Fornell, C. (1982). *A second generation of multivariate analysis*. New York: Praeger.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18 (1), 39–50.

Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage Publications.

Freese, J., & Powell, B. (1999). Sociobiology, Status, and Parental Investment in Sons and Daughters: Testing the Trivers-Willard Hypothesis. *American Journal of Sociology*, 106, 1704-43.

Fuchs, L., & Woessmann, L. (2004). *What Accounts for International Differences in Student Performance? A Re-examination using PISA Data*. Working paper no. 1235, Labour Markets.

Fuligni, A. J., Eccles, J. S., Barber, B. L., & Clements, P. (2001). Early adolescent peer orientation and adjustment during high school. *Development Psychology*, 37(1), 28-36.

Fuligni, A. J., & Stevenson, H. W. (1995). Time use and mathematics achievement among American, Chinese, and Japanese high school students. *Child Development*, 66(3), 830-842.

Furman, W., & Buhrmester, D. (1992). Age and sex difference in perceptions of networks of personal relationships. *Child development*, 163, 103 - 115.

Gager, C. T., & Sanchez, L. (2004). *Whose time is it? The effect of gender, employment, and work/ family stress on children's housework*. Presented at annual meeting American Sociology Association. San Francisco.

- Gal, I., & Garfield, J. B. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. B. Garfield, *The assessment challenges in statistics education*, eds. Amsterdam: IOS, 1-13.
- Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, 25, 201–239.
- Garson, D. (2007). *Inter-rater reliability: Estimation based on the correlation of scores between/among two or more raters who rate the same item, scale, or instrument*. (ON-LINE <http://www2.chass.ncsu.edu/garson/pa765/reliab.htm> May 2007).
- Gilligan, C. (1982). *In a different voice*. Cambridge, MA: Harvard University Press.
- Goldhaber, D., & Brewer, D. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational Evaluation and Policy Analysis*, 22(2), 129-145.
- Good, T. L., & Grouws, D. A. (1979). The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth-Grade Classrooms. *Journal of Educational Psychology*, 71, 355-362.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The Effect of School Resources on Student Achievement. *Review of Educational Research*, 66, 361–396.
- Grolnick, W. S., Ryan, R. M., & Deci, E. L. (1991). The inner resources for school performance: Motivational mediators of children's perceptions of their parents. *Journal of Educational Psychology*, 83, 508-517.
- Gronlund, N. E. (1982). *Constructing Achievement Tests*. Englewood Cliffs, NJ: Prentice-Hall.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*, 124-136.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis with Readings*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall.

Hanushek, E. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24*(3), 1141-78.

Harries, M. M., & Schaubroeck, J. (1990). Confirmatory modeling in organizational behavior/human resource management: Issues and applications. *Journal of Management, 16* (2), 337-360.

Harter, S. (1978). Effect on motivation reconsidered: Toward a developmental model. *Human Development, 21*, 34-64.

Hatry, H. P. (1994). *Collecting data from agency records. In handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass Publishers.

Hauser, R. M., & Huang, M. H. (1997). Verbal ability and socioeconomic success: A Trend Analysis. *Social Science Research, 26*, 331-376.

Heyneman, S., & Loxley, W. (1983). The effect of primary-school quality on academic achievement across twenty-nine high-and low-income countries. *American Journal of Sociology, 88*(6), 1162-1194.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1*, 69-82.

Hill, N. E., & Craft, S. (2003). Parent-school involvement and children's school performance: Mediated pathways among African American and Euro-American children. *Journal of Educational Psychology, 95*, 74-83.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In Brooks-Gunn, J., & Petersen, A.C. (eds.) *Girls at puberty: Biological and psychosocial perspectives*. Plenum, New York.

HKCEE (2006). (ON-LINE
http://eant01.hkeaa.edu.hk/hkea/topper.asp?p_coverdown=what_home.html July 2007)

Ho, S. C., & Willms, J. D. (1996). Effects of Parental Involvement on Eighth-grade Achievement. *Sociology of Education, 69*, 126-141.

Hofstede, G. (1991). *Cultures and Organisations: Software of the Mind*. Harper Collins Publishers, London House, New York.

Hofstede, G. (2004).
 (ON-LINE http://www.geert-hofstede.com/hofstede_dimensions.php July 2007).

Horn, J. L. (1991). Comments on 'Issues in Factorial Invariance'. In Collins, L.M., & Horn, J.L., *Best methods for the Analysis of Change*. Washington, DC: American Psychological Association, 114-125.

Horn, J. L., & McArdle, J. (1992). A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Experimental Aging Research, 18*, 117-144.

Hoover-Dempsey, K. V., & Sandler, H. M. (2005). *The Social Context of Parental Involvement: A Path to Enhanced Achievement*. Presented to Project Monitor, Institute of Education Sciences, U.S. Department of Education.

Hoover-Dempsey, K. V., Walker, J. M. T., Sandler, H. M., Whetsel, D., Green, C. L., Wilkins, A. S., & Closson, K. E. (2005). Why do parents become involved? Research findings and implications. *Elementary School Journal*, 106(2), 105-130.

House, J. D. (2003). Self-beliefs and science and mathematics achievement of adolescent students in Hong Kong: Findings from the Third International Mathematics and Science Study (TIMSS). *International Journal of Instructional Media*, 30(2), 195 -212.

Hoyles, C., Healy, L., & Pozzi, S. (1994). Groupwork with computer: An overview of findings. *Journal of Computer Assisted Instruction*, 10, 202-215.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks, CA: Sage. 158-176.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R.H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications*. Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

Hubbard, J. H. (1999). The Forced Damped Pendulum: Chaos, Complication and Control. *American Mathematical Monthly*, 106 (8), 741-758.

Hughes, J. N., & Kwok, O. (2007). The influence of student-teacher and parent-teacher relationships on lower achieving readers' engagement and achievement in the primary grades. *Journal of Educational Psychology*, 99, 39-51.

Iyengar, S. S., & DeVoe, S. E. (2003). Rethinking the Value of Choice: Considering Cultural Mediators of Intrinsic Motivation. In Nebraska Symposium on Motivation: *Cross-Cultural Differences in Perspectives on the Self*. (49)5. Murphy-Berman and J. Berman, eds. London: University of Nebraska Press, 149–91.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the Value of Choice: A Cultural Perspective on Intrinsic Motivation. *Journal of Personality and Social Psychology*, 76 (3), 349–66.

Jambulingam, T., Kathuria, R., & Doucette, W. (2005). Entrepreneurial orientation as a basis for classification within a service industry: The case of retail pharmacy industry. *Journal of Operations Management*, 23 (1), 23-42.

Johnson, D. (1994). *Research Methods in Educational Management*. Harlow, Longman.

Johnson, D. W., & Johnson, R. T. (2004). *Using Cooperative Learning in Math. Teaching and Learning Middle Grade Mathematics*. Key College Publishing.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.

Jöreskog, K. G., & Sörbom, D. (1986a). *PRELIS – A program for multivariate Data Screening and Data Summarization. A Preprocessor for LISREL*. Mooresville, Indiana: Scientific Software, Inc.

Jöreskog, K. G., & Sörbom, D. (1986b). *LISREL 6: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Mooresville, Indiana: Scientific Software, Inc.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with SIMPLIS command language*. Chicago: Scientific Software, Inc.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8.5 Structural equation modeling with SIMPLIS command language*. Lawrence Erlbaum, Hillsdale, NJ.

Kaplan, D. (1995). Statistical power in structural equation modeling. In R.H. Hoyle (ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*, 100-117. Newbury Park, CA: Sage Publication, Inc.

Kelloway, E. K. (1998). *Using LISREL for Structural Equation Modeling, A Researcher's Guide*. Newbury Park, CA: Sage.

Kerlinger, F. (1986). *Foundations of Behavioral Research*. New York; Holt, Rinehart and Winston.

Kiamanesh, A. R. (2004). *Self-concept, home background, motivation, attitude, attribution, and their effects on Iranian students' science achievement*. Third International Biennial SELF Research Conference.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. N.Y.: Guildford Press.

Kohn, M. (1989). *Class and Conformity: A Study in Values; With a Reassessment, 1977*. Chicago: University of Chicago Press.

Koutsoulis, M. K., & Campbell, J. R. (2001). Family processes affect students' motivation, and science and math achievement in Cypriot high schools. *Structural Equation Modeling*, 8, 108-127.

- Kuhn, T. S. (1970). *The structure of scientific revolutions*, 2nd ed., Chicago, University of Chicago Press.
- Lee, J. C. K., Lam, W. P., & Li, Y. Y. (2003). Teacher evaluation and effectiveness in Hong Kong: Issues and challenges. *Journal of Personnel Evaluation in Education*, 17(1), 41-65.
- Lee, V. E., Marks, H. M., & Byrd, T. (1994). Sexism in single-sex and coeducational secondary school classrooms. *Sociology of Education*, 67, 92-120.
- Leiken, R., & Zaslavsky, O. (1999). Cooperative Learning in Mathematics. *Mathematics Teacher*, 92(3), 240-7.
- Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and Competence in German and American University Students. A Comparative study Based on Self-Determination Theory. *Journal of Educational Psychology*, 96(1), 68-84.
- Levin, T., & Long, R. (1981). *Effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Li, K. C. (2003). *SEM welcomes PISA findings* (ON-LINE <http://www.info.gov.hk/gia/general/200307/01/0630245.htm> May 2007).
- Light, P. (1993). *Collaborative Learning with Computers*. In *Language, Classrooms and Computers*. London, Routledge.
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.

Lo, N. K., Tsang, W. K., Chung, C. M., Chung Y. P., Sze, M. M., Ho, S. C., & Ho, M. K. (1997). *A survey of the effectiveness of Hong Kong secondary school system: Final report of Earmarked Grant Research Project*. Hong Kong: Hong Kong Institute of Educational Research. The Chinese University of Hong Kong.

Loehlin, J. C. (1998). *Latent Variable Models*. Lawrence Erlbaum, Hillsdale, New Jersey.

Lundberg, S. J., McLanahan, S., & Rose, E. (2007). Child Gender and Father Involvement in Fragile Families. *Demography*, 44(1), 79-92.

Ma, H. K., Shek, D. T. L., Cheung, P. C., & Lam, C. O. B. (2000). Parental, peer, and teacher influences on the social behavior of Hong Kong Chinese adolescents. *The Journal of Genetic Psychology*, 161(1), 65-78.

MacCallum, R. M. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107-120.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490-504.

Mael, F. A. (1998). Single-sex and coeducational schooling: relationships to socioemotional and academic development. *Review of Educational Research*, 68, 101-129.

Malhotra, N. K. (2004). *Marketing research: An applied orientation*. Prentice Hall, Englewood Cliffs, New Jersey.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, 79, 280-295.

Marsh, H. W. (1992). *Self-Description Questionnaire II: Manual*. New South Wales, Australia: University of Western Sydney, Macarthur, Faculty of Education, Publication Unit.

Marsh, H. W., & Craven, R. G. (2002). The Pivotal Role of Frames of Reference in Academic Self-concept Formation: The Big Fish Little Pond Effect. In F. Pajares & T. Urdan (Eds.), *Adolescent and Education*. (2) 83-123. Greenwich, CT. Information Age.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal Effects of Self-Concept and Performance from a Multidimensional Perspective- Beyond Seductive Pleasure and Unidimensional Perspectives. *Association for Psychological Science*, 1(2), 133-163.

Marsh, H. W., Hau, K. T., & Kong, K. W. (2002). Multilevel causal ordering of academic self-concept and achievement: Influence of language of instruction (English vs. Chinese) for Hong Kong students. *American Educational Research Journal*, 39, 727-763.

Marsh, H. W., Martin, A. J., & Hau, K. T. (2006). A multiple method perspective on self-concept research in educational psychology: A construct validity approach. In M. Eid & E. Diener (Eds.) *Handbook of multimethod measurement in psychology*, 441-456. Washington, DC: American psychology Association.

Marsh, H. W., & Peart, N. (1988). Competitive and cooperative physical fitness training programs for girls: Effects on physical fitness and on multidimensional self-concepts. *Journal of Sports and Exercise Psychology*, 10, 390-407.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic Self-Concept, Interest, Grades, and Standardized Test Scores: Reciprocal Effects Models of Causal Ordering. *Child Development*, 76(2), 397-416.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement -- structural equation models of longitudinal data. *Journal of Educational Psychology*, 89, 41-54.

Massen, G. H., & Bakker, A. B. (2001). Suppressor variables in path models. *Sociological Methods and Research*, 30, 241-270.

Mayer, S. (1997). *What Money Can't Buy: Family Income and Children's Life Chances*. Cambridge, MA: Harvard University Press.

Medley, D. M. (1977). *Teacher Competence and Teacher Effectiveness: A Review of Process-Product Research*. Washington, D.C.: American Association of Colleges for Teacher Education.

Medsker, G. J., Williams, L. J., & Holahan, P. J. (1994). A Review of Current Practices for Evaluating Causal Models in Organizational Behavior and Human Resources Management Research. *Journal of Management*, 20, 439-464.

Meehan, B. T., Hughes, J. N., & Cavell, T. A. (2003). Teacher-Student Relationships as Compensatory Resources for Aggressive Children. *Child Development*, 74 (4), 1145-1157.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525-543.

Miserandino, M. (1996). Children who do well in school: Individual differences in perceived competence and autonomy in above-average children. *Journal of Educational psychology*, 88, 203-214.

Monk, D., & King-Rice, J. (1994). Multi-level teacher resource effects on pupil performance in secondary mathematics and science: The role of teacher subject matter

preparation. In R. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29-58). Ithaca, NY: ILR Press. (ED 377 756)

Moriarty, B., Douglas, G., Punch, K., & Hattie, J. (1995). The importance of self-efficacy as a mediating variable between learning environments and achievement. *British Journal of Educational Psychology*, 65, 73-84.

Mueller, R. O. (1996). *Basic Principles of Structural Equation Modeling: An Introduction to LISREL and EQS*. New York: Springer.

Mujis, D., & Reynolds, D. (2001). *Effective Teaching: Evidence Based Practice*. Paul Chapman Publishing: London.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., Chrostowski, S. J., & Smith, T. A. (2000). *TIMSS 1999 International Mathematics Report: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. The International Study Center Boston College, Lynch School of Education.

Nash, R. (2005). Explanation and quantification in educational research: the arguments of critical and scientific realism. *British Educational Research Journal*, 31(2), 185-204.

Nasser, F. M. (2004). Structural Model of the Effects of Cognitive and Affective factors on the Achievement of Arabic-Speaking Pre-service Teachers in Introductory Statistics. *Journal of Statistics Education*, 12 (1), 1-24.

Natriello, G. (1987). The Impact of Evaluation Processes on Students. *Educational Psychologist*, 22, 155- 175.

Nix, G., Ryan, R. M., Manly, J. B., & Deci, E. L. (1999). Revitalization through self-regulation: The effects of autonomous and controlled motivation on happiness and vitality. *Journal of Experimental Social Psychology*, 35, 266-284.

Nonoyama, Y. (2005). *A Cross-National, Multi-Level Study of Family Background and School Effects on Educational Achievement*. Teachers College, Columbia University.

Norwich, B. (1987). Self-Efficacy and Maths Achievement: A Study of Their Relation. *Journal of Educational Psychology*, 79 (4), 384-387.

Nowell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960-1994: an analysis of differences in mean, variance, and extreme scores. *Sex Roles*, 39, 21-43.

Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill, New York.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. McGraw-Hill, New York.

OECD (2003). *The Second HKPISA Report (PISA 2003)*. Hong Kong: HKPISA Center.

OECD (2004). *Learning for Tomorrow's World – First Results from PISA 2003*. Hong Kong: HKPISA Center.

O'Leary-Kelly, S. W., & Vokurka, R. J. (1989). The empirical assessment of construct validity. *Journal of Operations Management*, 16, 387-405.

Oliver, R., Omari, A., & Knibb, K. (1997). Creating collaborative computer-based learning environments with the World Wide Web. In R. Kevill, R. Oliver & R.

Phillips, (Eds.). *What works and why. Proceedings of the 14th Annual Conference of the Australian Society for Computers in Tertiary Education*, 444-449. Perth, WA: Academic Computing Services.

Oppenheim, A. N. (1993). *Questionnaire design, interviewing and attitude measurement*. Pinter, London.

Papanastasiou, C. (2002). Effects of Background and School Factors on the Mathematics achievement. *Educational Research and Evaluation*, 8(1), 55-70.

Pajares, F. (1997). Gender differences in mathematics self-efficacy beliefs. In A. Gallagher & J. Kaufman (Eds.). *Mind the Gap: Gender differences in mathematics*. Boston, MA: Cambridge University Press.

Pajares, M., & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Maths Performance of Entering Middle School Students. *Contemporary Educational Psychology*, 24, 124-139.

Pajares, F., & Miller, M. D. (1994). Role of self-Efficacy and Self-Concept Beliefs in Mathematical Problem Solving. A Path Analysis. *Journal of Educational Psychology*, 86(2), 193-203.

Pajares, F., & Kranzler, J. (1995). Self-Efficacy and Self-Concept Beliefs in Mathematical problem solving: A Path Analysis. *Journal of Educational Psychology*, 86, 193-203.

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and School Success: Self-efficacy, Self-concept, and School Achievement. In R. Riding & S. Rayner (2001). *Perception*, 239-266. London: Ablex Publishing.

Pajares, F., & Schunk, D. H. (2002). Self and self-belief in psychology and education: A historical perspective. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education*, 3-21. San Diego, CA: Academic Press.

Pajares, F., & Valiante, G. (1999). Grade Level and Gender Differences in the Writing Self-Beliefs of Middle School Students. *Contemporary Educational Psychology*, 24, 390-405.

Pastorelli, C., Caprara, G. V., Barbaranelli, C., Rola, J., Rozsa, S., & Bandura, A. (2001). Structure of children's perceived self-efficacy: A Cross-national study. *European Journal of Psychological Assessment*, 17, 87-97.

Patrick, H., Ryan, A. M., & Kaplan, A. (2007). Early Adolescents' Perceptions of the Classroom Social Environment, Motivational Beliefs, and Engagement. *Journal of Educational Psychology*, 99(1), 83-98.

Peckham, P. D., & Roe, M. D. (1977). The Effects of Frequent Testing. *Journal of Research and Developmental Education*, 10, 40-50.

Phan, H., & Walker, R. (2000). *Maths Self-efficacy in Primary Schools: Evidence of a Hierarchical Structure*. Paper Presented at the Inaugural International Self-Concept Conference. Self Research Centre, October.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33-40.

PISA (2003). (ON-LINE

http://www.oecd.org/pages/0,2966,en_32252351_32236173_1_1_1_1_1,00.html

March 2007).

PISA Technical Report (2003). (ON-LINE

<https://www.pisa.oecd.org/dataoecd/49/60/35188570.pdf> Aug. 2007).

Policy Address by Chief Executive (2005). (ON-LINE

<http://www.info.gov.hk/gia/general/200510/12/P200510120104.htm> March 2007).

Pong, W. Y., & Chow, J. C. S. (2002). On the pedagogy of examinations in Hong Kong. *The Journal of teaching and Teacher education*, 18(1), 139-149.

PTU News (2004). *Yellow Ribbon Campaign to Stop the Killing*. Hong Kong Professional Teachers' Union, 98(469).

PTU News (2006). *Solidarity, Empathy and Tenacity*. Hong Kong Professional Teachers' Union, 132(503).

Puhakka, E. (1999). *Application of Two Stage Cluster Sampling in Finnish Data of IEA Civic Education Study 1999*. Institute for Educational Research, University of Jyväskylä.

Purkey, W. (1988). *An overview of self-concept theory for counselors*. Eric Clearing House on Counseling and Personnel Services, Ann Arbor, Michigan.

Purkey, S. C., & Smith, M. S. (1983). Effective Schools: A Review. *Elementary School Journal*, 83, 427-452.

Raley, S., & Bianchi, S. (2006). Sons, Daughters and Family Processes: Does Gender of Children Matter? *Annual Review Sociology*, 32, 401-421.

Ramserier, P. (2001). Motivation to learn as an outcome and determining factor of learning at school. *European Journal of Psychology of Education*, 16(3), 421-439.

- Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of Maths Self-Efficacy in the Structural Model of Maths Achievement. *Journal of Educational Psychology*, 85(1), 41-48.
- Rao, N., Moely, B. E., & Sachs, J. (2000). Motivational beliefs, study strategies and mathematics attainment in high-and low-achieving Chinese secondary school students. *Contemporary Educational Psychology*, 25(3), 287-316.
- Ratelle, C. F., Lacrose S., Guay, F., & Senecal, C. (2005). Perceptions of Parental Involvement and Support as Predictors of college Students' Persistence in a Science Curriculum. *Journal of Family Psychology*, 19 (2), 280-293.
- Reeve, J. (1998). Autonomy support as an interpersonal motivation style: Is it teachable? *Contemporary Educational Psychology*, 23, 312-330.
- Reines-Eudy, R. (2000). Using structural equation modeling to test for differential reliability and validity: An empirical demonstration. *Structural Equation Modeling*, 7(1), 124-141.
- Reynolds, D., & Teddlie, C. (2001). Reflections on the critics, and beyond them. *School Effectiveness and School Improvement*, 12(1), 99-123.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (1998). *Teachers, Schools and Academic Achievement*. Paper presented at the January 1998 Annual Meeting of the American Economic Association, Chicago, IL.
- Rose, D., & Sullivan, O. (1996). *Introducing data analysis for social scientists*. Open University Press, Buckingham.
- Rosenshine, B. (1983). Teaching Functions in Instructional Programs. *The Elementary School Journal*, 83, 335-351.

Ryan, R. M., & Deci, E. L. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55 (1), 68–78.

Ryan, R. M., Deci, E. L., & Grolnick, W. S. (1995). Autonomy, relatedness, and the self: Their relation to development and psychopathology. In D. Cicchetti & D. J. Cohen Eds., *Developmental psychopathology: Theories and methods*. New York: Wiley. 618-655.

Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom: Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology*, 50, 550-558.

Salomon, G. (1984). Television is easy" and print is "tough: The differential investment of mental effort in learning as a function of preconceptions and attributes. *Journal of Educational Psychology*, 76, 647-658.

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement. Research Report*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1998). The Tennessee Value-Added Assessment System (TVAAS): A Quantitative, Outcomes-Based Approach to Educational Assessment. In J. Milliman (Ed.), *Grading Teachers, Grading Schools: Is Student Achievement a Valid Evaluation Measure?* Thousand Oaks, CA: Corwin Press.

Sapsford, R., & Abbott, P. (1996). Ethics, Politics and Research. In R. Sapsford and V. Jupp. *Data Collection and Analysis*. Sage Publishing: London.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistics for moment structure analysis. *Psychometrika*, 66, 507-514.

Schau, C., & Mattern, N. (1997). Assessing students' connected understanding of statistical relationship. In I. Gal & J. B. Garfield, *The assessment challenges in statistics education*, eds. Amsterdam: IOS, 91-104.

Scheerens, J., & Creemers, B. P. M. (1989). Conceptualising school effectiveness. *International Journal of Educational Research*, 13(7), 691-706.

Schein, E. (1992). *Organizational culture and leadership (2nd Ed.)*. San Francisco: Jossey-Bass Publishers.

Schickendanz, J. A. (1995). Family Socialization and Academic Achievement. *Journal of Education*, 177, 17-38.

Schiefele, U., & Csikszentmihalyi, M. (1995). Motivation and ability as factors in mathematics experience and achievement. *Journal for Research in Mathematics Education*, 26(2), 163-181.

Schmidt, W., & Finnigan, J. (1992). *The Race without a Finish Line: America's Quest for Total Quality*. Jossey-Bass, San Francisco, CA.

Schreiber, J. B., & Chambers, E. A. (2002). After school pursuits, ethnicity, and achievement for 8th and 10th grade students. *Journal of Educational Research*, 96(2), 90-100.

Schunk, D. H. (1983). Developing children's self-efficacy and skills: The role of social comparative information and goal setting. *Contemporary Educational Psychology*, 8, 76-86.

Sekaran, U. (1992). *Research methods for business*, 2nd ed., New York: Wiley.

Shah, R., & Goldstein, S. M. (2006). Use of Structural Equation Modeling in Operations Management Research: Looking back and forward. *Journal of Operations Management*, 24(1), 148-169.

Shavelson, R. J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology*, 74, 3-17.

Shek, D. T. L., Lee, T. Y., & Chan, L. K. (1998). Perceptions of parenting styles and parent-adolescent conflict in adolescents with low academic achievement in Hong Kong. *Social Behavior and Personality*, 26(1), 89-98.

Shultz, K. S., Hoffman, C. C., & Palmon, R. R. (2001). *Using Archival Data for I-O Research: Advantages, Pitfalls, Sources, and Examples*. Annual Conference of the Society for Industrial and Organization Psychology (SIOP), San Diego, CA.

Shumway, S., Saunders, W., Stewardson, G., & Reeve, E., (2001). A comparison of classroom interpersonal goal structures and their effect on group problem-solving performance and student attitudes toward their learning environment. *Journal of Industrial Teacher education*, 38 (3), 6-24.

Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behaviour*, 23, 297-326.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417-453.

Skaalvik, E. M., & Hagvet, K. A. (1990). Academic achievement and self concept. An analysis of causal predominance in a developmental perspective. *Journal of Personality and Social Psychology*, 58, 292-307.

Skaalvik, E. M., & Rankin, R. J. (1992). Math and Verbal Achievement and Self-concepts: Testing the Internal/External Frame of Reference Model. *Journal of Early Adolescence*, 12, 267-279.

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85, 571-581.

Smithers, A. (2004). *England's Education. What can be learned by comparing countries?* Centre for Education and Employment Research, 1-55.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methods*, 13, 290-312.

Solberg, V. S., Gusavac, N., Hamann, T., Felch, J., Johnson, J., Lamborn, S., & Torres, J. (1998). The adaptive success identity plan (ASIP): a career intervention for college students. *The Career Development Quarterly*, 47, 48-92.

Sommer, K. L., & Baumeister, R. F. (2002). Self-evaluation, persistence, and performance following implicit rejection: The role of trait self-esteem. *Personality and Social Psychology Bulletin*, 28, 926-938.

Sorbom, D. (1974). A general method for studying differences in factor means and factor structures between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229-239.

Steenkamp, Jan-Benedict E. M., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25, 78-90.

Stevenson, H. W., & Lee, S. Y. (1990). *Contexts of achievement: A study of American, Chinese, and Japanese children*. Monograph of the Society for Research in Child Development, 55, (1-2, Serial No. 221).

Stiggins, R. J. (1985). Improving Assessment Where It Means the Most: In the Classroom. *Educational Leadership*, 43, 69-73.

Stiggins, R. J., Conklin, N., & Bridgeford N. J. (1986). *Insights into Classroom Assessment*. Portland, OR: Northwest Regional Educational Laboratory.

Stipek, D. J., & Gralinski, J. H. (1991). Gender Differences in Children's Achievement-related Beliefs and Emotional Responses to Success and Failure in Mathematics. *Journal of Educational Psychology*, 83, 361-371.

Stodalsky, S. S., Salk, S., & Glaessner, B. (1991). Student views about learning math and social studies. *American Educational Research Journal*, 28 (1), 89-116.

Strein, W. (1993). Structural Validity of an Academic Self-Concept Scale: An Illustrative Case Study of Large-Scale Archival data Set Use. *Measurement and Evaluation in Counseling and Development*, 26, 125-136.

Tam, T. K. (1997). Novice and Experienced Teachers' Instructional Activities in the Classroom. *Educational Research Journal*, 12 (1), 36-50.

TIMSS (2003). (ON-LINE <http://www.timss.bc.edu/timss2003.html> March 2007)

Thrupp, M. (1999). *Schools making a difference: Let's be realistic! School mix, school effectiveness and the school limits of reform*. Buckingham, Philadelphia: Open University Press.

Topping, K., Valtin, R., Roller, C., Brozo, W., & Dionisio, M. L. (2003). *Policy and Practice Implications of the Program for International Student Assessment (PISA) 2000*. Report of the International Reading Association PISA Task Force.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. New York: Academic Press.

Um, E. Y., Corter, J., & Tatsuoka, K. (2005). *Motivation, Autonomy Support, and Mathematics Performance: A Structural Equation Analysis*. Teachers College, Columbia University.

Usher, R. (1996). *Understanding Educational Research*. London: Routledge.

Vallerand, R. J. (2001). A hierarchical model of intrinsic and extrinsic motivation in sport and exercise. In G. C. Roberts (Ed.), *Advances in motivation in sport and exercise*. Champaign, IL: Human Kinetics.

Vallerand, R. J., Blais, M. R., Briere, N. M., & Pelletier, L. H. (1989). Construction and validation of the Academic Motivation Scale. *Canadian Journal of Behavioral Science*, 21, 323-349.

Vallerand, R. J., Blais, M. R., Briere, N. M., & Pelletier, L. H. (1993). On the assessment of intrinsic, extrinsic, and amotivation in education: Evidence on the concurrent and construct validity of the Academic Motivational Scale. *Educational and Psychological Measurement*, 53, 159-172.

Vallerand, R. J., Fortier, M., & Guay, F. (1997). Self-determination and persistence in a real-life setting: toward a motivational model of high school dropout. *Journal of Personality and Social Psychology*, 5, 1161-1176.

Vallerand, R. J., & Losier, G. F. (1999). An integrative analysis of intrinsic and extrinsic motivation in sport. *Journal of Applied Sport Psychology*, 11, 142-169.

Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.

Verma, V. (2002). *Comparability in International Survey Statistics*. International Conference on improving Surveys, 25-28. Copenhagen.

Vroom, V. H. (1964). *Work and Motivation*. New York. John Wiley & Sons, Inc.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Walberg, H. J., Paschal, R. A., & Weinstein, T. (1985). Homework's Powerful Effect on Learning. *Educational Leadership*, 42, 76-79.

Walkerdine, V. (1998). *Counting Girls Out: girls and mathematics*. London, Falmer Press.

Wang, J. J. (2001). An Empirical Study of Gender Differences in Chinese Students' Science Achievement. *Journal of Educational Research*, 90(4), 252-255.

Wang, J., & Staver, J. R. (1997). An empirical study of gender differences in Chinese students' science achievement. *Journal of Educational Research*, 90(4), 252-255.

Weiss, K., & Krappmann, L. (1993). *Parental support and children's social integration*. Paper presented at the biennial meeting of the Society for Research in Child Development, New Orleans, LA. (ED 361 077)

Wenglinsky, H. (1998). Finance equalization and within-school equity: The relationship between education spending and the social distribution of achievement. *Educational Evaluation and Policy Analysis*, 20(4), 269-283.

Wentzel, K. R. (1998). Social relationships and motivation in middle school: The role of parents, teachers and peers. *Journal of Educational Psychology*, 90, 202-209.

Widaman, K., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, 8, 16-37.

Williams, G. C., Freeman, Z. R., & Deci, E. L. (1998). Supporting autonomy to motivate patients with diabetes for glucose control. *Diabetes Care*, 21, 1644-1651.

Williams, L., & Holahan, P. J. (1994). Parsimony-based Fit Indices for Multiple-indicator Models; Do They Work? *Structural Equation Modeling*, 1(2), 161-189.

Willingham, W. W., & Cole, N. S. (1997). Research on gender differences. In W.W. Willingham, & N.S. Cole, (Eds.). *Gender and Fair Assessment*. NJ, Lawrence Erlbaum.

Wilson, S., Floden, R., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations. A research report prepared for the U.S. Department of Education*. University of Washington, Center for the Study of Teaching and Policy, Seattle.

Wiseman, A. W., & Brown, D. S. (2002). *Does Teacher Preparation Really Matter? The Influence of Teacher Preparation on Student Achievement in the United States and Abroad*. Conference of the Oklahoma Association of Teacher Educators (OATE).

Wong, R. S. (1998). Multidimensional influences of family environment in education: The case of socialist Czechoslovakia. *Sociology of Education*, 71, 1-22.

Wong, K. C., Lam, Y., & Ho, L. M. (2002). The Effects of Schooling on Gender Differences. *British Educational Research Journal*, 28(6), 827-843.

Worsham, M. E., & Evertson, C. M. (1980). *Systems of Student Accountability for Written Work in Junior High School English Classes*. R&D Report No. 6105. Austin, TX: Research and Development Center for Teacher Education.

Wright, S. P., Horn, S. P., & Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(30). Boston: Kluwer Academic.

Yeung, A. C. L., Cheng, T. C., & Chan, L. Y. (2004). From Customer Orientation to Customer Satisfaction: The Gap between Theory and Practice. *IEEE Transactions on Engineering Management*, 51(1), 85-97.

Yeung, A. C. L., Lee, T. S., & Chan, L. Y. (2003). Senior management perspectives and ISO 9000 effectiveness: an empirical research. *International Journal of Production Research*, 41(3), 545-569.

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publishing.

Younger, M., Warrington, M., & Williams, J. (1999). The gender gap and classroom interactions: Reality and Rhetoric? *British Journal of Sociology of Education*, 20(3), 325-341.

Zhang, L. F., & Sternberg, R. J. (1998). Thinking styles, abilities, and academic achievement among Hong Kong university students. *Educational Research Journal*, 13(1), 41-62.

Zimmerman, B. J., & MartinezPons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82, 51-59.

Zuckerman, M., Porac, J. F., Lathin, D., Smith, R., & Deci, E. L., (1978). On the Importance of Self-Determination for Intrinsically Motivated Behavior. *Personality and Social Psychology Bulletin*, 4 (3), 443–46.