MEASUREMENT INVARIANCE OF INVENTORY OF SCHOOL MOTIVATION BETWEEN THE UNITED STATES AND CHINESE COLLEGE STUDENTS

By

LIHUA XU

Bachelor of Arts in English
Sichuan University
Chengdu, Sichuan
1998

Master of Arts in English Language and Literature
Sun Yat-sen University
Guangzhou, Guangdong
2003

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2009

MEASUREMENT INVARIANCE OF INVENTORY OF SCHOOL MOTIVATION BETWEEN THE UNITED STATES AND CHINESE COLLEGE STUDENTS

Dissertation Approved:

Dr. Laura Barnes, PhD

Dissertation Adviser

Dr. Katye Perry, PhD

Committee Chair

Dr. Janice Miller, PhD

Committee Member

Dr. Diane Montgomery, PhD

Outside Adviser

Dr. A. Gordon Emslie, PhD

Dean of the Graduate College

ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Laura Barnes, Dr. Katye Perry, Dr. Janice Miller and Dr. Diane Montgomery. Without them, this dissertation would have been impossible. I would like to extend my special thanks to Dr. Katye Perry for taking the initiative to chair my committee and for her support and advice during my graduate study.

This project would not have been completed without the love and support from my family. My father and siblings have always been indispensable in their support of my educational endeavors. My special thanks go to my eldest brother, Pengfei Xu, for his counsel and advice in doctoral studies. I owe a great debt of gratitude to my elder sister, Lijun Xu, for her continued attention to my parents.

I am grateful to all my friends for their encouragement during my graduate study. Specially, I would like to thank Larry Shawn Bassham for his assistance with and advice for my dissertation. His help with dissertation format, Endnote, and his intellectual insights on cross-cultural differences are very valuable and greatly appreciated.

I would like to thank Dr. Dale Fuqua for his much-needed reference to work in University Assessment and Testing. He has truly made a difference in my graduate study. Also, I would like to thank Dr. Mark Gavin for this suggestions and recommendations regarding using structural equation modeling. Without serving in my committee and with

no obligations to help me, he offered much-needed advice. Thanks for being accessible and for the quick turnaround to my questions.

I would like to express gratitude to Dr. Pam Bowers, Jennifer Stevens, Chris Ray, and the other staff members in University of Assessment and Testing for their cooperation and encouragement throughout the duration of this undertaking.

I would like to thank my Chinese fellows, Yan Yang, Xiaoling Ke, Grace Yan Johnson, Min Wei for helping me translate the measuring instrument. Special gratitude goes to Weiyan Lin and Danping Huang for assisting me to obtain the approval from their university officials for authority in recruiting Chinese participants for my research. Without their aid in data collection this dissertation would not be possible.

Finally, I am forever indebted to my survey participants. Although your names are not known, I recognize the value in your time and effort for my data collection.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii
LIST OF TABLES vii
LIST OF FIGURES viii

Chapter I. Introduction 1

Maehr's Personal Investment Model 5
Importance of Cross-Cultural Studies 6
Statement of Problem 8
Purpose of Study 10
Significance of Study 12
Preview of Remaining Chapters 12

Chapter II. Review of Literature 14

Definitions and Operationalization of Achievement Goal Theory 14
Limitations of Achievement Goal Theory 17
Issues with Achievement Goal Measurement 20
Single-item instrument. 21
Target population. 21
Domain specificity. 22
Construct validation. 22
Review of Some Popular Achievement Goal Measures 23
Maehr's Personal Investment Model 27
Validation Studies about ISM 29
Structural Equation Modeling and Confirmatory Factor Analysis 34
Measurement Invariance 36
Chapter Summary 42

Chapter III. Method 43

Participants 43
Procedures 44
Instrument 45
Inventory of School Motivation. 45
Demographic information 48
Instrumentation Translation 50
Analysis Procedures 52
Computer Software 53
Chapter Summary 54

Chapter IV. Results 55

Sample Characteristics 55
Descriptive Statistics 57
Psychometric Characteristics of ISM across the Two Samples 60
Single-sample Confirmatory Factor Analysis 62
Hierarchical Measurement Invariance Test across Cultures for each Subscale 65
Configural, metric and scalar invariance testing for the subscale Task. 67
Configural, metric and scalar invariance testing for the subscale Effort. 68
Configural invariance testing for the subscale Competition. 70
Configural invariance testing for the subscale Social Power. 71
Configural, metric and scalar invariance testing for the subscale Social Concern. 71
Configural, metric and scalar invariance testing for the subscale Affiliation. 73
Configural, metric and scalar invariance testing for the subscale Praise. 75
Configural invariance testing for the subscale Token. 76
Testing for Differences in Observed Means and Latent Factor Means 78
Summary 80

Chapter V. Discussion and Conclusion 82

Implications for the Theory 84
Implications for the ISM 86
Implications for Cross-Cultural Comparisons 92
Necessity of measurement invariance test. 92
Cross-cultural comparisons of latent means. 93
Practical Implications 96
Limitations 97
Suggestions for Further Research 98
Conclusion 100

References 102

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL 119
APPENDIX B: INSITUTIONAL REVIEW BOARD APPROVAL OF MODIFICATION 120
APPENDIX C-1E: EMAIL INVITATION TO FACULTY (ENGLISH VERSION) 121
APPENDIX C-1C: EMAIL INVITATION TO FACULTY (CHINESE VERSION) 122
APPENDIX C-2E: FOLLOW-UP EMAIL TO FACULTY (ENGLISH VERSION) 123
APPENDIX C-2C: FOLLOW-UP EMAIL TO FACULTY (CHINESE VERSION) 124
APPENDIX D-E: WEB INTRODUCTION SCRIPT (ENGLISH VERSION) 125
APPENDIX D-C: WEB INTRODUCTION SCRIPT (CHINESE VERSION) 126
APPENDIX D-M: INTRODUCTION TO WEB SURVEY -MODIFIED 127
APPENDIX E-E: SURVEY INSTRUMENT (ENGLISH VERSION) 128
APPENDIX E-C: SURVEY INSTRUMENT (CHINESE VERSION) 131
APPENDIX F: TRANSLATION EVALUATION FORM 133
VITA 139

LIST OF TABLES

TABLE PAGE

Table 1 *Descriptive Statistics for ISM Items across Cultures* 59

Table 2 *Reliabilities and Composite Reliabilities for 8 Subscales* 61

Table 3 *Tests of Measurement Invariance and Population Heterogeneity of ISM in American and Chinese Students* 62

Table 4 *Unstandardized and Standardized Coefficients for American and Chinese students* 64

Table 5 *Correlations among Factors within each Cultural Groups* 65

Table 6 *Step-wise Invariance Test for Subscale Task across Cultures* 67

Table 7 *Step-wise Invariance Test for Subscale Effort across Cultures* 69

Table 8 *Configural Invariance and Within Group CFA Test for Subscale Competition* .. 70

Table 9 *Configural Invariance and Within Group CFA Test for Subscale Social Power* 71

Table 10 *Step-wise Invariance Test for Subscale Social Concern across Cultures* 72

Table 11 *Step-wise Invariance Test for Subscale Affiliation across Cultures* 74

Table 12 *Step-wise Invariance Test for Subscale Praise across Cultures* 75

Table 13 *Configural Invariance and Within Group CFA Test for Subscale Token* 77

Table 14 *Group Differences for ISM Factors based on Observed Scores and Latent Factor Scores* 79

LIST OF FIGURES

FIGURE PAGE

Figure 1. Theoretical Structure for the ISM.....49

Figure 2. Participants' Classification in College 56

Figure 3. Parents' Income ($ vs. ¥) Stated by the American and Chinese Participants....57

Figure 4. Origin of Residence as Stated by the American and Chinese Participants 57

Chapter I

Introduction

Chinese culture has long been known for the value it places on education as the primary avenue to social and financial advancement. With the economic prosperity that has occurred in the last two decades in China, some people are saying education is less useful because educational qualifications are no longer positively related to income levels (1996). However, the vast majority of Chinese people still believe education provides a means for personal improvement. The centuries-old national examination in China provided various government agencies with civil officials. The examination system has not declined since then and is still the primary route for social mobility. Not everyone has the opportunity to go to high school or college, though. Only those who pass the entrance examinations are eligible to attend high school and college. The size of the population in China and the relative scarcity of resources determine that only those individuals who can survive the fierce competition through education can succeed. Only 20% of the Chinese population is admitted to college (Phelps, 2005).

Education is valued not only for the external rewards it can bring to individuals, but because it can satisfy individuals' internal and social goals. Chinese students work hard at school not only for their future job and financial security but to please their parents, to be a good daughter or son, and to bring honor to their families. With regard to social goals, students' academic success is the primary expectation of their parents. It can

bring pride, joy, and honor to their entire family; whereas, academic failure in school can let down their family and cause them to lose face/reputability (Salili, 1995). Attaining academic excellence may be considered the best way for a student to repay their parents (Tao & Hong, 2000). With regard to internal goals, under the influence of Confucian teaching, which emphasizes hard work, effort, and endurance, Chinese students are socialized to value hard work and excellence in education (Salili & Hau, 1994; Yang, 1986). When talking about the significance of education in Chinese culture, Lee (1996) states that, "internally, education is important for personal development, and associated with it is the notion of human perfectibility, which is believed to be achievable by everyone. Externally, education is important for social mobility, and is also believed to be achievable by whosoever aims to do so" (p. 39).

Human perfection is the highest purpose of life which requires life-long dedication and effort toward learning on the part of learners (Li, 2004). Influenced by Confucian tradition, Chinese students are motivated, "in addition to mastering knowledge in specific academic domains, to cultivate their inner virtue (*neisheng*) and to assume 'meritorious service' (*waiwang*)" (Li, 2004, p. 117).

Interestingly, though Chinese students presumably rank high in motivation for academic achievement within their own culture, this has not been demonstrated in the research literature on comparative education. For instance, in studies done by McClelland and his colleagues (McClelland, 1961), Chinese students were labeled as low in achievement motivation based on their scores on the Thematic Apperception Test. The test was originally developed by Morgan and Murray (1935) but was transformed by McClelland and his colleagues into a major tool for assessing achievement motivation. In

those studies, the role of culture was not taken into consideration. The concept of achievement motivation was assumed to be universal and its measuring instrument was assumed to be cross-culturally valid.

McClelland and his colleagues (Atkinson, 1964; McClelland, 1961) are the pioneers and flagship figures in the study of achievement motivation, and the field owes a debt to them. Their definition of achievement motivation is based on Murray's (1938) definition, "the desire or tendency to do things as rapidly and/or as well as possible… to accomplish something difficult…to overcome obstacles and attain a high standard…to excel one's self… to rival and surpass others" (p. 164). Striving for excellence and the will to achieve are normal human behavior characteristics. This definition includes both mastery (to do things as rapidly and/or as well as possible) and performance goal orientation (to rival and surpass others), which are two prevalent goal orientations in achievement goal research. However, the definition fails to take into account the social goals which are significant in a collectivistic culture like the Chinese culture where achievement is more than individual endeavors. Such social goals include but are not limited to the goals to meet parents' expectations, to please parents, to be a good daughter or son, and to bring honor to one's family.

McClelland and his colleagues' theory about achievement motivation and the measure, Thematic Apperception Test, is based on an individualistic ideology of achievement. Therefore, evidence for the validity of this measure on a collectivist culture such as China should be obtained if the TAT is to be used in cross-cultural studies. Maehr (1978) observed that "McClelland has created an ethnocentric approach to motivation, an approach that simply compares other cultures to a western prototype without doing

justice to the potential for excellence that exists within these cultures" (p. 130). Such studies often "had the effect of 'elevating' the goals, perceptions, and behaviors of Western cultural groups to the status of universal norms" (Maehr & McInerney, 2004, p. 63). Traditional western theories suggest that the achievement motivation trait is developed through childrearing practices which stress independence, mastery, and competitiveness (McClelland, 1963; Salili, 1996a). However, Chinese parents emphasize proper behavior, strict discipline, and place less emphasis on the child's expression of opinion, independence, creativity, and all-round personal development (Ho & Kang, 1984; Maehr & McInerney, 2004; Salili, 1995).

Lack of culturally sensitive theory and corresponding measures in the achievement motivation field should not impede cross-cultural research. The consistent gap in achievement between American elementary and high school students and their Asian peers, particularly in the field of sciences and mathematics, has led researchers and educators to search for the possible reasons. This becomes increasingly urgent with globalization where technologically relevant skills are in great demand. The good news is that research in cross-national achievement gaps in sciences and mathematics has fit hand in glove with the research in achievement motivation (J. G. Elliot & Bempechat, 2002).

There is a belief that cultural values mediate achievement behavior. Stevenson, in collaboration with other scholars (Chen, Lee & Stevenson, 1995; Stevenson & Stigler, 1992), has contributed substantially to this area of study. They reported that Asian students were said to hold the view that effort is more important than ability as the cause of success and failure, while American students were strong believers in the importance of innate ability. However, Marsh (1984) has not found that effort correlates positively

with achievement outcomes. Instead, he found that high achievement is associated with attributing success to ability and failure not to lack of ability (See Bempechat & Drago-Severson, 1999). Cross-cultural research in achievement goal theory will help answer the question concerning the role of effort and ability in school performance for cultures not included in the original research.

Maehr's Personal Investment Model

Maehr's (1984; Maehr & Braskamp, 1986) personal investment model seems to have addressed the limitations of cultural and social relevancy. Despite its western origin, it is designed to be applicable in cross-cultural settings (Watkins, McInerney & Lee, 2002). It is a social cognitive theory and recognizes the role of cultural and situational factors in influencing individuals' thoughts and perceptions which in turn affects their choice, persistence, and change in levels of activity involved in achievement tasks. To be more specific,

> Personal Investment (PI) Theory is concerned with how persons choose to invest their energy, talent, and time in particular activities. PI theory is particularly relevant in investigations into how individuals of varying social and cultural backgrounds relate to different achievement situations such as schooling (Maehr & McInerney, 2004, p. 73).

Sociocultural factors play an important role not only in what is worthy of individuals' personal investment but also how individuals should invest themselves. In addition, PI theory emphasizes the subjective meaning that individuals assign to achievement situation based on their belief systems developed culturally; that is, whether individuals will invest themselves in particular activities depends on what the activities

mean to them (Maehr, 1984; Maehr & Braskamp, 1986; Maehr & McInerney, 2004). Furthermore, the model suggests multiple goals for motivated action that are applicable in a cross- cultural setting. In particular, "it conceptualizes achievement motivation in terms that recognize the possibility of diverse modes of achievement behavior across cultures and groups" (Maehr, 1984, p. 74). Thus, it seems to provide a theoretical framework for cross cultural comparison of group differences.

Based on Maehr's theory, McInerney and his colleagues proposed a hierarchical, multidimensional model of goal orientations to represent a set of goals relevant for both Western and non-Western students. The Inventory of School Motivation (ISM; McInerney & Sinclair, 1991) was developed to assess the constructs relevant to the model in educational settings. Those constructs include *Sense of Self* (sense of purpose and sense of competence) and *Personal Incentives* (task involvement, effort, praise, competition, social power, token rewards, social concern, and affiliation).

Importance of Cross-Cultural Studies

Cross-cultural studies of achievement goal theories are important in the sense that they provide a means to test the external validity and generalizability of the theory, model and the measure (Marsh & Hau, 2004). As Sue (1999) argued, psychological research has not made full use of cross-cultural studies which made it impossible to generalize their interpretations and evaluate the applicability of their theories. Similarly, Ali and McInerney (2005) argue that the "psychological basis of achievement motivation, when integrated with principles of cultural anthropology and cultural psychology, will move both theory and research forward" (p. 2).

In order for the Inventory of School Motivation to be useful for meaningful cross-cultural comparison on achievement motivation, its measurement invariance across cultures needs to be tested. That is, ISM must have similar measurement qualities across cultures. As mentioned earlier, achievement gap on mathematics and science tests between U.S. and some Asian countries could potentially be attributed to differences on achievement motivation. Measurement equivalence allows researchers to have confidence in: first, measure constructs are applicable across cultures; second, scale items are interpreted consistently by respondents across cultures; third, rating scales are calibrated similarly across cultures; finally, observed mean differences reflect the mean differences of underlying latent traits (Drasgow, 1984).

Culture can influence construct comparability/measurement equivalence in at least two ways. First, the psychological instrument developed under one culture may not be able to measure the same construct in another culture. Referring to psychological assessments conducted in Asia, Sue and Chang (2003) pointed out that research on cultural differences and similarities assumes that we have valid and equivalent tools with which to evaluate these findings. Cultural values and beliefs can greatly affect item responses to measures assessing social and psychological constructs that are developed and administered. Second, cultural tendency to respond in a particular way (e.g. frequent use of the low and high end of the response scale) might cause nonequivalence.

Chinese and American cultures are different in a lot of ways; however, similarities between the two cultures make the test of measurement invariance relevant and meaningful. The differences between these two cultures in achievement goal orientations are a matter of quantity not quality. In a bipolar continuum with collectivism

and individualism as two extremes, Chinese culture is generally categorized as a collectivist culture and American culture as individualist. Although American culture is said to promote independence, autonomy and self-reliance, socially oriented goals still play an important role in motivating students, just as learning and performance goals. For instance, in the study of Chinese, Canadian Chinese and Canadian students, Salili, Chiu and Lai (2001) found that family-related goals (e.g. studying to please parents) and teacher-oriented social goals both drive students to do well in the classroom. Canadian culture is believed to be very similar to American culture in terms of individualist ideology. Hence, not surprisingly, Maehr and Yamaguchi (2001) suggest that the general goal theory framework is applicable substantially across widely divergent cultural groups. In addition, research (Fyans, Salili, Maehr & Desai, 1983; Maehr & McInerney, 2004) indicates that there is a near-universal view of achievement and achievement motivation. The concepts of learning and performance goals originated in Western society; however, they are similarly relevant to the Chinese "Eastern" culture. The universality in view of achievement motivation and cross-cultural applicability of achievement goal orientation makes the study of measurement equivalency meaningful.

Statement of Problem

One of the major methodological limitations of early cross-cultural comparisons is that psychological instruments are assumed to measure the same thing across cultures. No attempts are taken to ensure that psychological constructs are culturally appropriate and that the behaviors or attitudes that psychological instruments are measuring are functionally, conceptually, and metrically equivalent (McInerney, Yeung & McInerney, 2001). The observed mean level differences in scores are interpreted as construct

differences. However, as Vandenberg and Lance (2000) point out, "if not tested, violations of measurement equivalence assumptions are as threatening to substantive interpretations as is an inability to demonstrate reliability and validity" (p. 6).

To demonstrate that the construct has cross-cultural validity, the first step is to establish that, theoretically, the construct is relevant to both cultures. (For instance, spelling ability is a construct that has meaning in the U.S. because of the plurality of English word origins but has little relevance in many non-English speaking cultures.) Assuming cross-cultural construct relevance, the next steps involve demonstrating the extent to which the instrument functions the same way across the groups. This is often accomplished through confirmatory factor analyses which assess the degree to which a common statistical model describes the statistical properties and structure of the instrument across the groups. These tests of invariance are often done in stages where the first question is whether the same number of constructs are measured in the two groups; then whether the pattern of item loadings are the same in the groups; then whether the magnitude of loadings are the same; then whether specifications concerning intercepts are the same, and so on. Once these equalities are demonstrated, then comparisons involving means and associations across the groups can be meaningfully interpreted and their relationships to other variables investigated.

Early factorial validation studies by McInerney and his associates have demonstrated that the Inventory of School Motivation (ISM) has reliability and validity for measuring motivational goal orientations in Western culture (McInerney & Sinclair, 1991; McInerney & Swisher, 1995). However, little attempt is made to demonstrate the applicability of the construct to groups other than the one for which ISM was developed

and standardized. An invariance test of the ISM was conducted by Ali and McInerney (2005) with high school students across seven cultural groups, including Hong Kong. The results suggested a stable and reliable factor structure across cultures and provided some evidence that the ISM scales are applicable to students from various cultural backgrounds. The problem that the current study is addressing is whether the eight-factor model of ISM is equivalent between U.S. and China in terms of factor structure, item factor loadings and item intercepts, and if yes, whether the factor means differ across cultures**.**

Purpose of Study

This study is different from Ali and McInerney's (2005) in several ways. First, despite the common Confucian heritages, Hong Kong has strong influence of individualism due to the over century-long colonization by Great Britain. Hong Kong is one of the biggest world economical and trade harbors. Therefore, in terms of cultural ideology, Hong Kong has a mix of both Eastern and Western cultures. Second, instead of studying high school students as Ali and McInerney did, this project is interested in the achievement goal orientations of university students. No study has been done with university students for their interpretations of ISM. Third, the Chinese version of ISM that Hong Kong students completed in Ali and McInerney's study is in Cantonese. This study will use a Mandarin Chinese version of ISM for its mainland Chinese participants. Mandarin is the official language across China. Finally, Ali and McInerney did not address the mean level similarities and differences in constructs of goal orientations across cultures. This lack is significant because such comparisons are able to specify the role of culture in shaping individuals' achievement goal orientations. For instance, the

significance of social goals (e.g. affiliation) in collectivistic cultures will provide evidence for the classical collectivism and individualism argument.

The present study was designed to accomplish the following goals. The first goal was to test whether the construct of achievement goal orientations as operationalized by McInerney's Inventory of Student Motivation is comparable between the two cultures. That is, the goal was to test whether the factor structure of achievement goal orientations support an eight-factor model containing Task Involvement, Effort, Competition, Social power, Affiliation, Social concern, Praise, and Token across two cultures. If the same factor structure held for each culture, next was to test whether factor loadings and intercepts also are equivalent. Establishing factor comparability, the second goal was to test whether there are mean differences in the eight factors across the cultures.

This was primarily a measurement invariance study. That is, the primary goal was to examine the suitability of the ISM as a measure of achievement motivation in studies comparing Chinese and American cultures. Though the second purpose was to test mean differences in the eight factors, the ability to do that would be constrained by the invariance findings. That is, comparisons can only meaningfully be made on subscales that are invariant across cultures. Because the primary focus was on measurement invariance, data were not being collected which would allow meaningful interpretations of construct differences in a nomological network (e.g., no other reliable validating instruments were administered). The factor comparisons were included to provide evidence for validity of measurement invariance in cross-cultural studies.

The first goal was accomplished through a measurement invariance check of the ISM with the multi-group confirmatory factor analysis. The second goal was contingent

on the fulfillment of the first goal. That is, only when cross-cultural measurement invariance is established can the between-cultural mean differences in constructs be conducted. Otherwise, there is no way to pinpoint the source of the difference with certainty, as differences could be due to attitudinal variation or as a function of the psychometric properties of the particular items that are administered.

Significance of Study

Maehr's Personal Investment Model (Maehr & Braskamp, 1986) provides a theoretical framework for cross-cultural achievement goal comparisons due to the fact that it takes into consideration the sociocultural role in achievement goal orientation and its inclusion of social goals. It provides a potential model to explain the significant achievement gaps found across nations and to assess the role of culture in education and academic achievement. Its measuring instrument, the Inventory of School Motivation, is demonstrated to be a reliable tool to assess achievement goal orientation in Western cultures. The demonstration of construct validation of ISM in Chinese culture will make it a valuable tool to evaluate the traditional effort and ability argument found in the cross-cultural quantitative comparisons of school achievement. In addition, measurement equivalence/construct comparability is tested relatively infrequently. This study will help demonstrate its potential merits in cross-cultural construct validations.

Preview of Remaining Chapters

In Chapter 2, this researcher will provide a review of the predominant achievement goal theory and a critical analysis of selected instruments for measuring achievement goal orientations, focusing on academic settings. Chapter 2 will also include a review of early achievement goal theory and its limitations will be addressed. Several

popular achievement goal measures will be reviewed as illustration of measuring problems. Theoretical framework of this research, Personal Investment model, will be introduced and psychometric validation studies for its measuring instrument, Inventory of School Motivation, will be described in detail. Further, steps for testing measurement invariance will be provided here. Chapter 3 will provide detailed description about participant sampling and procedures of online survey administration. The ISM will be introduced in details. Procedures for translating the ISM into Mandarin Chinese will be described. Chapter 4 will provide detailed description of participant characteristics and results of statistical analyses. Chapter 5 will discuss the results in the context of the objectives of this study. Implications of the findings for achievement goal theory, construct measure, and for cross-cultural comparisons will be discussed.

Chapter II

Review of Literature

This chapter presents a review of literature relevant to the study. The first section will review the definitions and operationalization of achievement goal theory, then follows the limitations of achievement goal theory. The next section will present the issues of achievement goal measures. To help illustrate these issues, several popular achievement goal measures will be reviewed. The theoretical framework of this study, Maehr's Personal Investment model, will be presented next. The following section is about the development and psychometric validation of Inventory of School Motivation, which is based on Maehr's Personal Investment model. The last two sections of this chapter will be review of methodology applied in this study, confirmatory factor analysis, structural equation modeling, and sequential test of measurement invariance.

Definitions and Operationalization of Achievement Goal Theory

The theory that has received the most research attention in the past two decades in the achievement motivation area is achievement goal theory (A. J. Elliot & Thrash, 2001; Pintrich & Schunk, 1996). It remains the predominant approach to achievement motivation in the contemporary literature (Day, Radosevich & Chasteen, 2003). Generally speaking, achievement goals reflect "individuals' desire to develop, attain, or demonstrate competence at an activity" (Harackiewicz, Barron & Elliot, 1998, p. 2).

Achievement goal theories are concerned with students' perceptions and beliefs of purposes of achievement behavior in an academic situation. Some students may believe the purpose of performing well is to please their parents; others may believe the purpose of doing well in school is to outperform others and show how smart they are; and still others may believe the purpose of doing well is to gain social recognition from their friends or peers. Whatever these beliefs are, research indicates that those purposes and goals influence cognitive strategies, which in turn affect the quality of one's achievement. Specifically, these cognitive strategies include procedures involved in conducting school assignments, ranging from analyzing demands, utilizing resources, and monitoring progress to making attributions for their success and failure (Covington, 2000).

Research on achievement goal theories over the last two decades has largely focused on two goals: learning and performance goals (Dweck, 1986; Dweck & Leggett, 1988). A variety of labels have been used by different theorists to refer to these two goals, such as Nicholls' (1984) task-involved and ego-involved goals, Ames' (1992) mastery and performance goals, and Maehr and Midgley's (1991) task-focused and ability-focused goals. However, despite the various labels, students who hold learning goals are generally oriented to developing their skills and mastering tasks for their own sake and their feelings of competency are associated with effortful improvement. For those individuals, their comparing norms are self-referenced. In contrast, those who hold performance goals are oriented to surpassing others and demonstrating that they are smarter than others are. For those individuals, their comparing norms are other-referenced.

The early popular dichotomous conceptualization of goal orientation has been refined with the aid of empirical evidence. The initial findings about the effect of individuals' holding mastery and performance goal orientation on academic outcomes do not survive the scrutiny. To be more specific, although more adaptive outcomes generally are associated with mastery goal orientation, mixed outcomes are found to be associated with performance goal orientation (Harackiewicz, Barron, Carter, Lehto & Elliot, 1997).

To better explain reality, Elliot and his colleagues proposed a trichotomous framework of the goal orientation. More specifically, Elliot and Church (1997) divided performance goal orientation into performance approach goals and performance avoidance goals and left the mastery goals intact. A performance-approach goal is defined as an attempt to show competence in comparison to others while a performance-avoidance goal is defined as an attempt to avoid incompetence in comparison to others (Barron, Baranik & Finney, 2006). Through a self-report instrument, Elliot (1999) found that mastery goals were associated with positive results such as more persistence in studying, performance approach goals with both negative and positive consequences, and performance avoidance goals with only negative consequences.

More recently, Elliot and McGregor (2001) proposed a 2 x 2 achievement goal framework. Based on the trichotomous classification, they propose to divide the mastery goals into mastery approach goals and mastery avoidance goals. Individuals who hold mastery approach goals focus on improving skills and become more competent while individuals who hold mastery avoidance goals focus on not losing skills and not becoming incompetent. According to Elliot and McGregor, perfectionists tend to hold mastery avoidance goals in that they typically avoid making mistakes or doing anything

wrong. Despite the conceptual distinctiveness between performance approach and performance avoidance goals, empirically the two measuring scales were found to be correlated strong and positively (Urdan & Mestas, 2006). This raised the argument whether students could distinguish the approach and avoidance concerns (Roeser, 2004).

Limitations of Achievement Goal Theory

Achievement goal theory as represented above has several major limitations. First, although none of the theorists articulate explicitly the unidimensionality of achievement goals, the original dichotomy classification assumes a bipolar mastery versus performance goal continuum and individuals are categorized in terms of one type of goal or the other. Although not explicit, Dweck's (1986) early research suggested that goal orientation is a unidimensional construct with strong performance goal orientations at one end and strong learning goal orientation at the other end of a single continuum. Whether one adopts learning goal orientation or performance goal orientation depends on their beliefs about the malleability of ability. One cannot hold both goals because they cannot believe that ability is fixed and malleable at the same time. One of the problems with the bipolar continuum is the ambiguity of midpoint; that is, it represents that individuals have both learning and performance goal orientations or it represents that one has neither orientation (Button, Mathieu & Zajac, 1996). Similarly, Ames (1992) proposed that out of the two achievement goals orientations, students either choose performance goals to get positive judgment about their ability to look smarter or mastery goals to develop their ability by understanding and mastering new knowledge. However, recent theorizing and research suggest these goals are not incompatible; learning goal and performance goal orientations are actually two independent dimensions; and that students may hold both

goals simultaneously (Archer, 1994; Maehr, 1984; McInerney, 1995; Pintrich & Garcia, 1991; Urdan & Maehr, 1995). In addition, Barron and Harackiewicz (2001) outlined four possible patterns (additive pattern, interactional pattern, specialized pattern and selective pattern) with experiential evidence to validate the existence of multiple goals.

Second, theories of achievement motivation have typically ignored social goals, the effect of social relationships on academic achievement motivation in their studies of motivation (Blumenfeld, 1992; Juvonen & Weiner, 1993; Urdan & Maehr, 1995). Recently more researchers have realized the importance of studying social goals in addition to academic goals to better understand motivational dynamics (Anderman & Anderman, 1999; Covington, 2000; Deci & Ryan, 2000; Dowson & McInerney, 2001; Patrick, Anderman & Ryan, 2002; Urdan, 1997; Wentzel, 2000). In addition, Salili (1995) found that despite their cultures and gender, British and Chinese participants equally aspire to achieve in both individualistic and affiliative (social/family oriented achievement goals) situations. Social goals, apart from the academic goals, can help "organize, direct and empower individuals to achieve more fully" (Covington, 2000, p. 178). Social goals comprise important aspects of students' behavior, affect, and cognition in achievement settings and they may directly influence students' psychological processes as they work hard to achieve academically (Dowson & McInerney, 2001).

Unlike academic goals, students' social goals deal with their social reasons to strive for academic achievement (Urdan & Maehr, 1995). A number of social goals have been discussed and examined including being pro-social and responsible (Wentzel, 1993), pleasing the teacher (Wentzel, 1999), the desire to work with friends and peers (Ryan, 2001), gaining social approval, and bringing honor to the family (Urdan & Maehr, 1995).

In addition, cross-cultural research regarding the impact of teachers and parents on students' motivation and learning outcomes suggest that social goals are important motivational factors besides mastery and performance goals (Blumenfeld, 1992; Salili, Chiu & Lai, 2001). Both learning and performance goals imply a strong individualist Western flavor as opposed to the social and collectivist Eastern approach (McInerney, Roche, McInerney & Marsh, 1997; Watkins, McInerney & Lee, 2002). Social oriented goals such as affiliation with other students or wanting to succeed to please one's parents or honor one's family are likely to be salient in more collectivist cultures such as China (Ho, 1986). Though with the increasing modernization of China the younger generations of Chinese are becoming more like their individualistic Western counterparts (Yu & Yang, 1987), evidence indicates that collectivistic values still play a predominant role in Chinese's achievement (Salili, Chiu & Lai, 2001). Recent studies have found that family and social groups, as well as personal goals, have a great influence on Chinese students' achievement behavior. Academic excellence is often motivated by filial piety (obedience to one's parents) and making one's family proud (Salili, 1995; Salili, Chiu & Lai, 2001; Tao & Hong, 2000; Yu & Yang, 1994).

Wentzel (1992) argues that the attainment of high-level achievement is unlikely without joint pursuit of social, learning and performing goals. These goals can be complementary in that . . .

> "learning goals focus a student's attention on producing action for skill development…, whereas performance goals remind them of the long-term consequences of those actions…. As with performance goals, active pursuit of social goals can also promote achievement in that goals to be cooperative and

compliant are likely to direct attention to the instructional process and thus, support the pursuit of mastery and learning goals" (p. 292).

The third limitation of achievement goal theory is that, different terminology in achievement goal theories makes it difficult to determine whether mixed empirical findings are due to instruments based on different conceptual frameworks or due to methodological differences, for example, sampling characteristics, type of task, or achievement context (Day, Radosevich & Chasteen, 2003). Specifically, learning achievement goals are used interchangeably with mastery goals and intrinsic goal motivation, and performance achievement goals are used interchangeably with ego goals and ability goals. According to Elliot and Thrash (2001), different characteristics of the goal definitions are connected to different outcomes, which might help explain mixed empirical findings, for performance goals in particular.

Issues with Achievement Goal Measurement

Due to the growing popularity of achievement goal orientation theory, numerous measures have been developed to assess individual differences in this aspect and they have become theoretically and statistically sophisticated. In the aspect of theory, achievement goal orientations have gone from single dimension to two then three and four dimensions. In the aspect of statistical development, confirmatory instead of exploratory factor analysis is considered as a more effective and useful construct validation instrument in measure development. Nevertheless, achievement goal measures still have notable limitations such as single-item formats, inappropriate domain specificity levels, target population, generalizability, and insufficient construct validation evidence (VandeWalle, 1997). Several of these limitations are detailed below.

Single-item instrument. With the conceptualization of achievement goal theory as a single dimension, Bandura and Dweck (1985) developed a single forced-choice item to assess goal orientation. Based on their responses, respondents were classified as those with learning goal orientation and those with performance goal orientation. Ames and Archer (1987) also used a single-item measure to differentiate participants' goal values by applying a forced-choice question. Single-item instruments do not allow for estimation of internal consistency; thus, they are problematic in terms of psychometric properties (VandeWalle, 1997). Therefore, the accuracy and reliability of these instruments are questionable. In addition, these single-item instruments do not allow assessing the strength of performance and learning goal orientations (VandeWalle, 1997). With the recent development of multi-dimensional achievement goal theories, this becomes less a problem, as individuals are allowed to have a set of goals varying in degree. However, the number of items within a dimension needs to be considered due to its association with psychometric properties (e.g. reliability). Generally, the more items on a measure, the higher the reliability of the measure is (Leong & Austin, 1996).

Target population. Early research with achievement goal theories focused on children and adolescents (Ames & Archer, 1987; Bandura & Dweck, 1985; Nicholls, Patashnick & Nolen, 1985). Validation studies are rarely done with older populations, such as college students (VandeWalle, 1997). In addition, the norm groups of these instruments are primarily middle class Caucasians. Little attempt is made to validate these instruments with populations of other cultural heritages. More efforts are required to better understand achievement goal theory and possible variations among different groups in this aspect.

Domain specificity. Domain specificity is another issue for achievement goal orientation instruments (VandeWalle, 1997). At one extreme, some goal orientation instruments are too domain specific. For example, to measure students' performance and learning goal orientations Archer (1994) surveyed students in a psychology class by asking them when they felt most successful. It is unknown, however, whether the responses to such context-specific items could be generalized to other contexts. At the other extreme, Button et al. (1996) created an instrument to measure global learning and performance goal orientations. Items such as "The opportunity to learn new things is important to me" and "The opportunity to do challenging work is important to me" are general and not contextualized in academic or work situations. This approach, therefore, may also be problematic. Chiu, Hong, and Dweck (1994) suggested that goal orientation is domain specific by pointing out that being mastery goal oriented in the academic field may not be generalized to the athletic field. In addition, domain specificity for achievement goal scales was supported when Baranik et al. (2007) failed to combine goal items across academic and work domains using confirmatory factor analysis. Consistent with Ajzen's (1987) recommendation in the aspect of domain specificity, VandeWalle (1997) called for more goal orientation measures to be "operationalized at a midlevel of specificity: at the level of major life domains such as academics, work, and athletics" (p. 1002).

Construct validation. Another issue with existing goal orientation instruments is that many instruments (e.g., Duda & Nicholls, 1992; Nicholls, Patashnick & Nolen, 1985) limit construct validation evidence to exploratory factor analysis and reliability analysis using Cronbach's alpha and those instruments lack evidence from a thorough

confirmatory factor analysis (VandeWalle, 1997). Confirmatory factor analysis (CFA) is an important validation tool for its objectivity in comparison to exploratory factor analysis where a lot of subjective judgment calls are involved. CFA provides a means for construct validation not only within a culture but also across cultures. For within-culture validation, it tests how well the hypothesized model in terms of relationships of constructs with measuring items can explain the observed data. For cross-cultural validation, it tests whether constructs are measured the same way across groups.

Review of Some Popular Achievement Goal Measures

The above pointed out a few issues with existing achievement goal measures and the following are illustrations of these issues using a few popular measures. Given that achievement goal theory has been one of the most predominant framework for understanding achievement motivation over the last few decades, it is not surprising that there are abundant instruments to measure student motivation in the literature; for example, the *Motivated Strategies for Learning Questionnaire* (Pintrich & Garcia, 1991); the *Motivational Orientation Scale* (MOS; Nicholls, 1989; Nicholls, Patashnick & Nolen, 1985); the *Patterns Adaptive Learning Survey* (PALS; Midgley et al., 1996); the *Inventory of School Motivation* (ISM; McInerney & Sinclair, 1991); and the *Learning and Performance Goal Orientation Scale* (Button, Mathieu & Zajac, 1996). According to Jagacinski and Duda (2001), some of the existing goal orientation scales in the literature measure goals in certain areas, such as the academic domain (Midgley et al., 1996; Nicholls, Patashnick & Nolen, 1985), sport context (Duda, 1998; Duda & Nicholls, 1992; Duda & Whitehead, 1998), and work settings (VandeWalle, 1997). Others (Button,

Mathieu & Zajac, 1996) have tried to develop more general goal orientation measures to be applicable across different domains.

Given the number of achievement goal orientation measures, it seems important to evaluate whether these scales are in fact measuring the same construct (i.e., task and ego orientations). In addition, it is also important to evaluate how well these popular measuring instruments of goal orientations measure the construct they claim to measure in terms of psychometric characteristics (e.g. internal consistency, factor validity etc.) (Jagacinski & Duda, 2001). For this study, we only focus on goal orientation measures toward academic goals in educational setting, that is, the MOS, PALS, ISM, and the *Learning and Performance Goal Orientation Scale*. The first three scales were formulated to assess goal orientations toward academics, whereas the last set of scales was developed to apply across domains. In this section the rest three measures except the ISM will be reviewed. The ISM and its development and psychometric validations will be discussed after the review of its theoretical framework. The purpose of such arrangement is two-fold. First, the ISM is developed based on Maehr's Personal Investment model. Therefore, it is logical to present it after its theoretical framework. Second, the purpose of this study is to test the measurement equivalence of the ISM cross-culturally. Hence, a more detailed discussion is warranted.

First, we start with Nicholls' *Motivation Orientation Scales*. Conceptually grounded in Nicholls's theory (Nicholls, 1984, 1989), a number of forms of MOS have been developed to assess goal orientations in the academic context for elementary school students through college students (Jagacinski & Duda, 2001). Instead of asking students directly about their goals, the MOS measures goals by asking students when they feel

most successful. The task items suggest that success results from working hard, learning new things, and figuring out new strategies; on the other hand, the ego items suggest that success results from overdoing others. Another characteristic of the MOS is that it merges Ego goals with Social goals to form the "Ego and Social Orientation" scale. Although the rating items share some common characteristics, they belong to different goal orientations. It seems that Nicholls is aware of their distinctions as he stated that the purpose of social goals "is to indicate virtuous intentions or personal commitment rather than ability" (Maehr & Nicholls, 1980, p. 242). There is no way to know how closely these items correlate with each other and with the constructs they are supposed to measure since Nicholls did not report the reliability coefficient for this scale. According to Urdan and Maehr (1995), the merging of social goals with ego goals can obscure the distinct effects of social goals on students' motivation and achievement in school.

Carol Midgley and her colleagues (1996) developed the PALS to assess a range of motivational constructs including personal achievement goal orientations, which are grounded in the Dweck and Leggett's (1988) theoretical framework. The earliest versions of PALS included task goal orientation scale, a performance scale, and an extrinsic scale but did not include a performance-avoidance scale. After 1997, the extrinsic scale was dropped and a performance-avoidance scale was added. Therefore, a trichotomous achievement goal theoretical framework has been adopted, including task goal orientation, performance approach orientation and performance avoidance orientation. The various forms of PALS make it hard to pinpoint what form is used in various studies due to the lack of specificity in the description of the measure used (Ross, Blackburn & Forbes, 2005). A reliability generalization study of 30 studies using PALS indicates that its

reliability is fairly consistent across studies. However, “caution must be used in interpreting scores from studies conducted at the elementary school, high school, and college levels, because these sample characteristics and contexts have not been fully explored in relation to reliability of PALS scores” (Ross, Blackburn & Forbes, 2005, p. 461).

Button et al. (1996) developed an instrument to provide a more general and stable assessment of goal orientations to be of use with working adults, when most goal orientation scales target their population at children in academic settings. This scale consists of 16 items with eight items to measure Learning Goals and eight to measure Performance Goal Orientations. The Button’s et al. items are more general and are intended to assess overall orientations assumed to generalize across different activities (Jagacinski & Duda, 2001).

Different from the goal measurements discussed above, Button’s et al. (1996) General Learning and Performance scales were developed and validated entirely with the use of confirmatory factor analysis. The General Learning and Performance scales were found to be independent constructs and both scales were correlated significantly with Dweck’s Theories of Intelligence measure (Button, Mathieu & Zajac, 1996). However, later research (Jagacinski & Duda, 2001) did not confirm these results. Button’s et al instrument draws several criticisms. First, goals are assessed at a general level rather than at a level specific to work as it claimed (Baranik, Barron & Finney, 2007). Second, Jagacinski and Duda (2001) pointed out that the General Performance scale did not balance well in terms of the reference group against which the items were constructed. To be more specific, items are more representative of the expected reactions of individuals

low in perceived ability than those high in perceived ability. However, Jagacinski and Duda (2001) suggested that with further refinement and validation, Button's et al instrument holds promises in being as predictive an instrument as the more context-specific assessments.

Maehr's Personal Investment Model

Maehr's (1984; Maehr & Braskamp, 1986) Personal Investment (PI) model use multiple goals to explain human motivation. It appears to be an extension of goal theory to specifically address the limitations. It is interested in why and how individuals choose to invest their energy and time in particular activities or courses of actions. According to PI theory, meanings of those activities or action determine personal investment. Individuals exhibit different patterns of personal investment because they perceive the investment situations differently. PI theory is a social cognitive theory which recognizes the importance of sociocultural and situational factors in determining the meaning of the situations to individuals, what is worthy of personal investment, and also how individuals should invest themselves.

PI theory designates three interrelated facets as components of meaning which are considered critical to determine personal investment in particular activities. They are (1) *personal incentives* associated with performing in a situation, (2) *beliefs about self* and (3) *perceived options or facilitating conditions* available in a situation. *Personal incentives* refer to the incentives that individuals find salient to themselves and in particular what defines success or failure for individuals in a particular situation. Those incentives include Task incentives, Ego incentives, Social incentives, and Extrinsic incentives. *Beliefs about self* refer to individuals' organized perceptions, beliefs and feelings

regarding who they are. *Perceived options or facilitating conditions* available in a situation refer to the behavioral alternatives that individuals perceive to be available and appropriate in a given situation and the environmental factors that are likely to impact on motivated behavior (McInerney & Van Etten, 2002). Each component of PI theory may be influenced by personal experiences, personality, age/life stage, the character of the performance situation and, importantly, the sociocultural environment in which tasks, situations, and individuals are rooted (Ali & McInerney, 2005).

In the field of achievement motivation research, PI theory antedated achievement goal theories; however, it incorporates within its framework what later becomes the focus of goal orientation research in educational settings (McInerney & Van Etten, 2002). Specifically, while much achievement goal theory research in the last two decades is concerned with dimensions of mastery and performance goal orientations and their effects on educational behavior, PI theory is the only theory from its inception to integrate social goal orientations, as well as beliefs about self and thoughts about situations as possible motivational determinants of behavior (Maehr & McInerney, 2004). Only recently are social goals considered integral in educational achievement research and attempts made to broaden the dichotomous mastery and performance goal orientations (Blumenfeld, 1992; Wentzel, 1993, 1999).

Based on Maehr's (1984; Maehr & Braskamp, 1986) theory, McInerney and his colleagues (McInerney, Marsh & Yeung, 2003) proposed a hierarchical and multidimensional model of goal orientations designed to reflect a set of goals relevant in educational settings of various sociocultural contexts. The Inventory of School Motivation (ISM; McInerney & Sinclair, 1991) was developed to assess the constructs

relevant to the model. The hierarchical structure of the model consists of eight specific goals at the base (i.e. task involvement, effort, competition, social power, social concern, affiliation, praise, and token rewards), four more general goals in the middle (i.e. mastery, performance, social, and extrinsic), and general motivation at the peak (Ali & McInerney, 2005).

Validation Studies about ISM

The Inventory of School Motivation (ISM; McInerney & Sinclair, 1991) was composed of a number of dimensions drawn from goal theory and self-concept theory. It was designed to describe individual and group motivational characteristics to compare and contrast groups. Further, it was intended to explain outcome variables, such as performance and learning strategies, and/or to predict future learning outcomes of groups with various characteristics. The ISM was originally created based on a range of motivation constructs drawn from Maehr's PI model (McInerney, 1992; McInerney, Roche, McInerney & Marsh, 1997; McInerney & Sinclair, 1991), which are important in educational settings across diverse groups. These motivation constructs included motivational goal orientations, sense of self, and perceived opportunities or action possibilities. McInerney's earlier work has been focused on validating factor structure of the ISM and using the derived ISM factor scales to predict a variety of school-related outcomes such as students' confidence in school, perceived value of school, and absenteeism etc. (McInerney, 1995; McInerney, Roche, McInerney & Marsh, 1997; McInerney & Swisher, 1995). The original ISM was revised to take account of the earlier analyses (as reported in McInerney, Marsh & Yeung, 2003; McInerney, Roche, McInerney & Marsh, 1997; McInerney, Yeung & McInerney, 2001). Empirical evidence

from both exploratory and confirmatory factor analytic studies supports that the ISM is reliable and valid (see McInerney, 1992, 1995; McInerney, Marsh & Yeung, 2003; McInerney, Roche, McInerney & Marsh, 1997; McInerney & Sinclair, 1991; McInerney, Yeung & McInerney, 2001). A detailed review of literature of these empirical studies is presented below.

McInerney and Swisher (1995) conducted a validation study of the ISM using Exploratory Factor Analysis. Through the analysis, they reduced the original item pool from 100 to 61 items yielding ten interpretable factors. All the items were saliently loaded on their factors with factor loadings exceeding .30. Tabachnick and Fidell (2001) recommended as a general rule .32 for the minimum loading of an item. These ten factors were striving for excellence, sense of competence, recognition, social concern, affiliation, group leadership, sense of purpose for the future, sense of purpose for schooling, competition and, task involvement. The majority of the scales had acceptable reliability coefficients of .70 or above.

Recent research conducted (McInerney, Marsh & Yeung, 2003) on student motivation to examine the multidimensional structure of achievement goal orientation measured by the ISM retained 35 items in a CFA model on 10 goal orientation constructs (effort, task, sense of purpose, praise, competition, power, token, social concern, social dependence, and affiliation). The results supported a multidimensional school motivation construct. The factor structure was well defined in that the goodness of fit indices were good (e.g. RMSEA= .04, RNI= .92, TLI= .91). The results also provided evidence for the hierarchical nature of goals. Three second-order goal factors were generated, including mastery, performance and social goals.

Besides the within-cultural construct validation, McInerney et al. (1997) aimed to determine whether the goals held by students from diverse cultural backgrounds being schooled within Western countries differ and to determine the relationship of these goals to school motivation achievement. For this purpose, they used confirmatory factor analyses to demonstrate the applicability and relevance of the multiple goals and sense-of-self dimensions to Australian Aboriginal (n= 496), Anglo Australian (n= 1173), immigrant Australian (n= 487), and Navajo Indian (529) subjects. Based on confirmatory factor analysis, task and striving for excellence were shown indistinguishable and therefore combined to form one scale measuring task effort. In general, construct validity evidences of ISM (with 40 items for multiple goal components and 23 items for sense-of-self component) are found in these cultural groups. However, the result of ANOVA does not support the traditional stereotypical view of western and non-western groups. It is not supported, in particular, that the Western groups are more task oriented, competitive, and power seeking, whereas the indigenous groups are more affiliative, socially concerned, noncompetitive, and non-power-seeking. Urbanization of children from traditional backgrounds through schooling and mass media was used to partially explain the difference.

McInerney et al. (2001) aimed to validate the motivation orientation scales of the ISM across Navajo (n= 760) and Anglo (n= 1012) high school students in the U.S. Confirmatory factor analysis supported the eight-factor structure of motivation orientations for the total sample and the Navajo and Anglo subsamples, although Navajo students did not distinguish well between the Effort and Task constructs. In the later multi-group Confirmatory factor analysis of initial 39 items; the model of equality

constraint of factor loadings did not fit the data very well. Further, none of the items for the Task factor was invariant across groups and therefore were removed, which resulted in seven factors. Without showing the procedures, McInerney, Yeung, and McInerney (2000) finally reduced the number of items to 30, with which the model of equality constraints fit the data. McInerney et al. suggested meaningful cross-cultural comparisons should use the 30 items that have similar meanings to both cultural groups even though the ISM Motivation Orientation scales are applicable to students of different cultural backgrounds. In addition, the study found that Anglo students in Australia and the Anglo students in the U.S. could be vastly different in their motivation orientation based on the comparison to the results obtained in the McInerney's et al. (1997) study.

Using confirmatory factor analysis, Ali and McInerney (2005) examined the cross-cultural generalizability of the factor structure for the Inventory of School Motivation using 43 items. The sample was high school students drawn from seven cultures, including Anglo-Australian (n= 2616), Migrant Australian (n= 1265), Aboriginal Australian (n= 906), Hong Kong Chinese (n=697), Navajo (n= 1776), Anglo-American (n= 884) and African (n= 819) cultural groups. The data used in this study come from a large data pool derived from a series of longitudinal studies conducted by McInerney and his associates (McInerney, 1995; McInerney, Roche, McInerney & Marsh, 1997; Watkins, McInerney & Lee, 2002). The hierarchical invariance model tests were performed, including the models with fixed number of factors, with equal factor loadings, and with equal factor loadings and equal factor variance/covariances. Ali and McInerney (2005) found that "the items in most of the goal orientation scales of the ISM are mostly invariant groups such that they are probably applicable to both Western and non-Western

cultures" (p. 9). Therefore, the findings seem to support that the ISM scales are applicable to students of cultural backgrounds involved. However, Ali and McInerney also caution against the direct comparison of some items for the Hong Kong Chinese and African students because these two groups do not appear to interpret those items (e.g. items in the Social Power and Token subscales) within the scale in an identical way. Apparently, more invariance tests of ISM are necessary before cross-cultural comparisons of achievement goal orientations.

The above studies discussed apply the first-order confirmatory factor analysis as the methodological tool. McInerney et al. (2003) tested the hierarchical goal theory model of school motivation with 774 Australian high school students. The measure used was the motivational goal orientation component in the ISM. Instead of the regular eight constructs, ten goal orientation constructs with 35 items were applied, including effort, task, sense of purpose, praise, competition, power, token, social concern, social dependence, and affiliation. The results suggest that there is relatively strong support for the hierarchical structure of goals. The three higher-order goals (Mastery goal, Performance goal and Social goal) appear salient, and there is support, although weak, for the third-factor motivation, General Motivation. The corresponding constructs loading on the three higher-order goals are effort, task and sense of purpose on Mastery Goal factor; praise, competition, power and token on Performance Goal factor; and social concern, social dependence and affiliation on Social Goal factor.

The above has reviewed the conceptual development of achievement goal orientation framework from a bipolar continuum with mastery goals at one end and performance goals at the other end, to trichotomous with performance goals divided into

performance approach goals and performance avoidance goals, then to 2 (mastery vs. performance goals) x 2 (approach goals vs. avoidance goals) divisions. The chapter has also reviewed the limitations of these conceptualizations of not including social goals as important motivational goals in academic settings. Then a few issues of achievement goal measures are discussed, including single-item formats, inappropriate domain specificity levels, target population and insufficient construct validation evidence. Illustrations using specific achievement goal measures are presented. The next section will review issues related to the particular methodology to be used in this study.

Structural Equation Modeling and Confirmatory Factor Analysis

Structural Equation Modeling (SEM) refers to a general approach of multivariate data analysis that models the relations between observed and latent variables. SEM is the multivariate data analysis method that has undergone the most refinement and extension over the years and has continued to be developed (Hershberger, 2003). In comparison to the traditional regression analysis, ANOVA or MANOVA, SEM has the advantage of taking the measurement error into consideration while comparing group differences.

A full-blown SEM is composed of a measurement model and a structural model. The measurement model refers to how underlying latent variables are defined by observed variables and the structural model refers to the relations on the latent variables (Byrne, 1998). CFA is used to build the measurement part. Different from Exploratory factor analysis, CFA tests a set of *a priori* hypotheses concerning the constructs and observed variables. Hypothesized models that define the relationships among variables can be tested simultaneously and the statistical fit of the model to the observed data can be evaluated using a variety of goodness-of-fit indices.

In confirmatory factor analysis, model fit indices are used to test the fit between the theoretical model and the empirical data. The model fit is judged based on parameters like the χ^2 goodness-of-fit statistic, RMSEA, SRMR and CFI. The χ^2 statistic assesses the magnitude of discrepancy between the sample and fitted covariance matrices and it indicates how well the model tested fits the data. When it is significant, it means that the model does not fit the data, which usually is not desirable. The use of the χ^2 statistic alone is problematic because it is sensitive to sample size and to deviation from multivariate normality in the item responses. When the sample size is large, the χ^2 statistic is always significant because the sample and fitted covariance matrices are never exactly the same (Rensvold & Cheung, 1998). However, comparative fit index (CFI) is not sensitive to sample size and therefore it is used together with the χ^2 statistic. Nevertheless, instead of other index differences, the χ^2 difference is used when comparing constrained and unconstrained models in invariance testing due to the fact that the traditional statistical significance testing can be used to determine whether the differences should be attributed to sampling error (Rensvold & Cheung, 1998). Root mean square error of approximation (RMSEA) reflects the degree of incongruity for a model per degree of freedom. Standardized root mean squared residual (SRMR) is the square root of the mean of the squared discrepancies between the predicted and observed covariance matrices for the standardized observed variables. The cut-off point used for RMSEA was .08 (Browne & Cudeck, 1993), SRMR was .08 (Hu & Bentler, 1995), and CFI was .95 (Hu & Bentler, 1995). When all three criteria are met, the model is regarded as an adequate fit to the data. A less strict standard is also used to judge model fit: the cut-off point for RMSEA – .08, SRMR – .08, and CFI – .90. Cheung and Rensvold (2002) showed that the number of

items per factor and the number of factors in the model affects most of goodness-of-fit indices. They suggested that when judging model fit based on some generally accepted criteria, model complexity should be taken into consideration. The more complex models are expected to yield smaller goodness-of-fit indices.

Measurement Invariance

SEM is a burgeoning statistical methodology that has distinct utility in addressing the need for culture and group membership in measurement to be examined. It can assess whether individuals from different cultures interpret a measure in the same conceptual manner and if there are gender, ethnic, or other individual differences that affect individuals from responding to a measure in the same ways (Vandenberg & Lance, 2000). This could be accomplished by applying multiple-group CFA. The use of multiple-group CFA has allowed an increasing number of studies to provide evidence of the measurement invariance for instruments across groups (Vijver & Leung, 1997).

Measurement invariance refers to the idea that "under different conditions of observing and studying phenomena, measurements yield measures of the same attributes" (Horn & McArdle, 1992, p. 117). It indicates that "(a) the constructs are generalizeable to each sociocultural context, (b) sources of bias and error (e.g., culture bias, translation errors, varying conditions of administration) are minimal, (c) cultural differences have not differentially affected the constructs underlying measurement characteristics" (Little, 1997, p. 56). If cross-group differences are established, but measurement invariance has not been explicitly evaluated, there is no way to pinpoint the source of the difference unambiguously. Differences could be due to attitudinal variation or as a function of the psychometric properties of the particular items that were administered. Therefore,

Vandenberg and Lance (2000) point out, "if not tested, violations of measurement equivalence assumptions are as threatening to substantive interpretations as is an inability to demonstrate reliability and validity" (p. 6).

Researchers should be cautious about the direct cross-cultural mean comparisons of item- or scale-level observed scores. Translated versions of instruments are often created to extend the use of an instrument in populations who are fluent in other languages. An instrument that is translated into alternate languages may produce variable results unless appropriate measures are taken to ensure that translated versions are comparable (Wild et al., 2005). Items translated into another language may not have the same meaning as originally intended. Poor translations due to inadequate procedures or no equivalent in the target language can result in metric nonequivalence (Robert, Lee & Chan, 2006). Differences in score distributions among versions of an instrument could indicate true differences in the attitudes and beliefs among individuals or reflect facets of the instrument that was used to measure the construct (Behling & Law, 2000). These two implications are quite distinct from one another – the former indicates the presence of actual variation across groups, while the latter reflects aspects of the measurement instrument that was used to obtain the data. In addition, while comparing how individuals from different cultures choose their responses, Chen, Lee, and Stevenson (1995) found that American students more often used the extreme values of the rating scale while Asian students preferred the midpoints more, suggesting that there are differences in calibration across cultures. So when measurement invariance is able to be established, it indicates that biases and errors mentioned above are at minimum.

Invariance testing performed using a structural equation modeling (SEM) framework is well suited to addressing the need to examine measurement instruments as a function of culture and shared belief systems. Sequential steps of invariance test are involved in testing measurement equivalence. According to Steenkamp and Baumgartner (1998), the requirement of the minimum level of invariance varies depending on the goals of cross-cultural studies. For the study to explore whether a construct can be conceptualized the same way cross-culturally, the minimum requirement is that the same pattern of factor loadings is found (Horn, McArdle & Mason, 1983). Metric invariance is desirable but not strictly necessary (Steenkamp & Baumgartner, 1998). That is, a significant factor loading for an item cross-culturally is adequate to show the item is related to the underlying construct. The magnitude of the loading does not have to be the same. For the study to make a quantitative comparison of means cross-culturally, metric and scalar invariance is necessary for at least two items per factor in the multidimensional scale (Meredith, 1993). For the cross-cultural study to examine the structural relationships of one construct with other constructs, full or partial metric invariance but not scalar invariance are required (Steenkamp & Baumgartner, 1998).

Configural Invariance refers to that the same pattern of fixed and free factor loadings of the items on their corresponding factors fits the data well in all cultures. Before the configural invariance test, the factor for each subscale has to be scaled. For scaling purposes, one factor loading in each subscale will be set to 1.0 in both groups to scale the latent variable. The item whose loading is set to 1.0 is called a reference indicator. Any item can be used as the reference indicator when testing for configural invariance; however, the use of different reference indicators will lead to different results

when testing for invariance at the item level (Cheung & Rensvold, 1999). Configural invariance does not imply equality of constraints across parameters (Mullen, 1995). Full or partial configural invariance is required before proceeding to the higher level of invariance (Steenkamp & Baumgartner, 1998). Partial configural invariance is less stringent than full configural invariance in the sense that some constraints are freed up. It is normally considered as the baseline model against which a higher level of invariance, metric invariance, is evaluated.

Metric Invariance refers to the idea that the equality constraints of factor loadings fit the data well cross-culturally. This is analogous to measuring height cross-culturally – we are using 'meter' as the unit of scale in both culture A and culture B. When factor loadings are not equal, it is similar to using 'meter' in culture A but 'foot' in culture B to measure height. The satisfaction of metric invariance means that item-level score differences cross-culturally can be traced to the differences in underlying construct. Full metric invariance is a worthy ideal but unrealistic (Horn, 1991). At least one item other than the one used to scale the latent variable must be metrically invariant in order to precede measurement invariance analysis (Byrne, Shavelson & Muthen, 1989; Steenkamp & Baumgartner, 1998). Modification indices (above 10) can be used to determine which cross-group equality constraint most significantly contribute to the lack of fit. Freeing that constraint can lead to a better fit. Metric Invariance acts as the baseline model against which the next step, Scalar Invariance, is evaluated. Its invariance shows evidence that the factors have the same meaning across groups.

In testing Scalar Invariance, item intercepts are constrained to be equal across groups in addition to their factor loadings. According to Gregorich (2006), item intercepts

reflect the "systematic, additive influences on responses to corresponding items that are constant in each group and are unrelated to the common factors" (p. 6). When intercepts are equal, it means that the origin of the scale is identical across groups. For example, suppose we are using a 1-5 scale with '1' representing 'strongly disagree', '2' 'disagree', and so on. When the intercept differs, it means that in one group '2' may represent 'strongly disagree' rather than 'disagree'. Therefore, this test can be used as a test for cultural driven response sets or additive response bias (Bollen, 1989; Cheung & Rensvold, 2000; Gregorich, 2006; Robert, Lee & Chan, 2006). Such cultural characteristics as social desirability, acquiescence, evasiveness, or humility may influence respondent scoring (Mullen, 1995). It is used to test whether the two groups use the response scale in a similar way (Campbell, Barry, Joe & Finney, 2008). In addition, this test can be used as a test for another potential culture-based nonequivalent source, frame-of-reference (Heine, Lehman, Peng & Greenholtz, 2002; Robert, Lee & Chan, 2006). Frame-of-reference refers to the norm groups against which individuals are comparing when judging their standing on an item. Robert et al. (2006) suggest that culture could influence individuals' item endorsement because of the frame-of-reference. However, this threat of scalar invariance is most likely a downside of using Likert type of scale rather than a culture issue.

Scalar invariance is required for cross-cultural mean comparisons of constructs. Normally only those items with metric invariance are tested for scalar invariance. Scalar invariance implies that cross-cultural differences in means of the items are due to the mean differences in the underlying construct (Gregorich, 2006; Steenkamp & Baumgartner, 1998). According to Cheung and Rensvold (2002), the items with

intercepts freely estimated are excluded automatically from the estimation of latent means. In comparing group mean differences in constructs of interest, analysis of mean and covariance structures (MACS) has to be used, where mean-level information is analyzed in addition to the typical variance-covariance information (Little, 1997). Mean differences can be obtained by setting intercepts for the reference group to zero while without constraining the intercepts for the other group.

Measurement invariance can be tested at higher levels such as error variance, covariance between factors and common variance etc. As stated earlier, however, scalar invariance is sufficient for construct mean comparisons across groups. The equality constraints are becoming more stringent as the invariance test proceeds. χ^2 difference is used as an indicator of measurement invariance when comparing nested models (e.g. model with equality constraints of factor loadings is nested in the model without the constraints). When the χ^2 test is statistically significant, the less parsimonious model (i.e. the model with fewer constraints) is preferred; otherwise, the more parsimonious model is preferred. Therefore, in the sequential invariance tests, a nonsignificant χ^2 difference test is desirable as it indicates the constraints do not decrease the model fit. However, same as the usual χ^2 test, the χ^2 difference test is sensitive to the sample size. When the sample size is large, null hypothesis is easily rejected. Therefore, besides the χ^2 difference test, other goodness-of-fit statistics will be used as well. For the current study, Δ CFI will be used additionally to evaluate the relative good fit of different models. A value of Δ CFI smaller than or equal to -0.01 indicates that the null hypothesis of invariance should not be rejected (Cheung & Rensvold, 2002).

Chapter Summary

This chapter outlines the definition of achievement goal theory and its mainstream operationalization develops from one-dimensional model, to two, three, and currently four-dimensional model. The chapter reviewed the limitations with achievement goal theories. The one-dimensional operationalization causes a lot of problems, which considers achievement goal orientations as a continuum with mastery and performance goals at the two extremes. The second issue of mainstream achievement goal theories is exclusion of social goals. Next, the issues with achievement goal measures (e.g. single-item instrument, domain specificity, and restricted target population) were discussed and several popular measures were presented as illustration. Next, the theoretical framework, Maher's Personal Investment model, was reviewed. Maher's model appears to have addressed the limitations of mainstream goal theories, which makes it promising for cross-cultural studies. Developed based on this framework, Inventory of School Motivation is the focus of this study. As a consequence, its development and validation was reviewed in details. At the end, quantitative research methodologies, CFA and SEM, and the hierarchical measurement equivalence tests were reviewed.

Chapter III

Method

The purpose of this chapter is to five-fold. First, it is to describe the characteristics of the sample of participants from the two cultural groups as well as the medium through which surveys were conducted. Second, the survey instrument, Inventory of School Motivation (ISM), is described, including its hierarchical structure, definition of each domain, and number of items in each domain. Next, translation procedures of ISM from English to Chinese are described. Finally, this chapter discusses the statistical strategies used in the present research to fulfill the objectives posed in the first chapter.

Participants

The target population of this study was college undergraduate students in the United States and in China. Convenience sampling techniques were used to recruit students. Surveys were administered online in both countries.

The sample of the participants in the U.S. was recruited at a large land-grant Midwestern university. Most participants were from the Midwest U.S. who attended a four-year or above, public, comprehensive, research university with high undergraduate enrollment (The Carnegie Foundation, 2006). Participants were recruited from undergraduate English courses, a biology course, and psychology courses through an online subject pool management system. They were general education courses on campus which were offered to students from all disciplines. Most of the

participants (89.5%) were within the range of 18-22 years of age.

Participants in China were recruited in an attempt to ensure that they were similar to their U.S. counterparts. In doing so, a few characteristics were considered, including intensity of research, university type (i.e. private or public), students' region of origin within the country, undergraduate enrollment, and major disciplines, and selectivity of students. Two universities in the Southern China were selected and permission was granted from the directors in charge of student affairs to recruit participants from their universities. One university was engineering and technology oriented and the other one was comprehensive in disciplines. These two public universities were research universities and, in combination, offered disciplines varying in selectivity. Their students were mainly from the Southern China. The participants in this study were recruited from College English classes, a general education class, with the help of colleagues known to the researcher who were college English instructors. Most (89.1%) of the participants' age was within the range of 18-22 years.

Procedures

Two online surveys, one in English and the other in Chinese, were created. These two surveys were equivalent in both format and content. In the U.S., an email invitation was sent to the directors of English Composition and Biology courses in a large Midwestern university asking them to help recruit the students. A brief description of the study, a Web link, http://frontpage.okstate.edu/coe/lihuaxu1, to the online survey, and research compliance information (Appendix C-1E) were provided in the invitation. In addition, participants were recruited from SONA system. Likewise, in China, general education instructors in the two universities were emailed the survey invitation

(Appendix C-1C), with the Web link included. Then instructors passed along the email invitation to their students.

Both the American and Chinese participants were directed to an index page, which gave an introduction and informed them of the voluntary nature of the survey. Those who agreed to complete the survey chose a hyperlink to the survey instrument from the index page (Appendices D-E and D-C). However, those individuals who declined to take the survey were directed to a declination page by a hyperlink. Those participants who successfully submitted their responses after completing the survey were directed to a confirmation page that expressed the researcher's appreciation.

Two follow-up email invitations were sent after the introductory email. These emails reminded them that an online survey about school motivation was being conducted and their participation was appreciated (Appendices C-2E and C-2C).

Instrument

Inventory of School Motivation. The Inventory of School Motivation (ISM; McInerney, Roche, McInerney & Marsh, 1997; McInerney & Sinclair, 1991; McInerney, Yeung & McInerney, 2001) was used in this study. ISM items were developed using the Personal Investment theoretical model as a guide and items from Inventory of Personal Investment (Maehr & Braskamp, 1986) as a base to reflect an educational context and to investigate the school motivation in cross-cultural settings. There are 43 items in this instrument, measuring eight perceived school motivation goals underlying four general goal orientations:

Task (Mastery): Task involvement and Effort

Ego (Performance): Competition and Social power

Social Solidarity: Affiliation and Social concern

Extrinsic rewards: Praise and Token rewards

Task Involvement (4 items): These items measured the amount to which students are totally involved in what they are doing. They regard the study as exciting and fun, and they enjoy adventure and novelty. To them, successful people like challenges and like to solve problems. They feel satisfied and positive about themselves when they accomplish something others could not do, when they understand something for the first time, and when they are responsible for their accomplishments. Therefore, *Task Involvement* measures the student's interest in the task of learning and wanting to improve understanding.

Example: I try harder with interesting work.

Effort (7 items): Referred that students continually thinking of ways to improve themselves, spending long hours of study doing a good job, competition with themselves, not minding studying when others are having fun. They value study and take pride in their study.

Example: I try hard to make sure that I am good at my schoolwork.

Competition (6 items): These items measured the amount that students gain satisfaction and feel positive about themselves when they win in learning situations. They don't feel bad when they beat someone in competition, and they are not afraid of competition even when it is strong.

Example: I want to do well at school to be better than my classmates

Social power (6 items): These items measured the amount that students' desire to have social status in the class.

Example: I often try to be the leader of a group.

Affiliation (3 items): These items measured the importance for students to work with others in groups. Students do their best work when they work with others. They prefer studying with others rather than alone.

Example: I try to work with friends as much as possible at school.

Social concern (5 items): These items measured the concern that students have for the welfare of his/her peers. Students gain satisfaction by sacrificing personal gains for others.

Example: It is very important for students to help each other at school.

Praise (5 items): These items measured the importance for students to receive verbal praise from others. Students study harder so that they can receive respect from their peers, and they do their best work when parents or teachers encourage them and tell them they did well.

Example: Having other people tell me that I did well is important to me.

Token (7 items): These items measured the importance for students to receive tangible rewards for achievement. Students work hard in order to get rewards from their parents and/or teachers. Rewards motivate them to study harder at school.

Example: Getting merit certificates would make me work harder at school.

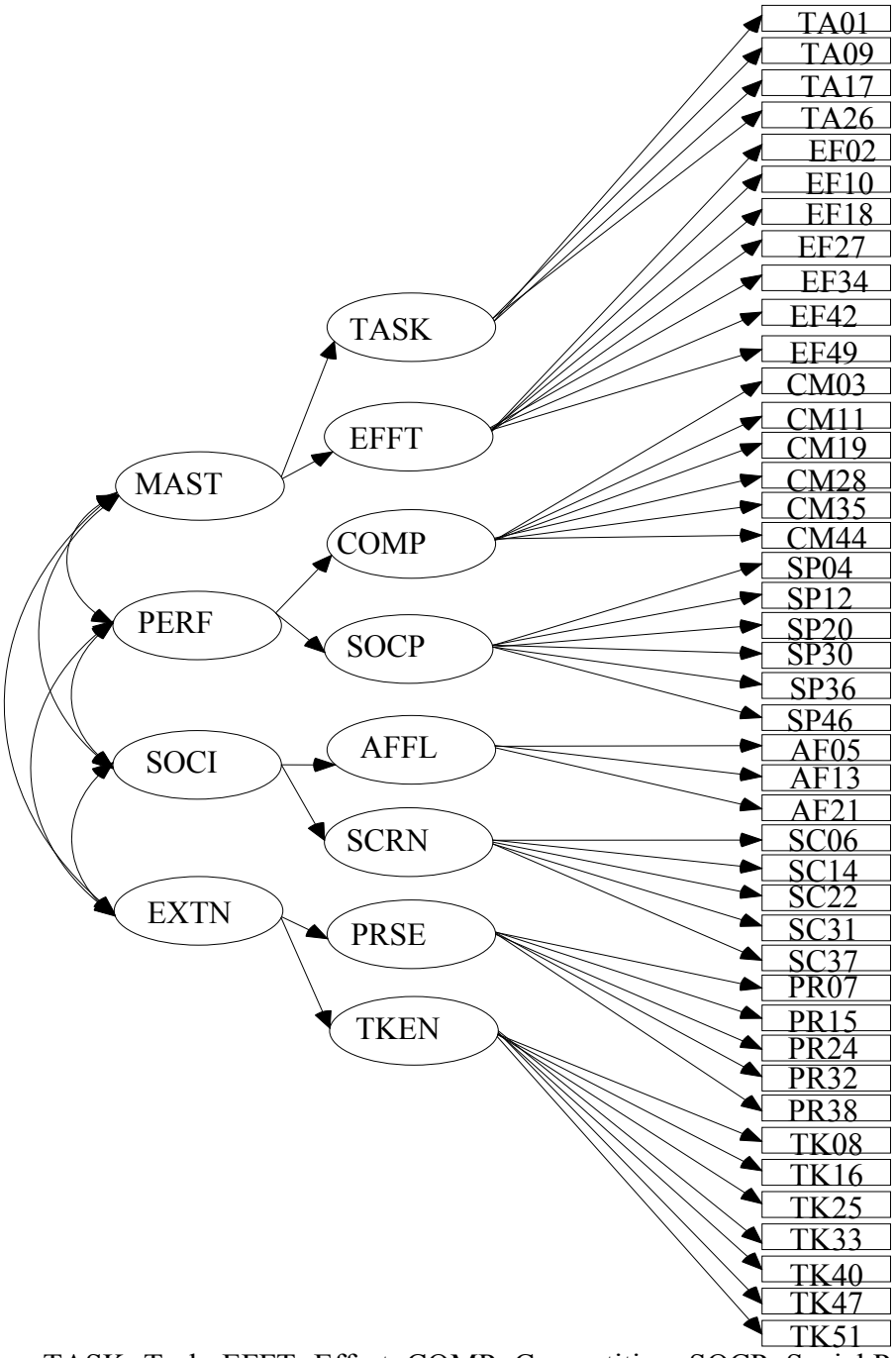
Items in ISM (McInerney & Sinclair, 1991) were answered using a Likert-type scale from strongly agree (1) to strongly disagree (5). To reduce the possibility of error and conform to traditional use of directions in the Likert scale, in this study the original 5-point Likert-scale was reversed. As a result, students respond to each item on a 5-point

Likert-scale (1= *strongly disagree* to 5= *strongly agree*). The responses to the items are coded in the way so that higher scores reflect higher levels of motivation.

Demographic information. Participants were asked their age, gender, race/ethnicity, classification and major in college, whether English/Chinese is their first language, place of original residence, and their parents' education and annual income.

Across the studies using ISM, the number of items included in each dimension varies from study to study, task (3-9 items: task and effort collapse to produce 9 items); effort (3-7 items); praise (4-8 items); competition (4-6 items); power (3-6 items); token (4-7 items); social concern (4-5 items); affiliation (3-4 items) (Ali & McInerney, 2005; McInerney, Marsh & Yeung, 2003; McInerney, Yeung & McInerney, 2001). The range of internal consistency reliability reported for these 8 dimensions are respectively .64- .84 (again .84 reported for the collapsed task and effort dimensions), .68- .81, .76- .83, .68- .78, .68- .80, .70- .79, .67- .72, .68 -.74. In general, the more items in a dimension, the higher the internal consistency reliability.

Figure 1. Theoretical Structure for the ISM.



Note. TASK=Task; EFFT=Effort; COMP=Competition; SOCP=Social Power; AFFL=Affiliation; SCRN=Social Concern; PRSE=praise; TKEN=Token; MAST=Mastery; PERF=Performance; SOCI=Social; EXTN=Extrinsic.

Instrumentation Translation

To the researcher's best knowledge, there was no Mandarin Chinese version of the ISM instrument available. Therefore, a Mandarin Chinese version of the ISM was developed for Chinese participants. In an attempt to obtain linguistic equivalence and translation equivalence, the translation and evaluation procedure was adapted from Vallererand's (1989) forward-backward translation process, Brislin's (1970) backward translation process, and Sperber, Devellis, and Boehlocke's (1994) instrument evaluation process. The following were the translation procedures used by this researcher for the Chinese version of Inventory of School Motivation.

1. *The forward translation of ISM from English to Chinese was performed by two bilingual translators.* Both translators achieved their Bachelor's and Master's degree in English and were experienced translators. To ensure the Chinese version was as closely equivalent as possible to the original English version, both forms and meanings of the items were emphasized. But meanings were given priority if both forms and meanings could not be maintained at the same time. For example, item *I like being given the chance to do something again to make it better* was translated into *I like doing the same thing one more time if given the opportunity, so that I can do better*. The two versions were considered to be the same in meaning but not in form. The translators did the translations independently. Afterwards, the two Chinese translations together with the original English version were given to two other equivalently experienced bilingual judges to check for errors and for their comments and suggestions. This method made use of the bilingual

judges "who compare[d] the source and translated versions of each test item and decide[d] whether any differences between translations could result in non-equivalence of meaning in the two populations of interest" (Brislin, 1970 cited in Hambleton & Bollwark, 1991, p. 16). Valuable suggestions were given so that disagreement was solved and consensus on the Chinese version was reached. One Chinese item was found to be ambiguous. It could be understood either as *I like to learn from others* or as *I like to study with others*. Therefore, the Chinese translation was reworded to eliminate the ambiguity. The judges' comments made the Chinese version more authentic.

2. *The Chinese version obtained from the forward translation was translated back into English by three other bilinguals.* This step is called backward translation. All three translators were native Chinese who achieved their Bachelor's and Master's degrees in English and were currently pursuing their doctoral degrees in the United States. None of these translators had ever seen the original English version of items. Consensus was reached through discussion.
3. *The two English versions of the items were evaluated by a group of seven monolinguals of English in terms of comparability of language and similarity of interpretability.* Comparability of language assesses the similarity of words, phrases, and sentences while similarity of interpretability evaluates the similarity of an item's meaning. Both were measured on a 7-point Likert scale (1= *Extremely comparable/ similar*; 7= *Not at all comparable/similar*) (Sperber, Devellis & Boehlocke, 1994). Items with mean scores of more than

3 were given back to the forward translators and then backward translators for further refinement. As a result, four items were retranslated and reassessed for their equivalency.

Analysis Procedures

The following is the restatement of research purposes listed in the Introduction chapter. After each purpose is the corresponding analysis plan. Objective I: To test whether the construct of achievement goal orientations as operationalized by McInerney's Inventory of Student Motivation is comparable between two cultures. That is, the ability of ISM items to form eight *a priori* factors would be tested using the American and Chinese subsamples. If the same factor structure holds for each culture, the next step was to test whether factor loadings and intercepts also were equivalent across the samples.

To fulfill this objective, both confirmatory factor analysis and multi-group confirmatory factor analysis methods were used. Specifically, a single group CFA was conducted to evaluate whether the eight-factor achievement goal model held for both cultures. The model fit was judged based on parameters like the χ^2 goodness-of-fit statistic, RMSEA, SRMR and CFI. The cut-off criterion used for RMSEA was .08 (Browne & Cudeck, 1993), SRMR was .08 (Hu & Bentler, 1995), and CFI was .95 (Hu & Bentler, 1995). When all three criteria were met, the model was regarded as an adequate fit to the data. A less strict standard was also used to judge model fit: the cut-off point for RMSEA was .08, SRMR, .08, and CFI was .90. Assuming findings of model tenability for each group, cross cultural measurement invariance test of ISM was conducted. That is, multi-group CFA was used to test the configural, metric, and scalar equivalency of eight

subscales across the cultures. Nested models were directly compared using the χ^2 difference test and CFI differences.

Objective II: To test whether there are mean differences in the eight factors (e.g., Task involvement, Effort, Competition etc.) across the cultures, given that factor comparability was first established.

This was accomplished by using the Mean and Covariance Structures analysis (MACS). The latent mean on a construct could not be directly estimated, and the between-group difference on latent mean was estimated by fixing the latent mean as 0 in one group. For example, to obtain the estimate of the mean difference between Chinese and American samples in the magnitude of Task goals, the American group was chosen as the referent group by constraining its mean to 0. The mean for the Chinese group was estimated and that estimate reflected the cross-cultural difference in the means of that construct.

Computer Software

Original raw data for American and Chinese samples were stored in EXCEL files. Then data were imported into SPSS16 for data cleaning and preliminary analyses, including descriptive statistics (means and standard deviation). Afterwards, LISREL8.8 for Windows (Jöreskog & Dag, 2006) was used to import SPSS external data as PSF. PRELIS was used to create covariance matrices, standard deviations, and mean matrices for the full eight-factor ISM scale and its subscales for both the American and Chinese samples. These data matrices were used in LISREL8.8 with maximum likelihood estimation to estimate model parameters and fit indices. Confirmatory factor analyses and

multi-group CFA were conducted for measurement equivalence across groups with maximum likelihood estimation procedures.

Chapter Summary

This chapter has outlined the methodology employed in the present research. The sampling and demographic information for participants in the United States and China have been presented. The instrumentation and the relevant translation procedures into Chinese version have been described. This chapter has also outlined the analytical procedures involved in conducting measurement invariance cross-culturally to fulfill the objectives posed in the Introduction chapter. In the next chapter the results will be presented.

Chapter IV

Results

This chapter presents the results of the study. First, the characteristics of the participants in both the American and Chinese culture are presented and compared. Second, psychometric properties of the ISM in the two groups are assessed. Next, Confirmatory Factor Analyses are conducted with the American and Chinese subsamples. Then the results of hierarchical invariance test using multiple-group Confirmatory Factor Analysis are presented. Finally, cross-cultural comparisons of means on the raw scale scores and when appropriate, comparisons on latent means are conducted.

Sample Characteristics

After the data were cleaned up, 343 valid responses were generated in the American sample and 392 were in the Chinese sample (duplicates were removed; subjects with reckless responses were removed, which means they chose the same response for all the items; only those who stated their first language is English/Chinese were kept in the data). Among those, 70.5% of the American participants were male; whereas, only 34.3% of the Chinese participants were male. In terms of age, 89.5% of the American participants and 89.1% of the Chinese participants were at the age of 18-22 years. In the American participants, 39.3% were freshmen, 22.6% sophomores, 24.6% juniors and 13.5% are seniors and above. Similarly, in the Chinese sample, 53.1% of

them were freshmen, 18.5% sophomores, 10.0% juniors, and 18.5% are seniors and graduates (see Figure 2). 65.6% of American participants reported their parents' income as $40,000 and above; only 13.7% of Chinese participants stated their parents' income as ¥40,000 and above (see Figure 3). Regarding origin of residence, 40.3% of the American participants were from the rural area, 44.4% from the suburban area and 15.3% from the urban area; whereas 43.6% of the Chinese participants were from rural, 17.1% from suburban and 39.2% from urban (see Figure 4). Based on those figures, it seems that the American participants were less likely to be freshmen than the Chinese participants, they came from more well-to-do families, and were less urban than the Chinese participants.

Figure 2. Participants' Classification in College

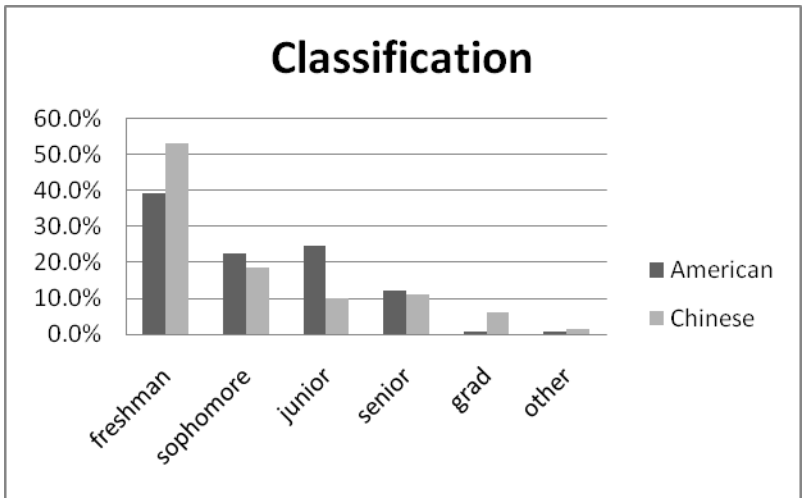


Figure 3. Parents' Income ($ vs. ¥) Stated by the American and Chinese Participants

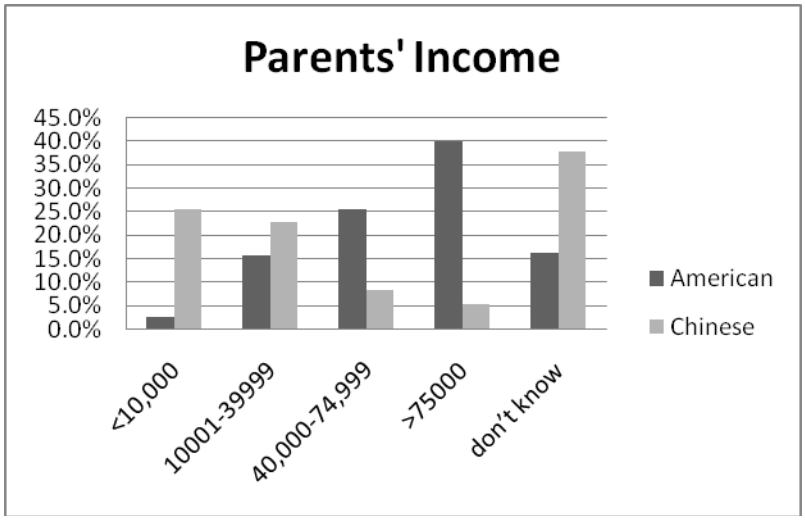
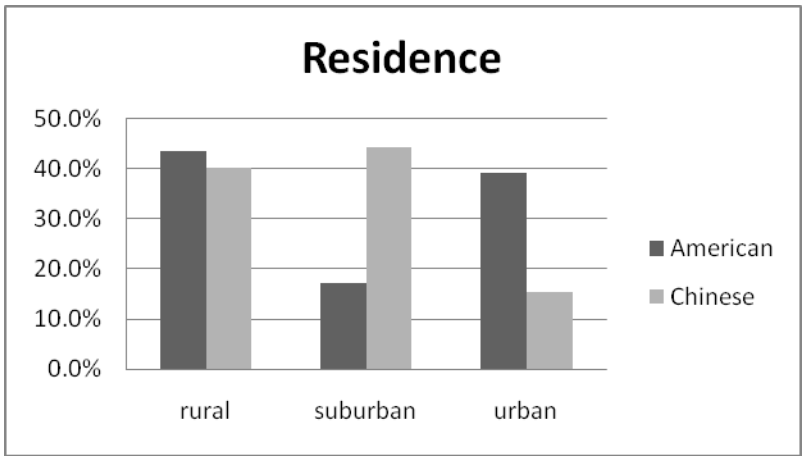


Figure 4. Origin of Residence as Stated by the American and Chinese Participants



Descriptive Statistics

Means, standard deviations, skewness, and kurtosis for each item broken out by culture are presented in Table 1. Judging from the item means of 4 (*Agree*) which was above midpoint (*Neither agree nor disagree*) of the scale, with some variation across the

cultures, participants seemed to endorse the items measuring the Task, Effort, Competition, Praise, Social Concern goal orientations. However, participants seemed to choose the midpoint of most items measuring Affiliation, Social Power, and Token goals. The descriptive statistics for each subscale indicated the same pattern. The mean of the Token subscale for the Chinese group is much higher than the American group (3.32 vs. 2.90), so is that for Competition (3.35 vs. 3.20). The mean of the Praise subscale for the American group is higher than that for the Chinese group (3.60 vs. 3.48).

Table 1

Descriptive Statistics for ISM Items across Cultures

Subscales		American (N = 343)				Chinese (N = 392)			
	Items	Mean	*SD*	S	K	Mean	*SD*	S	K
Task	TA01	4.41	0.694	-1.690	5.301	3.77	0.891	-0.677	0.410
	TA09	4.30	0.756	-1.476	3.514	4.39	0.649	-1.145	3.165
	TA17	4.25	0.720	-1.321	3.630	4.47	0.625	-1.345	4.065
	TA26	4.14	0.673	-0.761	1.803	4.11	0.646	-0.808	2.784
Effort	EF02	4.12	0.813	-1.109	1.770	4.29	0.725	-1.397	3.662
	EF10	4.05	0.779	-0.840	1.190	4.18	0.690	-0.910	2.292
	EF18	4.13	0.825	-1.012	1.341	4.33	0.700	-0.966	1.648
	EF27	3.38	0.968	-0.090	-0.731	3.52	0.863	-0.156	-0.069
	EF34	3.89	0.853	-0.697	0.497	3.47	0.882	-0.213	-0.215
	EF42	4.14	0.683	-0.739	1.570	4.26	0.621	-0.507	0.677
	EF49	4.02	0.751	-0.583	0.314	3.88	0.685	-0.423	0.716
Competition	CM03	3.80	0.973	-0.638	0.126	3.56	0.853	-0.190	0.017
	CM11	3.21	1.099	-0.091	-0.714	2.98	0.919	0.136	-0.002
	CM19	3.26	1.097	-0.029	-0.986	3.70	0.908	-0.655	0.294
	CM28	3.41	1.058	-0.356	-0.672	4.17	0.747	-0.991	2.209
	CM35	2.86	1.082	0.198	-0.783	3.29	0.941	-0.178	-0.303
	CM44	2.71	1.039	0.192	-0.617	2.41	0.996	0.503	0.011
Social Power	SP04	3.09	0.989	0.220	-0.538	2.96	0.980	0.055	-0.376
	SP12	3.35	1.022	-0.371	-0.422	3.08	1.038	-0.129	-0.532
	SP20	3.07	1.088	-0.016	-0.670	3.09	0.920	0.015	-0.204
	SP30	2.88	1.140	0.148	-0.793	2.80	0.924	0.049	-0.117
	SP36	2.38	1.042	0.599	-0.152	2.88	0.965	0.078	-0.333
	SP46	3.01	1.119	0.065	-0.821	2.94	0.903	0.043	-0.171
Affiliation	AF05	3.10	1.086	0.016	-0.728	3.26	0.853	-0.081	-0.080
	AF13	3.31	1.066	-0.315	-0.799	3.35	0.837	-0.213	-0.286
	AF21	2.93	1.147	0.032	-0.876	3.03	0.868	0.088	-0.244
Social Concern	SC06	3.81	0.812	-0.665	0.688	4.46	0.651	-1.356	3.602
	SC14	3.85	0.771	-0.708	0.720	3.56	0.688	-0.303	0.161
	SC22	3.93	0.819	-0.740	0.834	3.97	0.645	-0.492	1.302
	SC31	3.26	1.044	-0.370	-0.711	3.51	0.830	-0.440	0.015
	SC37	3.49	0.912	-0.648	0.015	3.16	0.845	-0.395	0.212
Praise	PR07	3.86	0.871	-0.687	0.479	3.58	0.905	-0.415	0.098
	PR15	3.27	1.012	-0.190	-0.568	3.26	0.902	-0.328	-0.012
	PR24	3.53	0.930	-0.501	-0.118	3.96	0.699	-0.398	0.526
	PR32	3.62	0.954	-0.518	-0.186	3.37	0.903	-0.488	0.127
	PR38	3.72	0.996	-0.770	0.044	3.23	0.929	-0.150	-0.202
Token	TK08	3.38	0.980	-0.325	-0.447	3.15	0.994	-0.013	-0.469
	TK16	3.08	1.029	-0.078	-0.696	3.09	0.910	-0.107	-0.004
	TK25	2.30	1.113	0.683	-0.322	2.96	0.988	-0.041	-0.431
	TK33	3.04	1.043	0.055	-0.584	3.26	0.876	-0.086	-0.249
	TK40	3.10	1.006	-0.003	-0.415	3.85	0.771	-0.521	0.347
	TK47	2.34	1.017	0.628	-0.123	3.04	0.885	0.014	-0.203
	TK51	3.11	1.020	-0.158	-0.514	3.88	0.755	-0.269	-0.259

Note. S = Skewness. K = Kurtosis

Regarding indices of item variability, the observed standard deviations for the American group are generally higher than those for the Chinese group with a few exceptions. This cross-cultural pattern is more obvious in the subscale variability. This seems to be consistent with the observation that Americans tend to choose more extreme ends of a scale but Asians including Chinese tend to choose midpoints of a scale (Chen, Lee & Stevenson, 1995). Within each culture, the subscales of Task, Effort and Social Concern have lower variability than the other subscales. Specifically, the observed standard deviations for the Task, Effort and Social Concern subscales range from .509 to .589 for the American group and from .451 to .471 for the American group. However, the standard deviations for the other subscales are over .700 for the American group and about .600 for the Chinese group.

Because maximum likelihood estimation procedures used in this study can lead to distorted results when normality assumption is violated, the normality indices of each item were evaluated in terms of skewness and kurtosis (Table 1). According to Curran, West, and Finch (1996), for univariate normality, skewness and kurtosis values of 0 to 2, and 0 to 7, respectively, can be taken as demonstrating sufficient normality. Absolute skewness values of all the items in this study fall into the range, with most items below 1 and a few items in the subscales of Task and Effort above 1. Absolute kurtosis values of all the items fall into the range of 0 to 7, among which most items have absolute kurtosis value below 1. The data appear to show sufficient normality.

Psychometric Characteristics of ISM across the Two Samples

Internal consistencies for the subscales across the cultures are presented in Table 2. For the American group, coefficient alpha ranged from .693 to .846. For the Chinese

group, coefficient alpha ranged from .523 to .822. The Task and Social Concern subscales in both cultures have lower than generally recommended coefficient alphas. This could be partly due to the short length of the two subscales (4 and 5 items respectively). In general, the coefficient alphas between the American and Chinese groups are comparable, except for the Task subscale (*i.e.* α = .699 vs. .523).

Despite its widespread popularity, coefficient alpha is known to under-estimate scale reliability when Tau equivalency among items is not established (Brown, 2006). Therefore, composite reliability was calculated in this study according to Raykov's (2004) CFA-based method of estimating scale reliability (see Table 2). As a whole, as compared to composite reliabilities, coefficient alphas seemed to underestimate the reliability of scales, even though by marginal differences.

Table 2

Reliabilities and Composite Reliabilities for 8 Subscales

Domain	American			Chinese		
	Coefficient Alpha	Composite Reliability	N	Coefficient Alpha	Composite Reliability	N
Task (4 items)	.699	.702	337	.523	.572	380
Effort (7 items)	.780	.800	336	.727	.726	371
Competition (6 items)	.807	.806	337	.727	.737	370
Social Power (6 items)	.846	.852	334	.822	.837	375
Affiliation (3 items)	.778	.781	342	.720	.722	382
Social Concern (5 items)	.693	.708	337	.639	.658	383
Praise (5 items)	.820	.827	335	.778	.744	374
Token (7 items)	.833	.838	337	.768	.775	371

Objective I: To test whether the construct of achievement goal orientations as operationalized by McInerney's Inventory of Student Motivation is comparable between the two cultures.

Single-sample Confirmatory Factor Analysis

Confirmatory factor analyses for ISM were conducted to test the measurement model for the two cultural groups. Specifically, the fit of a theoretically implied model was tested towards the data. In this study, the model of eight factors with their corresponding defining items was tested against the covariance matrices of both groups. Chi-square and other model fit indices were used to evaluate the fit. The model fit was estimated using maximum likelihood in LISREL 8.8. Raw data with pairwise deletion of missing data was used.

Model fit indices are presented in Table 3, including chi-square with *p*-value, RMSEA, CFI, and SRMR. For the American group, χ^2 (N = 343, df = 832) = 2669.11, RMSEA = .08 with 90% CI = (.077; .084), SRMR = .083 and CFI = .92. For the Chinese group, χ^2 (N = 392, df = 832) = 2933.27, RMSEA = .08 with 90% CI = (.077; .084), SRMR = .083 and CFI = .90. The model fit indices for the two groups are equal except for the χ^2. Based on the rule-of-thumb cut-off criteria for model fit (Browne & Cudeck, 1993; Hu & Bentler, 1995), the model was considered a border line fit.

Table 3

Tests of Population Heterogeneity of ISM in American and Chinese Students

	χ^2	df	RMSEA (90% CI)	SRMR	CFI
Single Group Solutions					
American (n = 343)	2669.11***	832	.080 (.077; .084)	.083	.92
Chinese (n=392)	2933.27***	832	.080 (.077; .084)	.083	.90

Note. RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval for RMSEA; SRMR = standardized root mean square residual; CFI = comparative fit index. *** $p < .001$

Table 4 presents the unstandardized and standardized loadings of 43 items across the cultures. All the unstandardized loadings for the eight subscales for both groups are

statistically significant. The standardized factor loadings for the subscales of Affiliation, and Praise for both groups range from .50 and above. Few items on the subscales of Task, Effort, Competition, Social Concern, Social Power and Token have standardized factor loadings under .50. The unstandardized factor loadings for the American group were generally higher than those for the Chinese group. However, when the variance of each item was taken into consideration, the magnitude of standardized factor loadings between the Chinese and American groups become more similar. The similarities of factor loadings were compared in detail in the measurement equivalence test section.

Table 4

Unstandardized and Standardized Coefficient for American and Chinese students

Subscales	Items	American		Chinese	
		Unstandardized	Standardized	Unstandardized	Standardized
Task	TA01	0.42	0.61	0.27	0.31
	TA09	0.41	0.54	0.30	0.47
	TA17	0.51	0.70	0.40	0.64
	TA26	0.39	0.58	0.37	0.57
Effort	EF02	0.42	0.51	0.34	0.47
	EF10	0.54	0.69	0.40	0.59
	EF18	0.55	0.66	0.44	0.62
	EF27	0.45	0.47	0.35	0.40
	EF34	0.53	0.62	0.37	0.42
	EF42	0.40	0.59	0.36	0.58
	EF49	0.50	0.67	0.40	0.58
Competition	CM03	0.48	0.50	0.53	0.62
	CM11	0.71	0.65	0.65	0.71
	CM19	0.71	0.65	0.52	0.57
	CM28	0.64	0.60	0.25	0.33
	CM35	0.83	0.77	0.57	0.61
	CM44	0.68	0.66	0.52	0.52
Social Power	SP04	0.62	0.63	0.66	0.67
	SP12	0.42	0.42	0.52	0.50
	SP20	0.91	0.84	0.72	0.78
	SP30	0.96	0.85	0.76	0.82
	SP36	0.54	0.52	0.50	0.52
	SP46	0.98	0.88	0.67	0.75
Social Concern	SC06	0.58	0.71	0.29	0.45
	SC14	0.53	0.69	0.45	0.66
	SC22	0.49	0.60	0.39	0.61
	SC31	0.49	0.47	0.42	0.50
	SC37	0.32	0.36	0.34	0.41
Affiliation	AF05	0.80	0.74	0.49	0.57
	AF13	0.73	0.68	0.58	0.69
	AF21	0.90	0.79	0.67	0.77
Praise	PR07	0.61	0.70	0.58	0.64
	PR15	0.64	0.63	0.64	0.71
	PR24	0.69	0.74	0.38	0.55
	PR32	0.79	0.82	0.58	0.64
	PR38	0.59	0.59	0.63	0.68
Token	TK08	0.63	0.65	0.57	0.57
	TK16	0.76	0.74	0.65	0.72
	TK25	0.64	0.58	0.51	0.52
	TK33	0.87	0.83	0.65	0.74
	TK40	0.41	0.41	0.41	0.54
	TK47	0.70	0.69	0.37	0.42
	TK51	0.65	0.64	0.37	0.49

Table 5 presents the correlations among factors within each culture. The magnitude of correlations varied from low (e.g. .10) to high (e.g. ≈ 1.00) with that of Social Concern and Competition being .10 for the American group and -.01 for the Chinese group.

Table 5

Correlations among Factors within each Cultural Group

		Chinese							
		Task	Efft	Comp	Socp	Affl	Scrn	Prse	Tken
American	Task		0.99	0.40	0.12	0.08	0.43	0.47	0.45
	Efft	0.91		0.50	0.28	0.23	0.58	0.37	0.38
	Comp	0.18	0.22		0.72	0.07	-.0.01	0.68	0.67
	Socp	0.10	0.24	0.65		0.25	0.22	0.55	0.53
	Affl	0.24	0.10	0.32	0.46		0.53	0.25	0.15
	Scrn	0.58	0.52	0.10	0.35	0.73		0.25	0.15
	Prse	0.47	0.35	0.42	0.37	0.32	0.44		0.98
	Tken	0.16	0.06	0.66	0.47	0.46	0.17	0.70	

Note. Efft = Effort, Comp = Competition, Socp = Social Power, Affl = Affiliation, Scrn = Social Concern, Prse = Praise, Tken = Token

As suggested by Brown (2006), a factor correlation of .85 is usually used as a cutoff criterion for detecting problematic discriminant validity. Based on this criterion, the factors Task and Effort, Praise and Token in the Chinese group seemed to have discriminant validity problem, so did the factors Task and Effort in the American group. This could be an indicator that Task and Effort are actually very similar constructs in both samples and so are the constructs of Praise and Token in the Chinese sample.

Hierarchical Measurement Invariance Test across Cultures for each Subscale

Several different aspects of measurement invariance were addressed in this study. First, given that the eight-factor model fit within each cultural group, although borderline yet adequate to proceed, subscale by subscale configural invariance was tested within each culture. Configural invariance restricted the factor structure to be the same across

the groups. It worked as a baseline model for the metric invariance. Then, for subscales that failed to reject the configural invariance, a series of increasingly-stringent nested invariance tests were conducted to examine various aspects of invariance. Metric invariance constrained all factor loadings to be equal across the groups. When the metric invariance model was compared to the configural invariance model and the chi-square difference was statistically significant, some equality constraints on factor loadings were freed up based on the substantive meaning and modification indices. Then this model with fewer equality constraints was compared with the configural invariance. The difference in chi-square was tested for statistical significance. As for the number of invariant items to be present for invariance test to proceed and for the interpretation to be meaningful, it is recommended that normally at least one item other than the reference item have invariant factor loading (Brown, 2006; Steenkamp & Baumgartner, 1998). The next step was, based on the model of metric invariance, the scalar invariance constrained on the intercept terms to be equal for those items that were found to have invariant factor loadings and let the nonequivalent items free. This last step in the step-wise invariance test was not necessary for testing whether the constructs of ISM mean the same thing across the cultures. However, it was conducted as an endeavor to detect the possible differences in item intercepts between the two cultures. These further equality constraints were tested for statistical significance by nested chi-square difference test. If the subscale - level factor loading invariance was achieved, the full scale of ISM would be constructed using the already verified equality constraints and free parameters. The model was compared with the baseline model – the configural invariance model. The equality

constraints imposed was assessed using the nested chi-square differences for statistical significance.

Configural, metric and scalar invariance testing for the subscale Task.

Configural, metric and scalar invariance for Task were tested across the two cultural subgroups, and the test of model fit is presented in Table 6. To scale the latent construct Task, TA17 was used as the reference indicator by fixing factor loading to 1.0 across the groups. Reference indicators in this research were chosen based on a few considerations: substantive meaning, comparability of factor loadings across cultural groups, and magnitude of unstandardized coefficient derived from single-group Confirmatory Factor Analysis (i.e. the closer it approaches 1.0, the better). The model for configural invariance fit the data very well, χ^2 (df = 4) = 11.09, p = .026, RMSEA = .07 with 90% CI = (.022; .120), SRMR = .022 and CFI = .98.

Table 6

Step-wise Invariance Test for Subscale Task across Cultures

Dimension	MI	χ^2	*df*	*p*	RMSEA (90% CI)	CFI	SRMR	$\Delta\chi^2$	Δdf	ΔCFI
Task	Configural Invar.	11.09	4	.026	.070 (.022; .120)	.98	.022			
	Metric Invar.	32.15	7	<.001	.099 (.066; .140)	.94	.067	21.06***	3	-.04
	Partial Metric Invar.	13.45	5	.019	.068 (.025; .110)	.98	.030	2.36	1	.00
	Scalar Invar.	33.92	7	<.001	.100 (.070; .140)	.93	.030	20.47***	1	-.05

Note. *** $p < .001$

Configural invariance was used as the baseline model for metric invariance. There was significant χ^2 difference between the metric invariance and configural invariance, $\Delta\chi^2$ = 21.06 (Δdf = 3) and ΔCFI = -.04. Therefore, partial metric invariance was tested when equality constraints of TA1 and TA9 were freed up. There was non-significant χ^2 difference between the partial metric invariance and configural invariance,

$\Delta\chi^2 = 2.36$ ($\Delta df = 1$) and $\Delta CFI = .00$. A value of ΔCFI smaller than or equal to -0.01 indicates that the null hypothesis of invariance should not be rejected (Cheung & Rensvold, 2002). This indicated that the factor loadings of TA1 and TA9 were different across the two cultures, while those of TA17 and TA26 were the same. This means that the observed item-level score differences of TA17 and TA26 between the American and Chinese participants can be traced to the differences in the latent variable Task. The same interpretation cannot be made for items TA1 and TA9.

With metric invariance established, scalar invariance was tested for the subscale. Other than the reference item TA17, TA26 were constrained equally across the cultures on its intercepts. There was significant χ^2 difference between the scalar invariance and metric invariance, $\Delta\chi^2 = 20.47$ ($\Delta df = 2$) and $\Delta CFI = -.05$. Based on chi-square difference and ΔCFI, the null hypothesis of scalar invariance was rejected. Therefore, results suggest that the subscale Task does not show scalar invariance cross-culturally.

Configural, metric and scalar invariance testing for the subscale Effort. Configural, metric and scalar invariance for Effort were tested across the two cultural subgroups, and model fit statistics are presented in Table 7. To scale the latent construct Effort, EF18 was used as the reference indicator by fixing factor loading to 1.0 across the groups. The model for configural invariance fit the data very well, χ^2 ($df = 28$) = 64.24, $p < .001$, RMSEA = .060 with 90% CI = (.040; .79), SRMR = .04 and CFI = .98.

Table 7

Step-wise Invariance Test for Subscale Effort across Cultures

Dimension	MI	χ^2	*df*	*p*	RMSEA (90% CI)	CFI	SRMR	$\Delta\chi^2$	Δdf	ΔCFI
Effort	Configural Invar.	64.24	28	<.001	.060 (.040; .079)	.98	.040			
	Metric Invar.	75.68	34	<.001	.058 (.040; .075)	.97	.055	11.44	6	-.01
	Partial scalar Invar.	84.76	38	<.001	.058 (.041; .075)	.97	.055	9.08	4	.00

With the configural invariance established, the model for metric invariance for Effort was tested across the two cultural subgroups. There was no significant χ^2 difference between the metric invariance and configural invariance with $\Delta\chi^2 = 11.44$ at $\Delta df = 6$ and $\Delta CFI = -.01$. This invariance indicates that factor loadings of the items measuring Effort were invariant across the two cultures. It means that the observed item-level score differences between the American and Chinese participants can be traced to the differences in the latent variable Effort.

With metric invariance, scalar invariance was tested for the subscale. There was significant χ^2 difference between the scalar invariance and metric invariance, $\Delta\chi^2 = 143.3$ ($\Delta df = 7$) and $\Delta CFI = -.08$. Based on chi-square difference and ΔCFI, the null hypothesis of scalar invariance was rejected. Modification indices were examined to explore the possible cause of misfit. It was suggested that a better model fit could be achieved with releasing the equality constraints of intercept invariance for EF18 (MI = 18.42), EF34 (MI = 73.46) and EF49 (MI = 31.91). Therefore, partial scalar invariance was tested when intercept invariance for EF18, EF34, and EF49 were freed up. There was non-significant χ^2 difference between the partial scalar invariance and metric invariance,

$\Delta\chi^2 = 9.08$ (Δdf = 4) and ΔCFI = .00. As a result, scalar invariance test of subscale Effort failed to reject the null hypothesis that item intercepts were the same across the groups.

Configural invariance testing for the subscale Competition. Configural invariance of Competition within each cultural group and across groups was tested and the model fit indices are presented in Table 8. To scale the latent construct Competition, CM19 was used as the reference indicator by fixing factor loading to 1.0 across the groups. The model fit indices for configural invariance indicated that there was model misspecification, χ^2 (df = 18) = 188.75, $p < .001$, RMSEA = .16 with 90% CI = (.14; .18), SRMR = .068 and CFI = .88. Therefore, the factor structure of Competition was tested within each culture. Results showed that there was model misspecification for both cultural groups. For the American group, χ^2 (df = 9) = 115.48, $p < .001$, RMSEA = .19 with 90% CI = (.16; .22), SRMR = .068 and CFI = .88. For the Chinese group, the model fit indices indicated model misfit, χ^2 (df = 9) = 73.27, $p < .001$, RMSEA = .14 with 90% CI = (.11; .16), SRMR = .068 and CFI = .90. Modification indices for item error terms within each culture suggested the factor structure was very different across the cultures. Further verification of factor structure is necessary using exploratory factor analysis. Therefore, further measurement invariance test was not conducted for this scale.

Table 8

Configural Invariance and Within Group CFA Test for Subscale Competition

Dimension	MI	χ^2	*df*	*p*	RMSEA (90% CI)	CFI	SRMR
Competition	Configural Invar.	188.75	18	<.001	.160 (.140; .180)	.88	.068
	CFA within American	115.48	9	<.001	.190 (.160; .220)	.88	.068
	CFA within Chinese	73.27	9	<.001	.140 (.110; .160)	.90	.068

Configural invariance testing for the subscale Social Power. Configural invariance of Social Power within each cultural group and across groups was tested and the model fit indices are presented in Table 9. To scale the latent construct Social Power, SP46 was used as the reference indicator by fixing factor loading to 1.0 across the groups. The model fit for configural invariance across the cultural groups suggested model misspecification, χ^2 (df = 18) = 151.19, $p < .001$, RMSEA = .14 with 90% CI = (.12; .16), SRMR = .057 and CFI = .95. Therefore, the factor structure of Social Power was tested within each culture. Results showed that there was model misspecification for both cultural groups. For the American group, χ^2 (df = 9) = 92.68, $p < .001$, RMSEA = .17 with 90% CI = (.14; .20), SRMR = .074 and CFI = .93. For the Chinese group, the model fit indices indicated model misfit, χ^2 (df = 9) = 58.50, $p < .001$, RMSEA = .12 with 90% CI = (.091; .095), SRMR = .057 and CFI = .96. Modification indices for item error terms within each culture suggested the factor structure was very different across the cultures. Further verification of factor structure is necessary using exploratory factor analysis. Therefore, further measurement invariance test was not conducted for this scale.

Table 9

Configural Invariance and Within Group CFA Test for Subscale Social Power

Dimension	MI	χ^2	*Df*	*P*	RMSEA (90% CI)	CFI	SRMR
	Configural Invar.	151.19	18	<.001	.140 (.120; .160)	.95	.057
Social Power	CFA within American	92.68	9	<.001	.170 (.140; .200)	.93	.074
	CFA within Chinese	58.50	9	<.001	.120 (.091; .150)	.96	.057

Configural, metric and scalar invariance testing for the subscale Social Concern. Configural, metric and scalar invariance for Social Concern were tested across the two cultural subgroups, and the model fit indices are presented in Table 10. To scale the latent construct Social Concern, SC14 was used as the reference indicator by fixing

factor loading to 1.0 across the groups. The model for configural invariance fit the data very well, χ^2 (df = 10) = 19.77, p = .032, RMSEA = .052 with 90% CI = (.015; .085), SRMR = .032 and CFI = .99.

Table 10

Step-wise Invariance Test for Subscale Social Concern across Cultures

Dimension	MI	χ^2	*df*	*p*	RMSEA (90% CI)	CFI	SRMR	$\Delta\chi^2$	Δdf	ΔCFI
Social Concern	Configural Invar.	19.77	10	.032	.052 (.015; .085)	.99	.032			
	Partial Metric Invar.	22.59	13	.046	.046 (.006; .077)	.99	.035	2.82	3	.00
	Partial Scalar Invar.	36.43	15	.051	.063 (.037; .089)	.98	.038	0.95	2	-.01

After configural invariance was established, metric invariance for Social Concern was tested across the two cultural subgroups. There was significant χ^2 difference between the metric invariance and configural invariance, $\Delta\chi^2$ = 11.62 (Δdf = 4) and ΔCFI = -.02. Examination of modification indices suggested that equality constraints on factor loading for item SC6 (MI = 28.26) should be released. Partial metric invariance was tested when equality constraint of factor loading for SC6 was freed up. There was no significant χ^2 difference between the partial metric invariance and configural invariance, $\Delta\chi^2$ = 2.82 (Δdf = 3) and ΔCFI = .00. This indicated that the factor loading of SC6 was different across the two cultures, while that of SC14, SC22, SC31 and SC37 were comparable. This means that the observed item-level score differences of SC14, SC22, SC31 and SC37 between the American and Chinese participants can be traced to the differences in the latent variable Social Concern. The same interpretation cannot be made for the item SC6.

With partial metric invariance, scalar invariance was tested for the subscale Social Concern. There was significant χ^2 difference between the scalar invariance and metric invariance, $\Delta\chi^2 = .128.9$ ($\Delta df = 4$) and $\Delta CFI = -.02$. Based on chi-square difference and ΔCFI, the null hypothesis of scalar invariance was rejected. Therefore, with the guidance of modification indices, the equality constraints of intercepts for items SC14 (MI = 70.63) and SC37 (MI = 37.74) were freed up, and there was significant χ^2 difference between the partial scalar invariance, $\Delta\chi^2 = 36.43$ ($\Delta df = 2$); however, $\Delta CFI = -.01$. As a result, scalar invariance test of subscale Social Concern failed to reject the null hypothesis that item intercepts were the same across the groups. Scalar invariance test indicated that SC22 and SC31 had the same intercepts cross-culturally.

Configural, metric and scalar invariance testing for the subscale Affiliation. Configural, metric and scalar invariance for Affiliation were tested across the two cultural subgroups, and the test of model fit is presented in Table 11. To scale the latent construct Affiliation, AF21 was used as the reference indicator by fixing its factor loading to 1.0 across the groups. Since the subscale consisted of only three items, statistically the model was saturated and not feasible for configural test. Since both Affiliation and Social Concern achievement goals address social perspective of individuals' goal orientations (McInerney & Sinclair, 1991), it is deemed appropriate to test the two constructs together. Therefore, for identification purposes, the Affiliation subscale was tested together with the Social Concern subscale, with each subscale represented by their defining items. This resulted in a two-factor model. In an attempt to test the Affiliation subscale primarily, equality constraints were not imposed on items measuring the Social Concern subscale. The model for configural invariance fit the data

very well, χ^2 (df = 38) = 108.82, $p < .001$, RMSEA = .071 with 90% CI = (.056; .087), SRMR = .044 and CFI = .96.

Table 11

Step-wise Invariance Test for Subscale Affiliation across Cultures

Dimension	MI	χ^2	*Df*	*p*	RMSEA (90% CI)	CFI	SRMR	$\Delta\chi^2$	Δdf	ΔCFI
Affiliation + Soconcn	Configural Invar.	108.82	38	<.001	.071 (.056; .087)	.96	.044			
	Partial Metric Invar.	113.22	40	<.001	.071 (.056; .086)	.96	.047	4.40	2	.00
	Partial Scalar Invar.	118.56	43	<.001	.069 (.055; .084)	.96	.047	5.34	3	.00

Metric invariance for Affiliation was tested together with Social Concern across the cultural groups. Without equality constraints on factor loadings of Social Concern items across the groups, there was no significant χ^2 difference for Affiliation between the metric invariance and configural invariance, $\Delta\chi^2 = 4.4$ (Δdf = 2) and ΔCFI = .00. This showed that the factor loadings of the Affiliation subscale were the same across the cultures. It means that the observed item-level score differences of Affiliation between the American and Chinese participants can be traced to the differences in the latent variable Affiliation.

Scalar invariance for Affiliation was tested together with Social Concern across the cultural groups. With item intercepts of the subscale Social Concern freely estimated across groups, there was no significant χ^2 difference for Affiliation between the scalar invariance and metric invariance, $\Delta\chi^2 = 5.3$ (Δdf = 3) and ΔCFI = .00. This showed that the intercepts of the Affiliation subscale were the same across the cultures. It means that the observed item-level score differences of Affiliation between the American and Chinese participants are due to differences on the underlying construct Social concern.

Configural, metric and scalar invariance testing for the subscale Praise.

Configural, metric and scalar invariance for Praise were tested across the two cultural subgroups, and the model fit indices are presented in Table 12. To scale the latent construct Praise, PR15 was used as the reference indicator by fixing factor loading to 1.0 across the groups. The model for configural invariance fit the data very well, χ^2 (df = 10) = 18.43, p = .048, RMSEA = .048 with 90% CI = (.004; .082), SRMR = .030 and CFI = .99. Thus, invariance of factor structure was supported. This was used as the baseline model for metric invariance test.

Table 12

Step-wise Invariance Test for Subscale Praise across Cultures

Dimension	MI	χ^2	*df*	*P*	RMSEA (90% CI)	CFI	SRMR	$\Delta\chi^2$	Δdf	ΔCFI
Praise	Configural Invar.	18.43	10	.048	.048 (.004; .082)	0.99	.030			
	Partial Metric Invar.	27.10	13	.012	.054 (.025; .083)	0.99	.046	8.67*	3	.00
	Partial Scalar Invar.	46.56	15	<.001	.076 (.052; .100)	0.98	.046	28.13***	2	-.01

Note. *** $p < .001$; * $p < .05$

With factor structure being the same, metric invariance for Praise was tested across the two cultural subgroups. There was significant χ^2 difference between the metric invariance and configural invariance, $\Delta\chi^2$ = 59.28 (Δdf = 4) and ΔCFI = -.03. With the guidance of modification indices, equality constraint of factor loading for PR24 (MI = 97.27) was freed up and partial metric invariance was tested. There was still significant χ^2 difference between the partial metric invariance and configural invariance, $\Delta\chi^2$ = 8.67 (Δdf = 3); however, ΔCFI = .00. Therefore, results failed to reject the null hypothesis that there were no differences between the partial metric invariance and configural invariance. This indicated that the factor loading of PR24 was different across

the two cultures, while that of PR7, PR15, PR32, and PR38 were the same. This means that the observed item-level score differences of PR7, PR15, PR32, and PR38 between the American and Chinese participants can be traced to the differences in the latent variable Praise. The same interpretation cannot be made for item PR24. These results provide support for the invariance on the pattern of factor loadings across the cultures.

With partial metric invariance established, scalar invariance for Praise was tested across the cultural groups. There was significant χ^2 difference for Praise between the scalar invariance and partial metric invariance, $\Delta\chi^2 = 65.87$ ($\Delta df = 3$) and $\Delta CFI = -.04$. Therefore, equality constraints on item intercepts of PR7 (MI = 14.92) and PR38 (MI = 31.84) were freed up based on their respective modification indices for tau-X. The nested χ^2 difference between this new partial scalar invariance and metric invariance was significant. However, $\Delta CFI = -.01$. As a result, scalar invariance test of the subscale Praise failed to reject the null hypothesis that item intercepts were the same across the groups. This means that the observed item-level score differences of Praise between the American and Chinese participants are due to differences on the underlying construct Praise.

Configural invariance testing for the subscale Token. Configural invariance for Token was tested across the two cultural subgroups, and the model fit indices are presented in Table 13. To scale the latent construct Token, TK33 was used as the reference indicator by fixing factor loading to 1.0 across the groups. The model fit indices for configural invariance indicated potential model misspecification, χ^2 (df = 28) = 202.91, p = .048, RMSEA = .13 with 90% CI = (.11; .15), SRMR = .079 and CFI = .93. Factor structure of Token within each cultural group indicated quite

different model misfit. For the American group, the model fit indices indicated an accepted model fit with χ^2 (df = 14) = 48.41, $p < .001$, RMSEA = .085 with 90% CI = (.060; .11), SRMR = .041 and CFI = .97. In contrast, for the Chinese group, the model fit indices suggested model misspecification, with χ^2 (df = 14) = 154.50, $p < .001$, RMSEA = .16 with 90% CI = (.14; .18), SRMR = .079 and CFI = .87. Therefore, the subscale Token was excluded from further invariance test.

Table 13

Configural Invariance and Within Group CFA Test for Subscale Token

Dimension	MI	χ^2	*df*	*p*	RMSEA (90% CI)	CFI	SRMR
Token	Configural Invar.	202.91	28	.048	.130 (.110; .150)	.93	.079
	CFA within American	48.41	14	<.001	.085 (.060; .110)	.97	.041
	CFA within Chinese	154.50	14	<.001	.160 (.140; .180)	.87	.079

In summary, results of these analyses suggest that the ISM is not functioning equivalently across the two cultural groups. The configural invariance test of subscales showed that the factor structure of Competition, Social Power and Token needed to be validated further within the culture. Further invariance test was conducted with the subscales of Task, Effort, Affiliation, Social Concern and Praise. To test metric invariance, factor loadings of each item of subscales were constrained to be equal across the groups. The resulted chi-square statistics and CFI were compared with that obtained in the configural invariance test. The metric invariance test indicated that the subscales Task, Effort, Social Concern, Affiliation, and Praise have the same meaning across the two cultural groups. Metric invariance acts as the baseline model for the more stringent invariance test, the scalar invariance test. In testing scalar invariance, intercepts of the items in subscales were constrained to be equal between the Chinese and American

groups. Besides the variance-covariance matrix, item means were analyzed in each subscale. The significance of intercept equality constraints was tested by using chi-square difference and ΔCFI. Only those items within each subscale that have similar factor loadings were given equality constraints on intercepts. Results indicated that the subscales Effort, Social Concern, Affiliation and Praise have the same intercepts across the cultural groups. For those subscales that have scalar invariance, mean differences on the latent factors was generated by setting the American group as the reference group.

Objective II: To test whether there are mean differences in the eight factors across the cultures, given that factor comparability was first established.

Testing for Differences in Observed Means and Latent Factor Means

Cross-cultural differences on the raw scale scores for the ISM subscales were computed. Table 14 shows the mean and standard deviation of the raw scale scores for American and Chinese participants. These results of the t-test for cultural differences showed that Americans had significantly higher raw scale scores than Chinese for Task, $t(715) = 2.51$, $p < .05$ and Praise, $t(707) = 2.35$, $p < .05$. Chinese had significantly higher raw scale scores than Americans for Competition, $t(705) = 2.78$, $p < .05$ and Token, $t(706) = 8.58$, $p < .01$. The effect size (Cohen's *d*) for Task, Competition, and Praise were all low at .19, .21, and .18, respectively. For Token, the effect size was medium .64. American and Chinese students had no significant differences on raw scale scores for Effort, Social Power, Affiliation, and Social Concern.

For an examination of latent mean differences, a prerequisite is invariance up to scalar level for the measurement model for the groups being compared. As reported earlier, there was scalar invariance for the subscales Effort, Affiliation, Social Concern

and Praise and therefore justifies the comparison of latent factor mean scores between the two cultures. Given the abstract nature of factor means, the absolute factor means of each group cannot be determined. Instead, the factor mean differences can be determined by arbitrarily fixing the factor mean of one group as 0 and the value generated for the second group is the factor mean differences between the two groups. In this study, the KAPPA or latent mean differences between American and Chinese for Effort, Affiliation, Social Concern and Praise were -.11, -.11, -.11 and .18, with t values of .2.67, 1.56, 2.67 and 3.12 respectively. The results indicate that Effort, Social Concern and Praise factors showed cross-cultural differences at $\alpha = .01$ level. Chinese have significantly higher latent mean scores than Americans on Effort and Social Concern, while Americans have significantly higher latent mean scores on Praise than Chinese.

Table 14

Group Differences for ISM Factors based on Observed Scores and Latent Factor Scores

ISM factors	Group mean (*SD*)		*df*	t-value	Effect size	KAPPA	t-value
	American	Chinese			(Cohen's *d*)		
Task (4 items)	4.28 (.509)	4.19 (.451)	715	2.51*	.19	NA	NA
Effort (7 items)	3.97 (.525)	4.00 (.454)	705	-.81	.06	-.11	-2.67**
Competition (6 items)	3.21 (.754)	3.35 (.581)	705	-2.78***	.21	NA	NA
Social Power (6 items)	2.96 (.805)	2.96 (.693)	707	.00	.00	NA	NA
Affiliation (3 items)	3.12 (.916)	3.21 (.680)	722	-1.51	.11	-.11	-1.56
Social Concern (5 items)	3.67 (.589)	3.73 (.471)	718	-1.52	.11	-.11	-2.67**
Praise (5 items)	3.60 (.725)	3.48 (.636)	707	2.36*	.18	.18	3.12**
Token (7 items)	2.90 (.728)	3.32 (.571)	706	-8.58***	.64	NA	NA

Note. *** $p < .001$ (2-tailed). ** $p < .01$ (2-tailed). * $p < .05$ (2-tailed). NA = not applicable

As seen from the table, the group means at the observed subscale level do not have statistically significant differences across cultures for Effort, Affiliation, and Social

Concern. However, when error variances are taken into consideration, group latent factor mean differences for Effort, Social Concern, and Praise are statistically significant between the two cultural groups.

Summary

This chapter has presented the results to address the two main purposes of this study: 1) to test whether the construct of achievement goal orientations as operationalized by McInerney's Inventory of Student Motivation is comparable between two cultures; 2) to test whether there are mean differences in the eight factors across the cultures, given that factor comparability was first established. The psychometric properties of the ISM have been tested and results have been presented. The results for reliability analyses indicate that in most cases the reliability was acceptable and variability was present between the two cultural groups.

Results of single group confirmatory factor analysis indicate that the eight-factor model fit each cultural group at the border line with the factors of Task and Effort being highly correlated. However, the full model does not perform the same cross-culturally. Results of measurement invariance test of the ISM subscales showed that the subscales Task, Effort, Social Concern, Affiliation, and Praise mean the same thing (i.e. metrically invariant) cross-culturally. Further, the subscales Effort, Social Concern, Affiliation, and Praise show scalar invariance. However, the factor structure of Competition, Social Power and Token may need to be further validated within the American culture before cross-cultural studies.

Group comparison of latent mean differences was performed for those subscales which reveal scalar invariance. Chinese participants scored significantly higher than

American peers on latent constructs Effort and Social Concern. The two groups did not show significant differences on Affiliation. American participants scored significantly higher than Chinese peers on Praise. A more in-depth discussion about the findings from this study will be presented in the next section.

Chapter V

Discussion and Conclusion

The purposes of this chapter are to highlight the major findings and then discuss the implications of these findings for the theories of achievement goal orientation, construct measurement, especially the ISM, and for cross-cultural comparisons. This chapter will also discuss the limitations of the present study and suggestions for further research.

In the past few decades, achievement goal orientation theories (Dweck, 1986; Dweck & Leggett, 1988) have mainly focused on mastery and performance goals and their effect on individuals' behaviors and achievement. However, Maher and Braskamp's (1986) Personal Investment (PI) Theory is based on social cognitive theory, and considers the influence of social-cultural contexts on how individuals choose to spend their energy and time in particular activities. Unlike other achievement goal orientation theories, PI theory includes social goals, which are considered important in collectivist cultures like China (Ng, 2009). In addition, PI theory recognizes that individuals can have multiple achievement goals at the same time. Hence, it should be applicable in both American and Chinese cultures. The Inventory of School Motivation (ISM, McInerney, Roche, McInerney & Marsh, 1997; McInerney & Sinclair, 1991), based on PI theory, was developed to investigate the nature of school motivation in cross-cultural settings.

The present study addressed two main objectives: 1) to evaluate whether the construct of achievement goal orientations as operationalized by McInerney's ISM was comparable between the two cultures by a step-wise measurement invariance test; 2) to test whether there were mean differences in the eight factors across the cultures.

In addressing the first objective, single-group Confirmatory Factor Analysis (CFA) within each culture and multi-group CFA across cultures were performed. Based on the significant loadings of items on their respective constructs and model fit indices (χ^2, RMSEA, CFI, SRMR), the eight-factor model was considered to fit each cultural group, although borderline yet adequate to proceed with hierarchical multi-group CFA. Results of the configural invariance test of the ISM subscales showed mixed findings, with the subscales of Task, Effort, Social Concern, Affiliation, and Praise showing invariant factor structure but not the subscales of Competition, Social Power, and Token. Therefore, the full eight-factor model is not considered to perform the same cross-culturally. Further validation of factor dimensionality of these three subscales is necessary using exploratory factor analysis. This was not conducted because it was not the focus of this study. Further measurement invariance tests of the subscales Task, Effort, Social Concern, Affiliation, and Praise showed support for partial metric invariance for all and partial scalar invariance for all except the Task subscale.

In addressing the second objective, mean and covariance structures were analyzed. Individual group latent mean was unable to be generated because the multi-group model requires additional constraints to identify a latent factor. This means, for identification purposes, that the intercept of the latent factor in one group needed to be constrained as 0. Consequently, latent mean differences between the groups were produced. Scalar

invariance was a prerequisite for meaningful latent mean comparisons. Hence, only the mean comparisons for the subscales Effort, Social Concern, Affiliation, and Praise were meaningful. Latent mean comparisons suggested that Chinese college students are more motivated to achieve in academic settings by the Effort and Social Concern goals than their American peers are, while American students are more motivated by the Praise goal than Chinese students are; the two groups of students do not differ in the Affiliation achievement goal. The results of latent mean comparisons across groups are contradictory to those obtained from the raw observed scores. For example, comparison of students' Effort achievement goal based on the raw scores revealed that there were no significant differences between the two groups. However, comparison of latent mean of the Effort achievement goal orientation showed that Chinese college students scored higher than American students. This indicates the importance of establishing construct validity and measurement equivalence in cross-group mean comparisons.

Implications for the Theory

Maehr and Braskamp's (1986) Personal Investment theory takes into consideration the social-cultural influence on individuals' perception and choice of incentives in their pursuit of achievement. According to PI theory, the disparities in socialization processes in different cultures lead to differences in values, goals, and meaning of success. PI theory conceptualizes achievement goal orientation in terms that recognize the possibility of diverse modes of achievement behaviors across cultures. The ability to account for social-cultural factors that impact individuals' motivation sets PI theory apart from McClelland's personality-trait theory, which was based on Western individualist ideology and practices and was not considered appropriate for motivational

studies in a collectivist culture like China. Achievement goal theories (A. J. Elliot & McGregor, 2001; Maehr & Midgley, 1991) have been mainly focused on how mastery and performance goal orientations affect individuals' cognitive strategies and academic success in school settings. PI theory, however, includes social goals as an extra dimension, which deals with the social reasons of trying to achieve in schools. Social goals, such as the desire to be affiliated with a group, to maintain group welfare, and to help others in this endeavor, are important aspects of individuals' behavior, affect, and cognition and, therefore, are considered important motivating factors in educational settings.

The findings from within-culture confirmatory factor analysis support the independent and multi-dimensional nature of achievement goals. This provides evidence for the limitations of Dweck's (1986) and Ames' (1992) early theories of achievement goal orientations. According to these theories, mastery and performance goals were two ends of a continuum and individuals were categorized either as mastery or performance goal oriented.

The factor structure analysis in this study indicates that social goals (social concern and affiliation) are important incentives that drive individuals to achieve regardless of their culture. This supports the importance of studying social goals along with academic goals in educational settings. Social goals are important motivational factors for understanding students' behaviors, affect, and cognition in achievement settings (Blumenfeld, 1992; Ryan, 2001; Salili, Chiu & Lai, 2001). The findings that social goals were positively related to mastery (Task and Effort) and performance goals

(Competition and Social Power) support that social goals to be cooperative and compliant support the pursuit of mastery and learning goals (Wentzel, 1992).

Implications for the ISM

Psychometric analysis findings support the belief that the majority of the subscales of the ISM have acceptable reliability in both cultures with the exception of the subscales Task and Social Concern, which have slightly lower levels of reliability. Validity analysis, as indicated by factor correlations in full-model CFA analysis, suggested that the latent constructs Task and Effort have discriminant problems in both cultural samples and that the constructs Praise and Token also have discriminant problems in the Chinese sample. The indistinguishableness of Task and Effort was also found in other literature (McInerney, Roche, McInerney & Marsh, 1997). It is recommended that the two subscales be combined in future studies. Alternatively, the subscale Task might be removed from the instrument because of its instability in terms of reliability.

Modification indices for error correlations obtained from single-group CFA within cultures were examined to detect the possible causes for the discriminant validity problem of the subscales of Praise and Token for the Chinese group but not for the American group. The modification index for correlated error terms between *Praise* 38 (Praise from my parents for my good schoolwork is important to me) and *Token* 25 (I work hard at school for presents from my parents) in the Chinese group is 53.42; whereas, the modification index in the American group is 7.50. The commonality between *Praise* 38 and *Token* 25 is that both items measure receiving recognition from parents because of good schoolwork. It is understandable that the error correlation between the

two items in the Chinese group is much higher than in the American group. Chinese parents place tremendous value on and have high expectation for their children's education, and they invest greatly in their education in terms of money and time (Shek & Chan, 1999). In return, students study hard at school to get good grades to please their parents. Getting praise and presents from their parents indicate they have done a good job, have met their parents' expectation, and their parents are proud of them. In comparison, American parents do not stress the importance of education as much as Chinese parents and they pay more attention to independence and development of an all-around individual. Additionally, American college students are more financially independent and a large proportion of them work long hours on or off campus to pay their tuition (Penn, Ray, Xu, Gross & Stevens, 2009). It is believed the high error correlation between the two items contribute to the discriminant validity problem between the factors Praise and Token in the Chinese sample.

The multi-dimensional structure of achievement goals is supported by the CFA results. These 43 items significantly loaded on their respective *a priori* factors. It is notable that the model worked well with the inclusion of social goals (affiliation and social concern) for academic achievement. This suggests that the social goals, at least as measured by the current items, are significant measures of motivation in school settings. Affiliation, the desire to be with friends at the school and to study with others instead of alone, represents students' social need in educational achievement. Social concern is an inclusive construct, the properties of which are cooperation and collectivism. The value of a cooperative learning structure for improving academic performance has long been known (Slavin, 1987). Individual accomplishments and self-actualization are emphasized

strongly in the American culture, and a way to reach these objectives is through competition. However, cooperation and collaboration are also essential for promoting more in-depth learning and long-term personal development. Academic living and learning communities are a popular organizational means in higher education to enhance students' learning strategies and cognitive abilities while also satisfying students' social needs and affiliation achievement goal orientations.

Having determined the non-invariant findings for the subscales Competition, Social Power, and Token, the question is then whether the differences represented the true differences in the measurement and structure of these subscales or whether the differences were due to certain types of biases. To explore the possible causes of the invariance test failure, modification indices were consulted in combination with the semantic interpretation of items. In the following paragraphs, the possible causes for the subscales' failure in the factor structure invariance test were explored. These possible causes included item bias and response style set. Suggestions for dealing with problematic items, when appropriate, were put forward for future studies.

In the case of the subscale Token, there appears to be item bias. Item bias can be caused by the influence of cultural specifics such as connotations associated with the item wording (Vijver & Tanzer, 2004). Differential sociocultural influences on certain items of a subscale for one group can cause non-invariance to occur. For example, item *Token 40* (Getting merit certificates helps me work harder at school) and item *Token 51* (If I got rewards at school I would work harder) have a strong association with effort/work ethic for the Chinese group and are interpreted differently than intended, as indicated by their significant modification indices with both the factors Task (32.01; 33.96) and Effort

(22.98; 26.95). This can be further supported by the size of the modification indices for their error terms with item *Competition 28* (I work harder if I'm trying to better than others) (11.09; 17.38) and the modification index for the error terms between each other (89.67). The phrase "work harder" in those items carries a strong tone of effort and work ethic, which is valued a great deal in educational settings in China. Therefore, item bias may confound the factor structure of the subscale Token for the Chinese group. Evaluation of modification indices for error terms and item loadings for the American group does not reveal the same pattern. Thus, those items are recommended to be rewritten before further use in cross-cultural comparisons involving a Chinese group. Ali and McInerney's (2005) study of Hong Kong Chinese resulted in similar findings producing evidence that Chinese students do not appear to interpret some items within the Social Power and Token subscales in the same way as the Australian students do.

Item bias seems to be displayed in the subscale Competition. Exploration of the possible misfit after single group CFA analysis for the Chinese subsample indicates that one competition item, *Competition 28*, (I work harder if I'm trying to be better than others) has a very strong connection with effort/task, as indicated by its modification indices for factor loading (MI Task = 73.60 and Effort = 75.40). Chinese culture stresses effort and a strong work ethic (Yang, 1986). Chinese students will work harder and put in more effort if they want to be better than others. Strong competition among students at all levels of school has reinforced the idea that hard work is intertwined with competition. It seems that the Chinese sample focused on the tone of the item and associated it with the construct Effort more than with Competition. The weak covariance of the item with the rest of the Competition items for the Chinese sample supports this hypothesis. The

evidence strongly suggested that item *Competition 28* be rewritten prior to its use in Chinese culture. Evaluation of the modification indices for item loadings for the American group did not reveal the same pattern, with the modification indices of the item on Factor Task as .04 and on Factor Effort as .89. The size of the modification indices for the American group suggested that American students do not associate "working harder if trying to be better than others" with effort/task achievement goals.

Response style set was another possible cause for non-invariance. It is hypothesized that the response style set leads to the non-invariance of item *SC6* on both the factor loading and intercept levels. Cheung and Rensvold (2000) suggest that the extreme ARS [Acquiescence Response Style] due to the tone of an item may cause factor loading invariance. Generally, ARS differences occur when one group systematically gives higher or lower responses. The extreme ARS for the Chinese group was demonstrated on one Social Concern item, *SC6*, (it is very important for students to help each other at school). The collectivist nature of Chinese culture teaches individuals to help each other (Song & Zhang, 2008). Chinese students strongly endorsed the item, as indicated by the item mean (4.46 for Chinese vs. 3.81 for American). However, item loading was quite different between the American and Chinese samples (.58 vs. .29). The small factor loading indicated that *SC6* did not correlate strongly with the construct Social Concern in the Chinese group. According to Cheung and Rensvold (2000), extreme ARS can affect factor loadings of the item.

TK40 (Getting merit certificates helps me work harder at school) seems to be another item that demonstrated the extreme ARS for the Chinese sample. Getting merit certificates is a popular means of recognition of academic excellence in Chinese culture.

It prompts students to study harder to get even better grades at school. Therefore, Chinese participants strongly endorsed this item (mean = 3.85). The extreme ARS affected the intercept of items on the construct Token but not the factor loading. This, although not tested, was confirmed by the similar factor loadings between the American and Chinese samples (.41 vs. .41) and the significantly different item means (3.10 vs. 3.85).

The invariance test for the subscales of Task, Effort, Social Concern, Affiliation, and Praise indicated that these subscales have a similar factor structure and equivalent scale metrics cross-culturally. The scalar invariance test resulted in intercept invariance in all the subscales except the Task subscale. The results further suggested that these subscales with the exception of the Task subscale might be used in the future for direct cross-cultural comparisons of college students' achievement goals between U.S. and China. Additionally, all of these subscales including the Task subscale can be used in testing their relation to other constructs in a nomological net. For example, antecedent causes of achievement goal orientations, such as parents' expectation of education for their children, parents' involvement in schooling, and classroom environments (cooperative vs. competitive) could be examined. Consequential effects of achievement goal orientations on students' engagement and success at school, such as perceived value of school, psychological well-being, performance anxiety, stress, display of adaptive learning behaviors, and use of meta-cognitive strategies (planning, monitoring, and regulation strategies) are suggested to be studied. Extrinsic goals have been associated with self-handicapping and avoidance of help-seeking (Urdan & Midgley, 2001). However, a study of African American students indicates that extrinsic goal motivations appear to have a positive effect on students' self-efficacy and self-regulated learning

(Ryan & Pintrich, 1997). Therefore, how extrinsic goals impact academic self-efficacy and academic outcomes requires further investigation to better understand the similarities and differences between American and Chinese students and the generalizability of extrinsic goals as a motivational mechanism. Research shows that pursuit of the social goals (affiliation goals) can be detrimental to academic achievement, engagement, and learning in the American culture (Anderman, 1999). However, in a collectivist culture, like China, although competition among students is very intense, students tend to work with each other to promote better academic performances. It is worth noting that in China, the desire to be affiliated with a group and wanting to be with friends could have a positive influence on students' achievement. The moderating role of cultures on these relationships is worthy of examination.

Implications for Cross-Cultural Comparisons

Quantitative comparisons across groups or cultures have more demands on measuring instrument than single-group research in terms of reliability and validity. Single-group research can have valid results with reliable and valid measures. Comparative research requires that the instruments measure constructs with the same meaning for individuals varying in demographic groups (Gregorich, 2006). Further, it requires that individuals from different groups use the rating scale similarly. Sources of biases (cultural bias, item bias and response style set) are minimal. Only when these requirements are satisfied fully or at least partially can the comparisons of latent means be defensible and trustworthy.

Necessity of measurement invariance test. Even when the raw means by chance result in the same conclusion as the latent means, it is necessary to perform the

measurement invariance test first (Steenkamp & Baumgartner, 1998). Sufficient measurement invariance is important for cross-cultural or cross-group comparisons. The conclusions derived from the comparison of latent means were not the same as those obtained by comparing raw means. For example, the independent t-tests for the subscales of Effort and Social Concern did not show a statistical significance at the observed mean level but latent mean comparisons showed a statistically significant difference ($p < .01$). This supports the merit of both the measurement invariance test and latent mean comparisons. Without comparisons on latent mean differences cross-culturally, the findings based on raw means were quite misleading. In addition, this verified the necessity of scalar invariance as a precursor of latent mean comparisons. Take the subscale of Effort as an example. All seven items showed metric invariance. However, the metric invariance did not guarantee that group mean comparisons were unbiased. In this instance, compare cross-culturally the composites by summing all six items as well as the three scalar invariance items. The seven-item composite suggested no significant group difference between the American and Chinese groups. However, the three-item composite suggested that the Chinese group reported significantly higher levels of effort than did the American group and this result corresponded to the findings from the latent mean comparisons. The group difference defined by the scalar invariant items verified the importance of establishing scalar invariance before testing latent mean differences.

Cross-cultural comparisons of latent means. Cross-cultural comparisons of latent means indicated that Chinese participants scored higher than their American counterparts on the Effort achievement goal. This seems to be supported by the literature, which contends that effort and hard work are important values in Chinese culture.

Chinese culture, based on Confucian beliefs, emphasizes excellence and, most importantly, hard work and endurance to achieve this (Watkins, McInerney & Lee, 2002). Other than being influenced by traditional culture, Chinese students face tremendous competition from peers and pressure from family to do well at school; therefore, they exert great effort trying to stay on the top of game to get through higher education. Most (72.0%) of these Chinese college students came from families with one or both parents working as farmers/laborers or other low-paid workers (as a referent, 22.6% of American students came from a family with similar background). Being able to come to college for the Chinese students was their primary, if not only, opportunity to have a well-paid, stable job and a brighter future. Therefore, it is not hard to believe this group of students were strong believers in the importance of effort and hard work in academic achievement. Although significantly lower than their Chinese counterparts, the American college students in this study endorsed the Effort achievement goal more strongly than other performance or extrinsic goals. The Effort achievement goal, therefore, appears to be the most important motivational goal for college students regardless of cultural background.

Cross-cultural comparisons of latent means indicated that Chinese participants scored higher than their American counterparts on the Social Concern achievement goal. This seems to be supported by literature (Salili, 1996b) which shows that Chinese students are collectivist in that they value the success of their school group and they are even willing to even sacrifice their own interest for the group's sake. Chinese students are taught that helping each other is a virtue. Further, students of the same major usually live together on campus in dormitories, and this environment facilitates tremendous collective activities, including academic discussion and collaboration. Although significantly lower

than their Chinese peers, the American students in this study endorsed a great deal the Social Concern achievement goal (mean = 3.67 out of 5, *Strongly Agree*). Social concern promotes students' academic achievement. Take the living and learning communities for example, in many colleges and universities in the U.S., these learning communities are integrated into educational practices to provide students with a quality education, to increase their engagement and learning outcomes (Zhao & Kuh, 2004) and to raise their retention and success rates. Psychologically, learning communities could probably help engender a sense of belonging while promoting positive feelings toward learning. Students in these communities work with their peers and/or mentors. One major goal of these communities is to cultivate students' respect for diversity and differences while strengthening their learning opportunities by teaching and helping others to learn. These learning communities have a strong element of social concern built into their philosophical mission.

Cross-cultural comparisons of the latent mean of Affiliation indicated that the Chinese participants did not score significantly differently from their American counterparts. Similarly, these two groups were not significantly different from each other on the observed score level as measured by the current items. Observed level difference on Affiliation seems to represent the difference on the latent level. This finding fails to support the belief that individuals in a collectivist culture, such as China, should be more affiliation oriented and motivated by cooperation in school settings; whereas individuals in individualist culture, such as the U.S., seek power and control over others and desire individual success through achieving personal goals (McInerney, Roche, McInerney & Marsh, 1997). As mentioned above, Chinese college students of the same cohort in the

same disciplinary major usually share the same dormitories, go to the same classes, participate in the same activities, and work on assignments together. Therefore, it is surprising that Chinese students did not score higher on the Affiliation achievement goals.

Cross-cultural comparisons of the latent mean of Praise indicated that the American participants scored higher than their Chinese counterparts. This was not surprising because praise in the American culture is a very popular means of encouragement from parents and teachers. Praise is used to boost individuals' self-esteem, and students are praised even for trivial accomplishments. Praise is an important motivational goal in American students' daily life and consequently, they become reliant on praise for achievement. In contrast, students in Chinese culture are taught the values of humility and self-deprecation (Salili, 1995). They are told not to feel proud when successful, but to focus on their shortcomings, improve their deficiencies and try to perfect themselves. Praise is rarely used and is mostly associated with exceptionally good work. Teachers and parents consider praise as detrimental to students' character if they are given too frequently (Salili, 1996b). Chinese students do not expect to be praised consistently from their parents or teachers. Therefore, it is expected that American students would score significantly higher than their Chinese peers on the amount of Praise achievement goals.

Practical Implications

To improve students' academic motivation and engagement in educational settings, practitioners should consider the importance of students' social goals in together with their academic goals. Mastery goals (task and effort) are not the only adaptive approach to learning. The importance of social goals in predicting learning engagement

and academic outcome deserves more attention. Policy makers and student affairs administrators in American culture may develop programs to develop group rather than individual learning situations, and structure learning around peer relations to increase students' success and retention in school settings. Research shows that competitive modes of learning are less successful than group learning (McInerney & Swisher, 1995).

Limitations

There are several limitations to this study, which need to be taken into account when interpreting the results. First, the ISM instrument was primarily validated among high school students. This study, to the researcher's best knowledge, is the first attempt to use the ISM with college students. Some items may be more appropriate for high school students than for college students. For example, one of the reviewers of the translation equivalence of the ISM questioned the appropriateness of one Token item (Getting merit certificates helps me work harder at school) for college students. This item probably needs to be rewritten prior to being applied to college students.

Second, online administration of the survey may be different from traditional paper-and-pencil survey methods. In comparison to the paper-and-pencil survey method, the web survey has the advantages of low cost, fast return speed, and minimized interviewer bias (Dillman, 2007), especially in international surveys. In this study, around 53% of the Chinese participants were college freshmen. While implementing data collection in China, it was noted that freshmen in one of the universities had restricted access to computers and the internet. These students were only allowed to access computers and the internet for academic purposes and only for a limited time. Therefore,

this group of participants may be different from general Chinese college students and raises the question of the representativeness of the sample.

Lastly, participants who volunteered and those who received extrinsic incentives (i.e. extra credit) to participate in the survey might be different in terms of their demographic characteristics and achievement goals. According to Sharp, Pelletier, and Levesque (2006), while offering course credit could improve the response rate, it also affects the sample characteristics in terms of motivation characteristics, and thus affects the representativeness of the sample. In this study, a small amount of course credit was offered to some American participants. But the differences between volunteers and non-volunteers (i.e. reward receivers) could not be studied since no identifiable information was collected from the participants to allow for any distinction between the groups. Consequently, the group differences on sample characteristics cannot be identified for the study.

Suggestions for Further Research

Based on the findings and limitations from this research, the following recommendations are made for future research. First, further validation of the factor structures of the Performance goal orientations (i.e. Competition and Social Power) and Token goals are necessary in the American culture before the cross-cultural invariance test. The discriminant validity problem between the Task and Effort achievement goals suggest the redundancy of the two subscales. It is suggested that the Task subscale be removed in the future studies as a part of the measurement.

Second, the combined methods of participant recruitment (voluntary vs. incentive received) in this study can potentially confound the sample characteristics and motivation

comparisons between groups. A consistent sample recruiting procedure is recommended for future studies in order to provide more clear and generalizable results.

Third, exploration of the possible causes of cross-cultural invariance failure suggests that some of the items should be rewritten to reduce the effect of culture specific connotation. It is suggested that further studies rewrite some of the incentive items to make them more appropriate for the participants.

Fourth, criterion variables such as school attendance, math self-efficacy, math GPA, psychological well-being, learning strategies (i.e. rote versus deep learning), and intention to finish college could be included to test the predictive validity of the achievement goal orientations using structural equation modeling. This will shed light on the differential relationships between goal orientations and academic outcomes and afford the opportunity to evaluate the possible moderation effect of culture on these relationships. Further, the achievement gap between American and Chinese students on mathematics and science might be partially due to their differences in achievement goal motivations. Studying achievement goal motivation together with other variables in a nomological network will help discover the causes of this achievement gap.

Lastly, examination of first-order multidimensional factor structure suggests that higher-order factor structure is possible to explain the high correlations (e.g. Phi = 0.91for the factors Task and Effort) among the first-order factor level. Specifically, the two factors posited to represent a second-order factor are more highly correlated with each other than with any other factors. McInerney and Sinclair (1991) developed the Inventory of School Motivation as a hierarchical and multi-dimensional achievement goal measuring instrument. Empirical evidence in this study appears to support the existence

of an underlying second-order structure. Verification of the hierarchical structure will allow for a more precise understanding of these goal orientations and allow for a more parsimonious study of students' motivation through fewer factors.

Conclusion

Measurement invariance testing is still a relatively new analysis method for cross-cultural construct validation. It tests whether the constructs can measure the same thing. Items on the scales must be interpreted in the same way across groups, and then a cross-cultural comparison of construct mean levels, correlates, and consequences can be interpreted in a meaningful way (Little, 2000). Bryne (2003) argued that this kind of analysis is especially appropriate for making cross-cultural comparisons.

The purpose of this study was to assess the measurement invariance of the Inventory of School Motivation across American and Chinese university students. To the researcher's best knowledge, this is the first measurement invariance study with the Chinese college students. This study has shown that confirmatory factor analysis and the measurement invariance test are powerful data analysis methods in cross-cultural comparisons.

The following conclusions can be drawn from this study. The reliability of the subscales of the Inventory of School Motivation is generally acceptable in both American and Chinese cultures. It appears that the subscales of Task and Effort are redundant in measuring students' mastery achievement goal in both cultures. The subscales of Token and Praise appear to measure the same construct in Chinese culture. Item bias and response style set due to cultural characteristics seemingly lead to Chinese students' non-invariant use of the rating scale and different interpretation of some items from American

students. Mean differences between direct cross-cultural comparison on the raw score levels based on some of the subscales are misleading.

References

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1-63). San Diego: Academic Press.

Ali, J. & McInerney, D. M. (2005). *Testing the invariance of a motivation model across seven cultural groups*. Paper presented at the AARE 2005 International Educational Conference.

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Ames, C. & Archer, J. (1987). Mother's beliefs about the role of ability and effort in school learning. *Journal of Educational Psychology, 79*, 409-414.

Anderman, L. H. (1999). Expanding the discussion of social perceptions and academic outcomes mechanisms and contextual influences. In T. C. Urdan (Ed.), *Advances in motivation and achievement: The role of context* (Vol. 11, pp. 303-306). Stamford, Connecticut: JAI Press.

Anderman, L. H. & Anderman, E. M. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology, 24*, 21-37.

Archer, J. (1994). Achievement goals as a measure of motivation in university students. *Contemporary Educational Psychology, 19*, 430-446.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandura, A. & Dweck, C. (1985). The relationship of conceptions of intelligence and achievement goals to achievement-related cognition, affect and behavior. Unpublished manuscript.

Baranik, L. E., Barron, K. E. & Finney, S. J. (2007). Measuring goal orientation in a work domain. *Educational and Psychological Measurement, 67*, 697-718.

Barron, K. E., Baranik, L. E. & Finney, S. J. (2006). *Mastery-avoidance at work*. Paper presented at the AERA's 2006 National Conference in San Francisco.

Barron, K. E. & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*, 706-722.

Behling, O. & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. London: Sage.

Bempechat, J. & Drago-Severson, E. (1999). Cross-national differences in academic achievement: Beyond etic conceptions of children's understandings. *Review of Educational Research, 69*, 287-314.

Blumenfeld, P. C. (1992). Classroom learning and motivation: Clarifying and expanding goal theory. *Journal of Educational Psychology, 84*, 272-281.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185-216.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

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

Button, S. B., Mathieu, J. E. & Zajac, D. M. (1996). Goal orientation in organizational research: A conceptual and empirical foundation. *Organizational Behavior and Human Decision Process, 67*(1), 26-48.

Byrne, B. M. (2003). Testing for equivalent self-concept measurement across culture. *International Advances in Self Research, 1*, 291-313.

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.

Campbell, H. L., Barry, C. L., Joe, J. N. & Finney, S. J. (2008). Configural, metric, and scalar invariance of the modified Achievement Goal Questionnaire across African American and White university students. *Educational and Psychological Measurement*.

Chen, C., Lee, S. Y. & Stevenson, H. W. (1995). Response style and cross cultural comparisons of rating scales among East Asia and North American students. *Psychological Science, 6*, 170-175.

Cheung, G. W. & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.

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*, 187-212.

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*, 233 - 255.

Chiu, C., Hong, Y. & Dweck, C. (1994). Toward an integrative model of personality and intelligence: A general framework and some preliminary steps. In R. J. Sternberg & P. Ruzgis (Eds.), *Intelligence and personality* (pp. 104-134). New York: Cambridge University Press.

Covington, M. (2000). Goal theory, motivation, and school achievement: An integrative Review. *Annual Review of Psychology, 51*, 171-200.

Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16-29.

Day, E. A., Radosevich, D. J. & Chasteen, C. S. (2003). Construct- and criterion-related validity of four commonly used goal orientation instruments. *Contemporary Educational Psychology, 28*, 434-464.

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.

Dillman, D. A. (2007). *Mail and internet surveys : the tailored design method* (2nd ed.). Hoboken, N.J.: Wiley.

Dowson, M. & McInerney, D. M. (2001). Psychological parameters of students' social and work avoidance goals: A qualitative investigation. *Journal of Educational Psychology, 93*, 35-42.

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues. *Psychological Bulletin, 95*, 134-135.

Duda, J. L. (1998). *Advances in sport and exercise psychology measurement*. Morgantown, W. Va.: Fitness Information Technology, Inc.

Duda, J. L. & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology, 84*, 290-299.

Duda, J. L. & Whitehead, J. (1998). Measurement of goal perspectives in the physical domain. In J. L. Duda (Ed.), *Advances in sport and exercies psychology measurement* (pp. 21-48). Morgantown, WV: Fitness Informatioin Technology.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Dweck, C. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychology Review, 95*, 256-273.

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist, 34*, 169-189.

Elliot, A. J. & Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 78*, 218-232.

Elliot, A. J. & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519.

Elliot, A. J. & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review, 13*, 139-156.

Elliot, J. G. & Bempechat, J. (2002). The culture and contexts of achievement motivation. *New Directions for Child and Adolescent Development, 96*, 7-26.

Fyans, L. G., Salili, F., Maehr, M. L. & Desai, K. A. (1983). A cross-cultural exploration into the meaning of achievement. *Journal of Personality and Social Psychology, 44*, 1000-1013.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*, S78-S94.

Hambleton, R. K. & Bollwark, J. (1991). Adapting tests for use in different cultures: Technical issues and methods. *Bulletin of the International Test Commission, 18*, 3-32.

Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T. & Elliot, A. J. (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology, 73*, 1284-1295.

Harackiewicz, J. M., Barron, K. E. & Elliot, A. J. (1998). Rethinking achievement goals: When are they adaptive for college students and why? *Educational Psychologist, 33*, 1-21.

Heine, S. J., Lehman, D. R., Peng, K. & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The reference-group effect. *Journal of Personality and Social Psychology, 82*, 903-918.

Hershberger, S. L. (2003). The growth of structural equation modeling: 1994 to 2001. *Structural Equation Modeling, 10*, 35-46.

Ho, D. Y. F. (1986). Chinese patterns of socialization: A critical review. In M. H. Bond (Ed.), *The psychology of the Chinese people* (pp. 1-37). Hong Kong: Oxford University Press.

Ho, D. Y. F. & Kang, T. K. (1984). Intergenerational comparisons of child-rearing attitudes and practices in Hong Kong. *Developmental Psychology, 20*, 1004-1016.

Horn, J. L. (1991). Comments on issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114-125). Washington, DC: American Psychological Association.

Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 117-144.

Horn, J. L., McArdle, J. J. & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *The Southern Psychologist, 1*, 179-188.

Hu, L. T. & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage.

Jagacinski, C. M. & Duda, J. L. (2001). A comparative analysis of contemporary achievement goal orientation measures. *Educational and Psychological Measurement, 61*, 1013-1039.

Jöreskog, K. G. & Dag, S. (2006). *LISREL 8 user's reference guide* (2nd ed.). Chicago, IL: Scientific Software International.

Juvonen, J. & Weiner, B. (1993). An attributional analysis of students' interactions: The social consequences of perceived responsibility. *Educational Psychology Review, 5*, 325-345.

Lee, W. O. (1996). The cultural context for Chinese learners: Conceptions of learning in the Confucian tradition. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 25-41). Hong Kong: Comparative Education Research Center.

Leong, F. T. L. & Austin, J. T. (1996). *The psychology research handbook: A guide for graduate students and research assistants*. Thousand Oaks, CA: Sage.

Li, J. (2004). "I learn and I grow big": Chinese preschoolers' purpose for learning. *International Journal of Behavioral Development, 28*, 116-128.

Little, T. D. (1997). Mean and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues. *Multivariate Behavioral Research, 32*(1), 53-76.

Little, T. D. (2000). On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. *Journal of Cross-Cultural Psychology, 31*, 213-219.

Maehr, M. L. (1978). Sociocultural origins of achievement motivation. In D. Bar-Tal & I. Saxe (Eds.), *Social psychology of education: Theory and research*. New York: John Wiley & Sons.

Maehr, M. L. (1984). Meaning and motivation. In R. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 1). New York: Academic Press.

Maehr, M. L. & Braskamp, L. A. (1986). *The motivation factor: A theory of personal investment*. Lexington, MA: Lexington.

Maehr, M. L. & McInerney, D. M. (2004). Motivation as personal investment. In D. M. McInerney & S. V. Etten (Eds.), *Research on sociocultural influences on motivation and learning: Big theories revisited* (Vol. 4, pp. 61-90). Greenwich, Conn: IAP.

Maehr, M. L. & Midgley, C. (1991). Enhancing student motivation: A school-wide approach. *Educational Psychologist, 26*, 399-427.

Maehr, M. L. & Nicholls, J. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 2, pp. 221-267). New York: Academic Press.

Maehr, M. L. & Yamaguchi (2001). Cultural diversity, student motivation and achievement. In F. Salili & C. Y. Chiu (Eds.), *Student motivation: The culture and context of learning* (pp. 123-148). New York: Kluwer Academic/Plenum.

Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of self-concept, and academic achievement. *Journal of Educational Psychology, 76*, 1291-1308.

Marsh, H. W. & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/ external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56-67.

McClelland, D. C. (1961). *The achieving society*. Princeton, NJ: Van Nostrand.

McClelland, D. C. (1963). Motivational patterns in Southeast Asia with special reference to the Chinese case. *Social Issues, 19*, 6-19.

McInerney, D. M. (1992). Cross-cultural insights into school motivation and decision making. *Journal of Intercultural Studies, 13*, 53-74.

McInerney, D. M. (1995). Goal theory and indigenous minority school motivational: Relevance and application. In M. L. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement: culture, motivation and achievement* (pp. 153-182). Greenwich, CT: JAI.

McInerney, D. M., Marsh, H. W. & Yeung, A. S. (2003). Toward a hierarchical model of school motivation. *Journal of Applied Measurement, 4*, 335-357.

McInerney, D. M., Roche, L. A., McInerney, V. & Marsh, H. W. (1997). Cultural perspectives on school motivation: The relevance and application of goal theory. *American Educational Research Journal, 34*, 207-236.

McInerney, D. M. & Sinclair, K. E. (1991). Cross cultural model testing: Inventory of School Motivation. *Educational and Psychological Measurement, 51*, 123-133.

McInerney, D. M. & Swisher, K. G. (1995). Exploring Navajo motivation in school settings. *Journal of American Indian Education, 34*, 28-50.

McInerney, D. M. & Van Etten, S. (2002). Big theories revisited: The challenge. In D. M. McInerney & S. Van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol. 2, pp. 1-12). Greenwich, CN: IAP.

McInerney, D. M., Yeung, A. S. & McInerney, V. (2000). *The meaning of motivation: Multidimensional and hierarchical perspectives*. Paper presented at the 81st Annual Meeting of the American Educational Research Association.

McInerney, D. M., Yeung, S. Y. & McInerney, V. (2001). Cross-cultural validation of the Inventory of School Motivation. *Journal of Applied Measurement, 2*, 135-153.

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

Midgley, C., Maehr, M. L., Hicks, L., Roeser, R., Urdan, T. C., Anderman, E. M., et al. (1996). *Patterns of Adaptive Learning Survey*. Ann Arbor: University of Michigan Press.

Morgan, C. & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology and Psychiatry, 34*, 289-306.

Mullen, M. R. (1995). Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies, 26*, 573-596.

Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

Ng, C. H. (2009). 'Learning for achievement' as a collective goal in re-culturing teaching and learning in Hong Kong classrooms. In C. H. Ng & P. D. Renshaw (Eds.), *Reforming learning: Concepts, issues and practice in the Asian-Pacific region* (Vol. 5, pp. 255-276). Dordrecht, Netherlands: Springer.

Nicholls, J. G. (1984). Achievement motivation: Conception of ability, subjective experience, task choice, and performance. *Psychology Review, 91*, 328-346.

Nicholls, J. G. (1989). *The Competitive Ethos and Democratic Education*. Cambridge, MA: Harvard University Press.

Nicholls, J. G., Patashnick, M. & Nolen, S. B. (1985). Adolescents' theories of education. *Journal of Educational Psychology, 77*, 683-692.

Patrick, H., Anderman, L. H. & Ryan, A. M. (2002). Social motivation and the classroom social environment. In C. Midgley (Ed.), *Goals, goal structures and patterns of adaptive learning* (pp. 85-108). Mahwah, NJ: Erlbaum.

Penn, J., Ray, C., Xu, L., Gross, T. & Stevens, J. (2009). *2009 National Survey of Student Engagement*. Stillwater, OK: Oklahoma State University.

Phelps, L. (2005). Academic achievement of children in China: The 2002 Fulbright experience (Publication no. 10.1002/ pits.20074). Retrieved February 10th, 2008, from Wiley InterScience www.interscience.wiley.com:

Pintrich, P. R. & Garcia, T. (1991). Student goal orientation and self-regulation in the college classroom. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 7, pp. 371-402). Greenwich, CT: JAI.

Pintrich, P. R. & Schunk, D. H. (1996). *Motivation in education: Theory, research, and application*. Englewood Cliff, NJ: Prentice Hall.

Raykov, T. (2004). Estimation of maximal reliabilities: A note on a covariance structure modeling approach. *British Journal of Mathematical and Statistical Psychology, 57*, 21-27.

Rensvold, R. B. & Cheung, G. W. (1998). Testing measruement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement, 58*(6), 1017-1034.

Robert, C., Lee, W. C. & Chan, K.-Y. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology, 59*, 65-99.

Roeser, R. (2004). Competing schools of thought in achievement goal theory? . In P. Pintrich & M. L. Maehr (Eds.), *Advances in motivation and achievement: Motivating students, improving schools: The legacy of Carol Midgley* (Vol. 13, pp. 361-392). Oxford, England: Elsevier.

Ross, M. E., Blackburn, M. & Forbes, S. (2005). Reliability generalization of the Patterns of Adaptive Learning Survey Goal Orientation Scales. *Educational and Psychological Measurement, 65*, 451-464.

Ryan, A. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development, 72*, 1135-1150.

Ryan, A. & Pintrich, P. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology, 89*, 329-341.

Salili, F. (1995). Explaining Chinese students' motivation and achievement: A socio-cultural analysis. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement: Culture, motivation and achievement* (Vol. 9, pp. 73-118). Greenwich, Connecticut: Jai Press.

Salili, F. (1996a). Achievement motivation: A cross-cultural comparison of British and Chinese students. *Educational Psychology, 16*, 271-279.

Salili, F. (1996b). Learning and motivation: An Asian perspective. *Psychology and Developing Societies, 8*, 55-81.

Salili, F., Chiu, C. Y. & Lai, S. (2001). The influence of culture and context on students' motivational orientation and performance. In F. Salili, C. Y. Chiu & Y. Y. Hong

(Eds.), *Student motivation: The culture and context of learning* (pp. 221-241). New York: Kluwer Academic/ Plenum.

Salili, F. & Hau, K. T. (1994). The effects of teachers' evaluative feedback on Chinese students' perception of ability: A cultural and situational analysis. *Educational Studies, 20*, 223-236.

Sharp, E. C., Pelletier, L. G. & Levesque, C. (2006). The double-edged sword of rewards for participation in psychology experiments. *Canadian Journal of Behavioral Science, 38*, 269-277.

Shek, D. T. L. & Chan, L. K. (1999). Hong Kong Chinese parents' perceptions of the ideal child. *Journal of Psychology, 133*, 291-302.

Slavin, R. E. (1987). Cooperative learning and the cooperative school. *Educational Leadership, 45*, 7-13.

Song, Y. & Zhang, Y. B. (2008). Cultural values in Chinese children's animation: A content analysis of the Legend of Nezha. *China Media Research, 4*(3), 18-27.

Sperber, A. D., Devellis, R. F. & Boehlocke, B. (1994). Cross-cultural translation: Methodology and validation. *Journal of Cross-Cultural Psychology, 25*, 501-524.

Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *The Journal of Consumer Research, 25*, 78-90.

Stevenson, H. W. & Stigler, J. W. (1992). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit Books.

Sue, S. (1999). Where have we gone wrong? *American Psychologist, 54*, 1070-1077.

Sue, S. & Chang, J. (2003). The state of psychological assessment in Asia. *Psychological Assessment, 15*, 306-310.

Tabachnick, B. G. & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

Tao, V. & Hong, Y. Y. (2000). A meaning system approach to Chinese students' achievement goals. *Journal of Psychology in Chinese societies, 1*(2), 13-38.

The Carnegie Foundation (2006). The Carnegie classification of institutions of higher education Retrieved 09/09/2009, from http://www.carnegiefoundation.org/classifications/

Urdan, T. C. (1997). Achievement goal theory: Past results, future directions. In M. L. Maehr & P. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 99-141). Greenwich, CT: JAI.

Urdan, T. C. & Maehr, M. L. (1995). Beyond a two-goal theory of motivation and achievement: A case for social goals. *Review of Educational Research, 65*(3), 213-243.

Urdan, T. C. & Mestas, M. (2006). The goals behind performance goals. *Journal of Educational Psychology, 98*, 354-365.

Urdan, T. C. & Midgley, C. (2001). Academic self-handicapping: What we know, what more there is to learn. *Educational Psychology Review, 13*, 115-138.

Vallerand, R. J. (1989). Vers une methodologie de validation trans-culturelle de questionnaires psychologiques: Implications pour la recherche en langue francaise [Toward a methodology of cross-cultural validation of psychological

questionnaires: Implication for research in French]. *Psychologie Canadienne, 30*, 662-680.

Vandenberg, R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement, 57*, 995-1015.

Vijver, F. V. d. & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

Vijver, F. V. d. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54*, 119-135.

Watkins, D., McInerney, D. M. & Lee, C. (2002). Assessing the school motivation of Hong Kong students. *Psychologia, 45*, 144-154.

Wentzel, K. R. (1992). Motivation and achievement in adolescence: A multiple goals perspective. In D. H. Schunk & J. L. Meece (Eds.), *Student Perceptions in the Classroom* (pp. 287-306). Hillside, NJ: Lawrence Erlbaum.

Wentzel, K. R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology, 85*(2), 357-364.

Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding students' academic success. *Journal of Educational Psychology, 91*, 76-97.

Wentzel, K. R. (2000). What is it I'm trying to achieve? Classroom goals from a content perspective. *Contemporary Educational Psychology, 25*, 105-115.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., et al. (2005). Principles of good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health, 8*, 94-104.

Yang, K. S. (1986). Chinese personality and its change. In M. H. Bond (Ed.), *The psychology of the Chinese people*. Hong Kong: Oxford University Press.

Yu, A. B. & Yang, K. S. (1987). Social-oriented and individual-oriented achievement motivation: A conceptual and empirical analysis. *Bullet of the Institute of Ethnology, Academia Sinica (Taiwan), 64*, 51-98.

Yu, A. B. & Yang, K. S. (1994). The nature of achievement motivation in collectivist societies. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi & G. Yoon (Eds.), *Individualism and collectivism: Theory, method, and applications* (Vol. 18, pp. 239-250). Thousand Oaks, CA: Sage.

Zhao, C.-M. & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education, 45*, 115-138.

APPENDIX A: INSTITUTIONAL REVIEW BOARD APPROVAL

Oklahoma State University Institutional Review Board

Date:	Monday, November 17, 2008
IRB Application No	ED08172
Proposal Title:	Measurement Invariance of Inventory of School Motivation between the United States and Chinese College Students
Reviewed and Processed as:	Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 11/16/2009

Principal Investigator(s):

Lihua Xu
107 Univ. Assess. & Testing
Stillwater, OK 74078

Laura Barnes
700 N. Greenwood
Tulsa, OK 74016

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☒ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Shelia Kennison, Chair
Institutional Review Board

APPENDIX B: INSTITUTIONAL REVIEW BOARD

APPROVAL OF MODIFICATION

Oklahoma State University Institutional Review Board

Date:	Monday, March 09, 2009	Protocol Expires:	11/16/2009
IRB Application No:	ED08172		
Proposal Title:	Measurement Invariance of Inventory of School Motivation between the United States and Chinese College Students		
Reviewed and Processed as:	Exempt **Modification**		

Status Recommended by Reviewer(s) **Approved**

Principal Investigator(s):

Lihua Xu
107 Univ. Assess. & Testing
Stillwater, OK 74078

Laura Barnes
700 N. Greenwood
Tulsa, OK 74106

The requested modification to this IRB protocol has been approved. Please note that the original expiration date of the protocol has not changed. The IRB office MUST be notified in writing when a project is complete. All approved projects are subject to monitoring by the IRB.

[X] The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

The reviewer(s) had these comments:

The modification request to allow recruitment via the SONA system is approved.

Signature :

Sheila Kennison, Chair, Institutional Review Board

Monday, March 09, 2009
Date

APPENDIX C-1E: EMAIL INVITATION TO FACULTY (ENGLISH VERSION)

My name is Lihua Xu. I am a doctoral student in Research and Evaluation in the College of Education at Oklahoma State University. I am recruiting approximately 500-600 students to participate in an online survey housed in the College of Education. The whole survey takes approximately 15 minutes. The title of this research is *Measurement Invariance of the Inventory of School Motivation across Chinese and American College Students*. The purpose of this study is to test whether the achievement motivation instrument can be interpreted similarly in Chinese and American cultures. This is significant in cross-cultural study of student motivation and education outcomes.

No participant's names will be asked for or used on any study data. Students' participation is voluntary and they can choose to terminate the survey at any time during the survey. The data will be stored in a password protected personal computer and only the researcher involved in this study have the access to them and they will be destroyed in five years. There are no known risks associated with this project which are greater than those ordinarily encountered in daily life.

For more information, you can contact Lihua Xu (lihua.xu@okstate.edu 405-744-6924). Should you have any questions about this project you may also ask the project advisor, Dr. Laura Barnes at Oklahoma State University (lbarnes@okstate.edu 918-594-8517). If you have any questions about students' rights as a research participant, you may contact Dr. Shelia Kennison, IRB Chair, 219 Cordell North, Stillwater, OK 74078, 405-744-1676 or irb@okstate.edu.

To go to the online survey, please click or paste in browser: http://frontpage.okstate.edu/coe/lihuaxu1

Please inform your students about this survey and encourage their participation. Please distribute the enclosed flyers to students in your class.

Thank you for your assistance with the survey.

Lihua Xu
Doctoral Student
Research and Evaluation
College of Education

Dr. Laura Barnes
Dissertation Advisor
Research and Evaluation
College of Education

APPENDIX C-1C: EMAIL INVITATION TO FACULTY (CHINESE VERSION)

亲爱的老师,

我是徐丽华，现在奥克拉河马州立大学攻读博士学位。为了完成毕业论文要求我需要收集 500 - 600 份关于学生学习动机的问卷调查。该论文的目的是比较中西文化下学生对该动机问卷的诠释是否一致。这对跨文化研究学习动机和学习成绩很有意义。

该问卷调查不询问被调查者的名字，整个问卷完成大概需要 15 分钟时间。问卷参与纯属自愿，调查者可以随时中断该调查。调查数据严格保密，只有研究者可以接触到。该数据会保存五年时间，之后会销毁。问卷参与不会产生比日常生活更大的危害。

如果您对该调查需要进一步的信息，请通过电子邮件 lihua.xu@okstate.edu 联系论文作者，或论文指导老师 (lbarnes@okstate.edu）。如果您对学生作为问卷参与者的权利有任何问题，请联系奥克拉河马州立大学的审查委员会。 该会的电子邮件地址为 irb@okstate.edu。

该问卷的网址为（**http://frontpage.okstate.edu\coe\lihuaxu2**）。

请积极鼓励您的学生参与，我对此衷心感谢。

此致

敬礼

徐丽华
博士生
教育学院
奥克拉河马州立大学
美国

APPENDIX C-2E: FOLLOW-UP EMAIL TO FACULTY (ENGLISH VERSION)

This is a reminder for you to encourage your students to participate in an online survey about achievement motivation.

My name is Lihua Xu. I am a doctoral student in Research and Evaluation in the College of Education at Oklahoma State University. I am recruiting approximately 500-600 students to participate in an online survey housed in the College of Education. The title of this research is *Measurement Invariance of the Inventory of School Motivation across Chinese and American College Students*. The purpose of this study is to test whether the achievement motivation instrument has similar psychometric properties in the two cultures. This is significant in cross-cultural study of student motivation and education outcomes.

No participant's names will be asked or used on any study data. The whole survey takes approximately 15 minutes. Students' participation is voluntary and they can choose to terminate the survey at any time during the survey. The data will be stored in a password protected personal computer and only the researcher involved in this study have the access to them and they will be destroyed in five years. There are no known risks associated with this project which are greater than those ordinarily encountered in daily life.

For more information, you can contact Lihua Xu (lihua.xu@okstate.edu 405-744-6924). Should you have any questions about this project you may also ask the project advisor, Dr. Laura Barnes at Oklahoma State University (lbarnes@okstate.edu 918-594-8517). If you have any questions about students' rights as a research participant, you may contact Dr. Shelia Kennison, IRB Chair, 219 Cordell North, Stillwater, OK 74078, 405-744-1676 or irb@okstate.edu.

To go to the online survey, please click or paste in browser: (http://frontpage.okstate.edu/coe/lihuaxu1)

Please inform your students about this survey and encourage their participation. Please distribute the enclosed flyers to students in your class.

Thank you for your assistance with the survey.

Lihua Xu
Doctoral Student
Research and Evaluation
College of Education

Dr. Laura Barnes
Dissertation Advisor
Research and Evaluation
College of Education

APPENDIX C-2C: FOLLOW-UP EMAIL TO FACULTY (CHINESE VERSION)

亲爱的老师,

此信的目的是提醒您鼓励学生参与学习动机的互连网问卷调查。

我是徐丽华，现在奥克拉河马州立大学攻读博士学位。为了完成毕业论文要求我需要收集 500 - 600 份关于学生学习动机的问卷调查。该论文的目的是比较中西文化下学生对该动机问卷的诠释是否一致。这对跨文化研究学习动机和学习成绩很有意义。

该问卷调查不询问被调查者的名字，整个问卷完成大概需要 15 分钟时间。问卷参与纯属自愿，调查者可以随时中断该调查。调查数据严格保密，只有研究者可以接触到。该数据会保存五年时间，之后会销毁。问卷参与不会产生比日常生活更大的危害。

如果您对该调查需要进一步的信息，请通过电子邮件 lihua.xu@okstate.edu 联系论文作者，或论文指导老师 (lbarnes@okstate.edu）。如果您对学生作为问卷参与者的权利有任何问题，请联系奥克拉河马州立大学的审查委员会。 该会的电子邮件地址为 irb@okstate.edu。

该问卷的网址为（**http://frontpage.okstate.edu\coe\lihuaxu2**）。

请积极鼓励您的学生参与，我对此衷心感谢。

此致

敬礼
徐丽华
博士生
教育学院
奥克拉河马州立大学
美国

APPENDIX D-E: WEB INTRODUCTION SCRIPT (ENGLISH VERSION)

Inventory of School Motivation Survey

The survey is to measure college students' achievement motivation using Inventory of School Motivation. The inventory was originally developed in the western culture. In this study, data will be collected in different cultures to test whether the instrument can be used with similar psychometric properties.

This survey will take approximately fifteen minutes to complete. Your participation is strictly voluntary. Participation in the survey indicates that you are at least 18 years old. The data will be stored in a password protected personal computer and only the researcher involved in this study have the access to them and they will be destroyed in five years.

There is no risk to those responding to this survey. The information collected cannot in any way be traced to respondents, as the survey design program used to build this instrument is not capable of tracking respondents or tying information to individual participants.

If you have any questions or concerns, please contact me or my dissertation advisor. Our contact information is listed below. It is recommended that you please print a copy of this page for future reference.

If you have questions about your rights as a research volunteer, you may contact Dr. Shelia Kennison, IRB Chair, 219 Cordell North, Stillwater, OK 74078, 405-744-1676 or irb@okstate.ed.

To participate in the survey, please click Agree to Participate button.

Thank you for your willingness to participate in this research study.

Lihua Xu
Primary Investigator
802 W. Highpoint Dr., Apt. 8
Stillwater, OK 74075
405-612-2450
lihua.xu@okstate.edu

Laura Barnes, Ph.D.
Dissertation Advisor
2444 Main Hall, OSU-Tulsa
700 N. Greenwood Ave.
Tulsa, OK 74106
918-594-8517
laura.barnes@okstate.edu

APPENDIX D-C: WEB INTRODUCTION SCRIPT (CHINESE VERSION)

学习动机调查

请在回答前仔细阅读

非常感谢您参与此次问卷调查。该问卷是调查学生的学习动机，是在国外编制的。此次在国内收集数据，目的是比较中西文化下学生对该动机问卷的诠释是否一致，具体讲，是想测试该问卷是否有类似的心理测量学系数。这对跨文化研究学习动机和学习成绩很有意义。

此次问卷调查不询问被调查者的名字，整个问卷完成大概需要 15 分钟时间。问卷参与纯属自愿，参与者必须在 18 周岁及以上。参与者可以随时中断该调查。调查数据结果严格保密，只有研究者本人可以看到。该数据会保存五年时间，之后会销毁。问卷参与不会引起比日常生活中遇到的更大的危害。

如果您对该问卷调查需要进一步的信息，请通过电子邮件 lihua.xu@okstate.edu 联系论文作者，或论文指导老师 （lbarnes@okstate.edu）。作为问卷参与者，如果您对自己的权利有任何问题，请联系奥克拉河马州立大学的审查委员会。 该会的电子邮件地址为 irb@okstate.edu。

非常感谢您的参与，谢谢您的宝贵时间。

Lihua Xu
Primary Investigator
802 W. Highpoint Dr., Apt. 8
Stillwater, OK 74075
405-612-2450
lihua.xu@okstate.edu

Laura Barnes, Ph.D.
Dissertation Advisor
2444 Main Hall, OSU-Tulsa
700 N. Greenwood Ave.
Tulsa, OK 74106
918-594-8517
laura.barnes@okstate.edu

APPENDIX D-M: INTRODUCTION TO WEB SURVEY -MODIFIED

The purpose of this study is to measure college students' achievement motivation using Inventory of School Motivation. The inventory was originally developed in the western culture. In this study, data will be collected in China and the United States to test whether the instrument can be appropriately used in both cultures.

This survey will take approximately fifteen minutes to complete. Your participation is strictly voluntary and you can choose to terminate the survey at any time without penalty. Participation in the survey indicates that you are at least 18 years old. Data from this study will be stored in a password protected personal computer and only the researcher involved in this study will have the access to them and they will be destroyed in five years.

Participation in this survey fulfills the course credit requirement in some courses. A small amount of course credit (0.5 unit) is offered for your participation. This course credit requirement may be fulfilled alternatively in two other ways: 1) attending Undergraduate Research Colloquia, or 2) researching and writing 3-4 page papers on designated research topics.

There are no known risks associated with this project which are greater than those ordinarily encountered in daily life. The information collected cannot in any way be traced to respondents, as the survey design program used to build this instrument is not capable of tracking respondents or tying information to individual participants.

If you have any questions or concerns, please contact me or my dissertation advisor. Our contact information is listed below. If you have questions about your rights as a research volunteer, you may contact Dr. Shelia Kennison, IRB Chair, 219 Cordell North, Stillwater, OK 74078, 405-744-1676 or irb@okstate.edu.

Thank you for your willingness to participate in this research

Lihua Xu
Primary Investigator
802 W. Highpoint Dr., Apt. 8
Stillwater, OK 74075
405-612-2450
lihua.xu@okstate.edu

Laura Barnes, Ph.D.
Dissertation Advisor
2444 Main Hall, OSU-Tulsa
700 N. Greenwood Ave.
Tulsa, OK 74106
918-594-8517
laura.barnes@okstate.edu

APPENDIX E-E: SURVEY INSTRUMENT (ENGLISH VERSION)

INVENTORY OF SCHOOL MOTIVATION

Please respond the following items on a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, 5=strongly agree).

TA1. I like being given the chance to do something again to make it better
TA9. I try harder with interesting work
TA17. I like to see that I am improving in my schoolwork
TA26. I need to know that I am getting somewhere with my schoolwork

EF2. I don't mind working a long time at schoolwork that I find interesting
EF10. I try hard to make sure that I am good at my schoolwork
EF18. When I am improving in my schoolwork I try even harder
EF27. The harder the problem the harder I try
EF34 I try hard at school because I am interested in my work
EF42. I work hard to try to understand new things at school
EF49. I am always trying to do better at my schoolwork

CM3. Winning is important to me
CM11. Coming first is very important to me
CM19. I like to compete with others at school
CM28. I work harder if I'm trying to be better than others
CM35. I want to do well at school to be better than my classmates
CM44. I am only happy when I am one of the best in class

SP4. I work hard at school so that I will be put in charge of a group
SP12. I want to feel important in front of my school friends
SP20. At school I like being in charge of a group
SP30. It is very important for me to be a group leader
SP36. I work hard at school because I want the class to notice me
SP46. I often try to be the leader of a group

AF5. I do my best work at school when I am working with others
AF13. I try to work with friends as much as possible at school
AF21. I prefer to work with other people at school rather than alone

SC6. It is very important for students to help each other at school
SC14. I like to help other students do well at school
SC22. I care about other people at school
SC31. I enjoy helping others with their schoolwork even if I don't do so well myself
SC37. It makes me unhappy if my friends aren't doing well at school

PR7. Praise from my teachers for my good schoolwork is important to me

PR15. Praise from my friends for my good schoolwork is important to me
PR24. At school I work best when I am praised
PR32. I want to be praised for my good schoolwork
PR38. Praise from my parents for good schoolwork is important to me

TK8. I work best in class when I can get some kind of reward
TK16. I work hard in class for rewards from the teacher
TK25. I work hard at school for presents from my parents
TK33. Getting a reward for my good schoolwork is important to me
TK40. Getting merit certificates helps me work harder at school
TK47. Praise for good work is not enough I like a reward
TK51. If I got rewards at school I would work harder

DEMOGRAPHIC INFORMATION

1. Gender: male__ female__
2. Age: below 18 years old__ 18-22 years old__ 23-27 years old__ 28-32 years old__ over 32 years old__
3. Race/ethnicity: White __ African American__ Native American ____ Asian American____ International____
4. What is your classification in college: freshman__ sophomore__ junior__ senior __graduate __
5. What is your major by college: arts_ sciences__ engineering __ agriculture __ business __ education __not decided yet__ other, please specify__
6. Is English your first language: yes__ no__
7. What is your highest degree planned at college: Bachelor's degree__ Master's degree__ Doctorate degree__
8. What is your ACT/SAT score __
9. What is your average grade in high school: A__ B__ C__ D__ F__
10. What is your average grade in college: A__ B__ C__ D__ F__
11. What is your place of original residence: rural__ suburban__ urban__
12. Your parents' estimated annual income: below $10,000___ $10,001- 39,999__ $40,000-74,999__$75,000 or more__ don't know
13. Your parents' education:

a. Father：high school and below__ college__ post-college__

b. Mother: high school and below__ college__ post-college __

14. Your parents' occupation:

a. Father: professional (e.g. architects, engineers, teachers etc.)__ clerical and sales (e.g. public servants, typist etc.)__ skilled (e.g. manager, computer technician, electrician etc.)__, semi-skilled (farm worker, clothing factor worker etc.)__, unskilled (e.g. truck driver, laborer, etc.)__ unemployed__, other__

b. Mother: professional (e.g. architects, engineers, teachers etc.)__ clerical and sales (e.g. public servants, typist etc.)__ skilled (e.g. manager, computer technician, electrician etc.)__, semi-skilled (farm worker, clothing factor worker etc.)__, unskilled (e.g. truck driver, laborer, etc.)__, unemployed__, other__

APPENDIX E-C: SURVEY INSTRUMENT (CHINESE VERSION)

学习动机量表

请根据自己的个人情况，据实回答下列问题（选项：完全不同意 不同意 中立 同意 完全同意）

TA1. 我喜欢有机会同一事情再做一次，以便做得更好。
TA9. 对于喜欢得工作，我会更卖力地做。
TA17. 我喜欢看到自己的功课有进步。
TA26. 我需要知道自己的功课有所进展。

EF2. 对于有意思的功课，我不介意长时间地学习。
EF10. 我会尽量把功课学好。
EF18. 如果我的功课有进步，我会更加努力。
EF27. 问题越难，我越努力。
EF34 我努力学习，因为我对功课感兴趣。
EF42. 我努力去理解在学习上的新东西。
EF49. 我总是尽量把功课做得更好。

CM3. 赢对我来说很重要。
CM11. 取的第一对我来说很重要。
CM19. 我喜欢在学习上与他人竞争。
CM28. 如果我想比别人强时，我会更努力。
CM35. 我想功课好，以便超过我的同学。
CM44. 我只有在班里是尖子生的时候才会高兴。

SP4. 我认真学习，以便别人会让我管理一个小组。
SP12. 我想在同学面前觉得了不起。
SP20. 我喜欢在学习上担任小组领导。
SP30. 能够成为小组领导对我来说很重要。
SP36. 我在学习上很努力，因为我想同学注意到自己。
SP46. 我经常努力成为小组领导。

AF5. 我同别人一起学习时成绩最好。
AF13. 我在学校尽可能多跟朋友学习。
AF21. 在学校我更喜欢跟别人一起学习而不是单独学习。

SC6. 在学校里同学之间互相帮助很重要。
SC14. 我喜欢帮助其他同学取得好成绩。
SC22. 在学校我关心别人。
SC31. 即使我自己的功课不是很好，我也喜欢帮助别人学习功课。
SC37. 如果我的朋友学习成绩不好，我会不开心。

PR7. 因为功课好受到老师的表扬对我来说很重要。
PR15. 因为功课好受到朋友的表扬对我来说很重要。
PR24. 在学习上如果我受表扬时，我学得最好。
PR32. 我想别人表扬我功课学得好。
PR38. 因为功课好受到父母的表扬对我来说很重要。

TK8. 当我得到某种奖励时，我在班上学得最好。
TK16. 我在班上努力学习以便得到老师的奖励。
TK25. 我在学校努力学习以便得到父母的礼物。
TK33. 因成绩好而得到奖励对我来说很重要。
TK40. 得到奖状会使我更努力学习。
TK47. 对好成绩光表扬是不够的，我喜欢奖励。
TK51. 如果在学习上得到奖励，我会更努力。

个人信息：

1. 您的性别： 男__女__
2. 请问您的年龄是（周岁）： 18 岁以下__ 18-22 岁__ 23-27 岁__ 28-32 岁__32 岁以上__
3. 民族：汉族__ 少数民族__ 外国学生__
4. 你所在年级：大一 __ 大二__ 大三__ 大四__研究生__
5. 你的专业是：文科__，理科__，工科__，商科__，教育__，没有确定____，其他，请注明__
6. 请问汉语是你的母语吗：是__ 否__
7. 请问你准备读的最高学位是：大学本科学位__硕士学位__博士学位__
8. 请问你的高考成绩是：__
9. 请问你高中的平均成绩：60 分及以下__61-70 分__ 71-80 分__ 81-90 分__ 91-100 分__
10. 请问你大学的平均成绩：60 分及以下__61-70 分__ 71-80 分__ 81-90 分__ 91-100 分__
11. 请问你家庭居住地是：农村__ 城市郊区__ 城市市区____
12. 请问你的父母年收入是：10000 元以下__ 10001-39999 元__ 40000-74999 元__ 75000 以上__不清楚__
13. 请问你父母的最高学历是（包括在读学历）：
    a. 父亲：初中及以下__高中/中专/技校__大学本科__硕士及以上__
    b. 母亲：初中及以下__高中/中专/技校__大学本科__硕士及以上__
14. 请问你父母的职业是：
    a. 父亲：立法者、高级官员和管理人员__ 专业人员__技术和辅助专业人员__职员__服务人员和商店与市场销售人员__农业和水产业技术工人__手（工）艺人和有关行业的工人__设备与机械操作工和装配工_简单劳动职业者_军队_
    b. 母亲：立法者、高级官员和管理人员__ 专业人员__技术和辅助专业人员__职员__服务人员和商店与市场销售人员__农业和水产业技术工人__手（工）艺人和有关行业的工人__设备与机械操作工和装配工_简单劳动职业者_军队_

APPENDIX F: TRANSLATION EVALUATION FORM

Please read the two sentences from the left and rate its comparability and interpretability to your right for each question.

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
TA1	I like being given the chance to do something again to make it better	I like doing the same thing one more time if given the opportunity, so that I can do better	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TA9	I try harder with interesting work	I will try harder at things that I like	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TA17	I like to see that I am improving in my schoolwork	I like to see progress in my coursework.	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TA26	I need to know that I am getting somewhere with my schoolwork	I need to know that I am making progress in my coursework.	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF2	I don't mind working a long time at schoolwork that I find interesting	I don't mind working long hours on interesting subjects	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF10	I try hard to make sure that I am good at my schoolwork	I will try my best to do a good job in my coursework	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF18	When I am improving in my schoolwork I try	I will work harder if I have made progress on my school work	1	2	3	4	5	6	7	1	2	3	4	5	6	7

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
	even harder															
EF27	The harder the problem the harder I try	The more difficult the questions are, the harder I try	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF34	I try hard at school because I am interested in my work	I study hard because I am interested in the school work	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF42	I work hard to try to understand new things at school	I try hard to understand new things in studies	1	2	3	4	5	6	7	1	2	3	4	5	6	7
EF49	I am always trying to do better at my schoolwork	I always try my best to have a better academic performance	1	2	3	4	5	6	7	1	2	3	4	5	6	7
CM3	Winning is important to me	Winning is important to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7
CM11	Coming first is very important to me	To be number one is very important to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7
CM19	I like to compete with others at school	I like to compete against others in study	1	2	3	4	5	6	7	1	2	3	4	5	6	7
CM28	I work harder if I'm trying to be better than others	I will try harder when I want to surpass others	1	2	3	4	5	6	7	1	2	3	4	5	6	7
CM35	I want to do well at school to be better than	I want to be good at my coursework so that I can outdo my	1	2	3	4	5	6	7	1	2	3	4	5	6	7

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
	my classmates	classmates														
CM44	I am only happy when I am one of the best in class	I am happy only when I am one of the top students in class	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP4	I work hard at school that I will be put in charge of a group	I work hard so that people let me be a team leader	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP12	I want to feel important in front of my school friends	I want to feel great in front of my classmates	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP20	At school I like being in charge of a group	I like to be a group leader in study	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP30	It is very important for me to be a group leader	Being able to become a group leader means a lot to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP36	I work hard at school because I want the class to notice me	I work very hard in my study because I want to draw my classmates' attention	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SP46	I often try to be the leader of a group	I often try hard to become a group leader	1	2	3	4	5	6	7	1	2	3	4	5	6	7
AF5	I do my best work at school when I am working with others	I perform the most when I study with others	1	2	3	4	5	6	7	1	2	3	4	5	6	7
AF13	I try to work with	I try to study with my friends at														

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
	friends as much as possible at school	school as much as possible	1	2	3	4	5	6	7	1	2	3	4	5	6	7
AF21	I prefer to work with other people at school rather than alone	At school I like to study with others rather than by myself	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SC6	It is very important for students to help each other at school	It is very important that students help each other at school	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SC14	I like to help other students do well at school	I like helping other fellow classmates to achieve better grade	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SC22	I care about other people at school	I care about others at school	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SC31	I enjoy helping others with their schoolwork even if I don't do so well myself	I like to help others with their coursework even though I am not very good myself	1	2	3	4	5	6	7	1	2	3	4	5	6	7
SC37	It makes me unhappy if my friends aren't doing well at school	I won't be happy if my friends do not have a good grade	1	2	3	4	5	6	7	1	2	3	4	5	6	7
PR7	Praise from my teachers for my good schoolwork is important to me	Getting praise from teachers because of my excellence in studies means a lot to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
PR15	Praise from my friends for my good schoolwork is important to me	Getting praise from friends because of my excellence in studies means a lot to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7
PR24	At school I work best when I am praised	I learn the best when I receive praise about my academic performance	1	2	3	4	5	6	7	1	2	3	4	5	6	7
PR32	I want to be praised for my good schoolwork	I want to be praised for my good academic performance	1	2	3	4	5	6	7	1	2	3	4	5	6	7
PR38	Praise from my parents for good schoolwork is important to me	Getting praise from parents because of my excellence in studies means a lot to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK8	I work best in class when I can get some kind of reward	I learn the best when I receive a certain award	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK16	I work hard in class for rewards from the teacher	I work hard in my class to win my teacher's award	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK25	I work hard at school for presents from my parents	I work hard at school to win my parents' presents	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK33	Getting a reward for my good schoolwork is important to me	Winning an award for my good academic performance means a lot to me	1	2	3	4	5	6	7	1	2	3	4	5	6	7

			Comparability of Language (Form)							**Similarity of Interpretability (Meaning)**						
	Original English Version	**Back Translation**	Extremely comparable=1 Moderately comparable=4 Not at all comparable=7							Extremely similar=1 Moderately similar=4 Not at all similar=7						
TK40	Getting merit certificates helps me work harder at school	Winning an award certificate makes me study harder	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK47	Praise for good work is not enough I like a reward	Praise alone is not enough. I like getting an award for good academic performance	1	2	3	4	5	6	7	1	2	3	4	5	6	7
TK51	If I got rewards at school I would work harder	I will make greater efforts if I get awarded for my study	1	2	3	4	5	6	7	1	2	3	4	5	6	7

VITA

LIHUA XU

Candidate for the Degree of Doctor of Philosophy

Thesis: MEASUREMENT INVARIANCE OF INVENTORY OF SCHOOL MOTIVATION BETWEEN THE UNITED STATES AND CHINESE COLLEGE STUDENTS

Major Field: Educational Psychology with emphasis on Research and Evaluation

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Educational Psychology at Oklahoma State University, Stillwater, Oklahoma in December, 2009.

Completed the requirements for the Master of Arts in English Language and Literature at Sun Yat-Sen University, Guangzhou, Guangdong in 2003.

Completed the requirements for the Bachelor of Arts in English at Sichuan University, Chengdu, Sichuan in 1998.

Experience:

Graduate research associate in the office of University Assessment and Testing from 2007 to present; Graduate research associate in the College of Education from 2005 to 2007; Graduate teaching assistant in Department of English from 2003 to 2004.

Professional Memberships:

AEA, AERA, NCME

Name: Lihua Xu Date of Degree: December, 2009

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: MEASUREMENT INVARIANCE OF INVENTORY OF SCHOOL MOTIVATION BETWEEN THE UNITED STATES AND CHINESE COLLEGE STUDENTS

Pages in Study: 138 Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Emphasis: Research, Evaluation, Measurement & Statistics

Scope and Method of Study: The study was to test measurement equivalence of the *Inventory of School Motivation* (ISM) between American and Chinese college students. To achieve this, an eight-factor model was tested for comparability across the two cultural groups using both single-group confirmatory factor analysis (CFA) and multi-group confirmatory factor analysis. In addition, contingent upon the degree of invariance, cross-cultural comparisons of latent means were conducted.

Findings and Conclusions: Results of CFA showed that the eight-factor model fit each cultural group was a borderline fit, yet deemed acceptable as indicated by various model fit indices. Results of multi-group CFA showed the subscales of Effort, Social Concern, Affiliation, and Praise have configural, metric and scalar invariance. The Task subscale has configural and metric invariance. Subscales of Competition, Social Power and Token did not show configural invariance. Therefore, the ISM did not perform equivalently cross-culturally. Results of cross-cultural latent mean comparisons indicate that Chinese participants scored significantly higher than American peers on the latent constructs Effort and Social Concern. The two groups did not show significant differences on Affiliation. American participants scored significantly higher than Chinese counterparts on Praise. Findings support the need to establish measurement equivalence prior to interpreting differences in means between culturally diverse groups.

ADVISER'S APPROVAL: DR. LAURA BARNES