Review of the Motivated Strategies for Learning Questionnaire (MSLQ) Using Reliability Generalization Techniques to Assess Scale Reliability

by

Robin Terrell Taylor

A dissertation submitted to the Graduate Faculty of
Auburn University
in Partial Fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 6, 2012

Keywords: Motivated Strategies for Learning Questionnaire (MSLQ), reliability generalization

Approved by

Margaret Ross, Chair, Professor & Educational Research Methodology and Analysis (ERMA)
Program Coordinator
David Shannon, Humana-Germany-Sherman Distinguished Professor, Educational, Foundations
Leadership and Technology
Sean Forbes, Associate Professor of Educational Foundations, Leadership and Technology

Abstract

Reliability generalization studies were conducted on the motivation and learning strategies scales of the Motivated Strategies for Learning Questionnaire (MSLQ) to typify score reliabilities for all scales on the instrument and to examine potential sources of measurement error across studies which used these scales. Average reliability coefficients range from a low of .61 for the learning strategies scale, help seeking, and a high of .88 for the motivation scale, self-efficacy of learning and performance. Overall, results of reliability generalization studies for both the motivation and learning strategies sections of the MSLQ demonstrate that the MSLQ can be used across a variety of different samples with reasonable confidence for obtaining generally reliable scores.

Acknowledgements

First, I offer thanksgiving to God for the blessings in my life, including the opportunity to pursue this degree. Second, I wish to thank my advisor and committee chair, Dr. Margaret (Betsy) Ross, for her encouragement, mentorship, and support throughout my academic endeavors at Auburn University. I am especially thankful for her guidance during the writing of this report. I also thank my committee members, Drs. David Shannon and Sean Forbes for their professional insights and thought-provoking instruction during my educational experiences at the university. In addition I would like to recognize the contribution of library staff members at Auburn University who were very supportive in helping locate and secure such a large number of manuscripts needed to complete this research.

I would also like to acknowledge the endless encouragement and support my family and friends provided throughout my pursuance of this degree. Their prayers, support, and confidence encouraged my commitment to complete this work. I especially want to thank my husband's mother, Virginia Taylor, and my mother, Donna Bevins, for their sacrifices to watch my son in order that I could maintain and often times regain focus on this research. Most importantly, I wish to recognize the contributions and sacrifices of my best friend and husband, Jason A. Taylor. His constant support, faith in my abilities, and endless encouragement allowed me to accomplish such an undertaking. I doubt I would have ever considered such an endeavor without his belief in my ability and encouragement to find and do something that I love. Thank you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations or Symbols

IG	intrinsic goal orientation
EG	extrinsic goal orientation
TV	task value
CB	control of learning beliefs
SE	self-efficacy for learning and performance
TA	test anxiety
R	rehearsal
E	elaboration
O	organization
CT	critical thinking
MSR	metacognitive self-regulation
TSM	time and study management
ER	effort regulation
PL	peer learning
HS	help seeking
RG	reliability generalization
M	mean
SD	standard deviation
W	Shapiro-Wilk test for normality statistic

p	p-value
η^2	eta squared
ω^2	omega squared
b	Beta coefficient
r_s	structure coefficient or Pearson's correlation coefficient
R^2	coefficient of determination
Adj. R^2	Adjusted coefficient of determination

CHAPTER I. INTRODUCTION

Validity and reliability are tools for addressing issues of an imperfect instrument in such a way that results from scores on these assessments help researchers accurately describe and understand constructs of interest. Constructs can be defined as "theoretical constructions, abstractions, aimed at organizing and making sense of our environment," (Pedhazur & Schmelkin, 1991, p. 52). Tests and instruments are used as measurement tools to allow researchers to quantify a construct of interest indirectly through observable variables. Reliability, often considered the accuracy of a measurement procedure, refers to the consistency and reproducibility of scores on the measurement procedure (Crocker & Algina, 1986). Scores on an instrument must demonstrate acceptable levels of reliability before validity, the degree in which scores provide meaningful interpretations for a construct of interest, can be verified (Crocker & Algina, 1986; Pehazur & Schmelkin, 1991; Nunnally, 1978).

Often social-behavioral scientists create tests to measure constructs like intelligence or learning and instruments to measure constructs such as motivation, self-efficacy, test anxiety, or self-regulated learning strategies. A common instrument, the Motivated Strategies for Learning Questionnaire (MSLQ), is often used to assess fifteen different scales related to components of motivation and self-regulated learning strategy use. The MSLQ is economically feasible to administer, readily available within the public domain, easily scored, and includes interpretations for individual student profiles. As such, a number of researchers have used the MSLQ for various research purposes.

The MSLQ, like most tests and instruments, includes reliability and validation data to help users of the instrument determine the appropriateness for using the instrument to measure these constructs. In Benson's (1998) review of the MSLQ, he stated, "Additional reliability and validity studies are needed to improve the subscale accuracy and to substantiate the dimensionality of the scale". Vacha-Haase (1998) introduced reliability generalization as an additional means for establishing typical reliabilities for measurement scales as well as exploring the generalizability of reliabilities across different study populations. The purpose of this dissertation research is to explore the generalizability and adequacy of sample specific reliabilities for both the motivation and learning strategies sections of the MSLQ. In addition, the construct validity of the MSLQ will be reviewed using Messick's (1995) framework of construct validity.

Reliability Generalization

Reliability generalization is a meta-analytic technique intended to assess the extent reliability scores vary across studies. Reliability generalization methods allow researchers to determine likely values for the reliability of a scale and to identify salient features of samples which may contribute to variations in reliabilities achieved across studies. Reliability generalization studies have likely advanced due to the poor reporting practices of researchers to address reliability for scores used in their studies.

Since Vacha-Haase's (1998) introduction to reliability generalization, an influx of studies have been used to explore the reliabilities achieved across different samples for numerous well-established psychological and socio-behavioral instruments. Reliability generalization studies can indicate groups for which instruments may be more or less suited;

which can have implications for improving an instrument or for adapting the instrument for different examinee populations.

Motivated Strategies for Learning Questionnaire Instrument Description

The MSLQ is a self-report instrument that includes 81 items developed to measure students' motivation orientations and use of learning strategies. The MSLQ is partitioned into a motivation section and a learning strategies section. The motivation section is comprised of three components: a value component which includes scales of intrinsic goal orientation, extrinsic goal orientation, and task value; an expectancy component which includes scales for control of learning beliefs and self-efficacy for learning and performance; and an affective component which includes a scale for test anxiety. The learning strategies section includes two components: a cognitive and metacognitive strategies component which includes scales for rehearsal, elaboration, organization, critical thinking, and metacognitive self-regulation; and a resource management strategies component which includes scales for time and study environment, effort regulation, peer learning, and help seeking (Pintrich, Smith, Garcia & McKeachie, 1991; Pintrich, Smith, Garcia & McKeachie, 1993).

The instrument was designed to be used with post-secondary students, and data presented in the manual (Pintrich et al., 1991) are based on a sample of 380 college students (356 students from a public 4-year university and 24 students attending a community college) within 37 classrooms covering 14 subject domains and 5 disciplines (natural science, humanities, social science, computer science, and foreign language). Although the MSLQ has been used widely by researchers measuring the motivation and learning strategy use of college students, many researchers have also employed the MSLQ to measure motivation and learning strategies use for students at the elementary, middle school and high school levels.

Instrument Validation

The instrument was validated over several waves of data collection. The authors used confirmatory factor analysis to estimate parameters and test the utility of the theoretical models for both motivation and learning strategy subscales (Pintrich et al., 1993). Confirmatory factor analyses allows the developers to specify which items are to be allocated to which factor. Fit indices were used to assess the fit between the observed data and the theoretical underpinnings of each scale. Cronbach's alpha was used to estimate the internal consistency for each of the fifteen MSLQ subscales. Alphas ranged from .52 for the help seeking scale to .93 for the self-efficacy scale. The developers of the instrument claim that the alpha coefficients for the MSLQ scales are robust and demonstrate good internal consistency (Pintrich et al., 1993).

Definition of Terms

Motivation. Motivation is grounded within various theoretical frameworks. Within the learning process, motivation is generally concerned with the learner's internal drive to succeed in academic tasks, and is often termed achievement motivation. Research for achievement motivation focuses in explaining why a learner chooses, expends effort, and persists on learning tasks (Dweck, 1999; Nicholls, 1984).

Intrinsic Goal Orientation. Goal orientation refers to why a learner engages in an academic task. Learners with intrinsic goal orientations possess real interest in the learning process and aspire to increase their knowledge of the subject matter (Dweck & Leggett, 1988).

Extrinsic Goal Orientation. Extrinsic goal orientation describes learners interest in engaging in a task due to causes outside the individual, such as to demonstrate their ability, to outperform others, and/or to receive some external benefit such as getting good grades, recognition, or a reward (Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997; Pintrich, 2000a).

Task Value. Task value refers to an individual's appreciation for a task's relevance. Task value relates to the degree of personal interest a learner has for a given task and includes beliefs about utility, relevance, and importance (Raynor, 1981; Schunk, 1991).

Control of Learning Beliefs. Control of learning beliefs are similar to concepts of locus of control, which infer the amount of control learners perceive they have over their learning ability (Rotter, 1966). Bandura (1994) defined control of learning beliefs as individuals exercising influence over their own motivation, cognition, affect, and behaviors.

Self-Efficacy of Learning and Performance. In general, self-efficacy refers to a person's judgments of their capabilities to perform an action successfully. Academic self-efficacy applies this general definition of efficacy to one's internal belief for executing and succeeding in academic tasks at designated success levels (Bandura 1986, 1989, 1994; Schunk, 1991).

Test Anxiety. Test anxiety is defined as an unpleasant feeling or emotional state manifested in a learner's performance on tests or other cognitive measures (Pintrich et al., 1991; Zeidner, 1998).

Self-Regulated Learning. Self-regulation refers to a student's proactive use of specific learning strategies to achieve learning goals. Students who practice self-regulative behaviors initiate and direct their learning effort to successfully acquire new knowledge (Pintrich 2000b, 2004; Zimmerman 1986, 2000)

Rehearsal. Rehearsal is considered a simple learning strategy which stores information to be learned into working memory through processes of naming, repeating, and reciting material for learning (Weinstein & Mayer, 1986).

Elaboration. Elaboration is a learning strategy in which a learner paraphrases or summarizes learning material to help the individual understand the material. This strategy is intended to build internal connections between one's prior knowledge and the new material. This strategy is considered a higher order learning skill because the strategy allows learners to store learned information into long-term memory (Weinstein & Mayer, 1986).

Organization. Organization, a higher order learning strategy, involves methods of outlining, taking notes, mapping or connecting key ideas in learning material (Weinstein & Mayer, 1986).

Critical Thinking. Critical Thinking is a higher order learning strategy which involves applying learned information to knowledge of new situations, i.e., relating subject matter to be learned to one's prior, personalized knowledge (Scriven & Paul, n.d.).

Metacognition. Metacognition refers to how one thinks about thinking, it encompasses methods of a learner's awareness and knowledge of their cognitive processes (Flavell, 1979, 1992; Pintrich, 2002).

Time and Study Management. Time and study management involves choosing environments that are conducive to learning (i.e., free from distractions) and effectively scheduling, planning, and managing one's study time (McKeachie, Pintrich, Lin & Smith, 1986; Pintrich et al., 1991).

Effort Regulation. The effort a student expends to reach his or her learning goals is termed effort regulation. Effort regulation enhances the ability of the learner to handle setbacks and failures within the learning process by correctly allocating resources and appropriate effort to increase more successful learning in the future (Chen, 2002).

Peer Learning. Peer learning involves using peers (friends, classmates, etc.) to collaboratively understand course material or information to be learned (Jones, Alexander & Estell, 2010).

Help Seeking. Help seeking can be an adaptive learning strategy that allows a learner to optimize learning by seeking help from local resources such as instructors, peers, tutors, or even additional textbooks (Ames, 1983; Karabenick, 1998; Zimmerman & Martinez-Pons, 1986).

Research Questions

- What is the typical reliability estimate for each scale on the motivation section of the MSLQ and how do the coefficients across studies vary from this estimate?
- Do study- or sample-specific variables affect reliability coefficients of MSLQ motivation scales?
    - Are there differences between reliability coefficients produced for motivation scales when researchers framed questions specific to a course subject versus asking respondents to generalize their motivation across the learning domain?
    - Are there differences between reliability coefficients produced for motivation scales when researchers used the questionnaire with college student populations versus using the instrument with younger student populations?
    - Are there differences between reliability coefficients produced for motivation scales when researchers used the questionnaire with student populations from the United States versus populations from other countries?
- What is the typical reliability estimate for each scale on the learning strategies section of the MSLQ and how do the coefficients across studies vary from this estimate?
- Do study- or sample-specific variables affect reliability coefficients of MSLQ learning strategies scales?

- Are there differences between reliability coefficients produced for learning strategies scales when researchers framed questions specific to a course subject versus asking respondents to generalize their learning strategy use across the learning domain?
- Are there differences between reliability coefficients produced for learning strategies scales when researchers used the questionnaire with college student populations versus using the instrument with younger student populations?
- Are there differences between reliability coefficients produced for learning strategies scales when researchers used the questionnaire with student populations from the United States versus populations from other countries?

Methodology

A literature search was conducted to retrieve any manuscript (journal article, book chapter, dissertation, conference proceeding, or other types) in which the author completed a substantive study using the MSLQ to collect data. Similar to meta-analytic techniques, each article retrieved was reviewed to a) determine if the study was an empirical study using the instrument, b) record the reporting of reliabilities for the research sample, and c) record study specific characteristics of each sample to use for statistical analyses. Data was recorded using Microsoft Excel 97 and all statistical procedures were completed using SPSS 19.

Limitations

A leading limitation of many reliability generalization studies completed to date is the exclusion of reliability coefficients in the studies. Reasons for why reliabilities were excluded from both reliability generalization studies completed for the motivation and learning strategies scales of the MSLQ include: a) studies that were reviewed but did not include reliability coefficients for their samples; b) studies that were reviewed and included reliability coefficients for their samples, but did not report reliabilities at the subscale level; and c) studies that could not be reviewed because studies could not be obtained. Rosenthal (1979) termed the tendency to publish research with positive results the ‘file drawer problem’. Rosenthal suggests that there may be numerous studies omitted from publication because of negative, or nonsignificant results. The generalizability of results was also limited due to the pervasive amount of missing data across sample descriptors.

A list was prepared of all ‘hits’ including the instrument name, *Motivated Strategies for Learning Questionnaire*, and its abbreviation, *MSLQ*. All attempts were made to locate each article included on the list. Articles that were not written in English or those that the lending library could not find were excluded from analyses. The list also contained a large number of dissertations. All dissertations that could be obtained free of charge were included in the study. In addition, approximately 20 more dissertations were purchased or rented to review how the MSLQ was used and record reliability data when appropriate. However, several dissertations were not reviewed either because the lending library did not allow the dissertation to circulate, or the cost of purchasing additional dissertations became prohibitive. All sample characteristics provided by authors were recorded to be used in the reliability generalization studies.

CHAPTER II. A REVIEW OF THE MOTIVATED STRATEGIES FOR LEARNING QUESTIONNAIRE

Introduction

Emphasis on learning and how educational institutions equip students to be productive workers in today's global and technologically advanced economy continues to be a major focal point for society. Today, technology is constantly advancing, and workforce demands require that employees continually adapt and increase their knowledge and skills set to compete in the market. Although institutions of formal education are still widely seen as instrumental in equipping students with the knowledge and skills for the workforce, modern views also emphasize the need to equip individuals to serve as active agents in their own learning process (Birenbaum & Rosenau, 2006; de la Harpe & Radloff, 2000; Lin, McKeachie & Kim, 2003). Individuals with a commitment to lifelong learning are apt to have an advantage in today's competitive marketplace.

Learning is simplistically perceived as the acquisition of knowledge. However, learning is a complex process and defining when learning has occurred remains difficult as the act itself is not often observable. Definitions of learning include Wittrock (1977) who defined learning as "processes involved in changing through experience. It is the process of acquiring relatively permanent change in understanding, attitude, knowledge, information, ability and skill through experience" (p. 9). Lefrancois (1995) defined learning as "all relatively permanent changes in behavior that result from experience but are not attributable to fatigue, maturation, drugs, injury, or disease" (p. 5). Gagne (1977) claimed, "Learning is a change in human disposition or capability, which persists over a period of time, and which is not simply ascribable to processes of growth" (p. 3).

Through research, many psychologists and education specialists have attempted to understand how human beings acquire and apply learned knowledge. Theoretical perspectives of learning have gradually progressed over time. During the start of the twentieth century, learning research was heavily influenced by the behavioral tradition of psychology. The behaviorist movement focused on explaining phenomena through empirical research and advanced as early psychologists strived to strengthen psychology as a science. Cognitivism advanced during the 1950s and replaced behaviorism as the major paradigm of understanding mental functions. Cognitivist theorists purport that learning involves more complex mental associations which are not reflected within overt behavior changes (Bredo, 1997; Shuell, 1986). Around the 1970s, constructivist theorists began to contend that learning is also a major function of social and cultural influences (Bredo, 1997).

These different concepts of learning became popular according to dominant interests of each era. The importance and relevance of each theory might best be portrayed in Bredo's words, "[These learning theories] are tools for organizing learning experiences in ways that are thought to be better in one way or another. Organizing learning based on one conception or another is likely to teach different meta-lessons about the nature of learning" (p. 39). Clayton (1965) also provides a summary for the implications of each theoretical perspective:

> Theories of learning are embodiments or applications of conceptions regarding the nature of mind. The history of education bears testimony to the fact that influential theories of the mind translate themselves at some point into educational practice. If we assume that the

> mind exists antecedently waiting to be trained, the natural result of this assumption is formal discipline. Or if we take for granted that the mind consists of a collection of impressions or mental states, we may then easily exalt the role of the teacher and formalize the process of instruction. In the one case the mind is, indeed, a source of energy or power, but the primary use of subject matter is to serve as gymnastic material. In the other case the emphasis falls primarily on the acquisition and organization of material, but with little regard for the development of individual capacity and interest. In both cases the conception of mind that is basic to the corresponding educational practice tends to set the mind apart as something to be trained or moulded. The selection and organization of subject matter is not determined by a purpose or aim that the learner is seeking to realize, but is imposed from without; with the result that education becomes formalized. If we abandon this ancient dualism, we limit the native equipment of the individual to a certain set or group of inborn tendencies or impulses, as determined by the structure of the nervous system. (pp. 266-267)

The current perspective for learning is framed within social-cognitive theory. Social-cognitive theory considers both the complexities of our mental processes and social influences while also stressing the role of the individual within the learning process. Social cognitive theory is generally credited to Albert Bandura and his research on social behaviors. The conceptual framework of social cognitive theory assumes a triadic reciprocality among behaviors, personal factors such as cognition, and environmental variables (Bandura, 1986, 2001; Schunk, 2004). The theory emphasizes students' self-regulation of their learning processes.

The curriculum of educational institutions, teacher preparation, intervention strategies, and current policies are all influenced by our theoretical perspectives of the learning process. Research to increase our understanding of learning, including research on academic achievement and determining factors related to learning, is abundant. This research involves copious numbers of different tests to measure learning achievements as well as instruments to measure different traits and variables associated with the learning process and subsequent academic success.

Measurement

A number of measurement protocols can be used to assess learning related factors. Winne and Perry (2000) describe seven measurement protocols (self-report questionnaires, structured interviews, teacher judgments, think aloud measures, error detection tasks, trace methodologies, and observations of performance) that have been used to assess self-regulated learning. The construction of most measurement tools begins with scientific inquiry aimed to precisely identify and describe phenomena of interest through the construction of truths or principles prescribing the phenomena. The ability to quantify phenomena depends on the accuracy and precision of a measurement tool, and the ability of research findings to be replicated (Brennan, 2001). Measurement tools that can precisely measure phenomena of interest in the physical sciences exist, but seldom do such tools exist within the social science setting. Within educational and psychological inquiry variables of interest like intelligence, academic success, motivation, self-efficacy, test anxiety, personality, meta-cognition, and self-regulation variables are typically measured indirectly.

Researchers often utilize self-report questionnaires in lieu of other measurement protocols because gathering and scoring data from self-report questionnaires is typically less resource intensive (Winne & Perry, 2000). Creating self-report questionnaires typically involves defining those traits for measurement then selecting items which relate to the trait of interest. Quantifications of these unobservable variables require a process of abstraction in which researchers measure attributes of these constructs to increase knowledge of the latent variables.

The term latent variable is used to refer to a construct inferred indirectly through a number of observable (indicator) variables (Hair, Black, Babin, Anderson, & Tatham, 2006; Meyers, Gamst & Guarino, 2006). A focus of measurement then is to use rules for assigning numerical values to represent quantities of these attributes (Nunnally, 1978). These numerical values can be summarized using mathematical operations which allow one to construct, validate, and investigate psychological measures.

The process of a) constructing measurement tools, b) employing mathematical methods to assist with construction, validation of interpretations, and investigation of psychological measures, c) assessing the level of reliability of interpretations based on the scores from an instrument, and d) assessing the level of validity with which we can make inferences from the measurement tool make up the field of psychometrics (Nunnally, 1970). Generally, the first steps to constructing a measurement instrument are to identify behaviors that represent the construct using theoretical rationales and then determining the set of operations that will isolate the attribute of interest and display it (Crocker & Algina, 1986). It is important that test developers write clear and concise items, make test instructions easily understood, and provide detailed prescriptions and conditions for administering the instrument (Nunnally, 1978).

Theoretically, there exists an unbounded universe of potential items to represent a construct of interest, and test developers must determine a finite pool of sample items appropriate to represent aspects of the attribute under study. Sets of items deemed to represent the construct should relate together in varying degrees (Streiner, 2003). Commonly, test developers have employed factor analysis to determine how well test items fit together. Factor analysis is a mathematical technique used to identify the underlying and interdependent structure among items/variables in an analysis. "The general purpose of factor analytic techniques is to find a way to condense (summarize) the information contained in a number of original variables into a smaller set of new, composite dimensions or variates (factors) with a minimum loss of information – that is, to search for and define the fundamental constructs or dimensions assumed to underlie the original variables," (Hair et al., 2006, p. 107).

Additional mathematical and statistical methods employed to explore the properties of item scores exist. For instance confirmatory factor analysis, a multivariate technique, enables researchers to simultaneously examine a series of interrelated dependence relationships among measured items and latent constructs (Hair et al., 2006). These mathematical methods can provide evidence to researchers for which items appropriately measure the latent construct and which items should be eliminated, attesting to the quality of the instrument as a measurement tool. Throughout the development phase for any measurement protocol, researchers must continually assess evidence for the quality of any test or instrument. Researchers must be concerned with the validity and reliability of scores from their measurement tools. Validity refers to the degree in which scores provide information relevant to the construct of interest, and reliability refers to the accuracy of a measurement procedure (i.e., how consistent and reproducible are the scores on a measurement procedure). A review of both concepts is discussed further.

Validity

Messick (1989) defines validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p. 13). Within educational and psychological disciplines, sources of validity evidence generally have been concerned with investigating content of a measurement in respect to the domain of reference;

inspecting the response behaviors of examinees to items/tasks on the measurement; exploring the internal and external structures of tests; considering differences in scores across contextual factors; and probing social consequences of interpretation and use of test scores (Crocker & Algina, 1986; Messick, 1989; Pedhazur & Schmelkin, 1991). Validation is an ongoing process which is highly dependent on the nature and quality of accumulated evidence gathered for the construct under study (Pedhazur & Schmelkin, 1991).

Historically, researchers have organized and grouped validity into several separate categories. For example, in 1954, Anastasi organized validity into face validity (what a test appears to measure), content validity (content relevance and representativeness), factorial validity (correlation between scores and a factor common to a group of measurements), and empirical validity (relation between test scores and a criterion). In 1949, Cronbach recognized two types of validity: logical (judgments for exactly what the test measures) and empirical (correlations between test scores and another measure) and modified his views in 1960 to consider content, criterion-related, predictive and concurrent, and construct validity. In 1966, the *Standards* published by the American Psychological Association recognized three categories of validity – content, criterion-related, and construct validity. Content validity is a reflection for how well subject matter included in a measurement relates to the intended content domain in which conclusions are made. It provides validation evidence of the relevance and representativeness of the content domain included within measurements. Criterion-related validity refers to the systematic relation of scores to other outcome criteria and is a reflection for how well scores on a measurement can be compared with an external criterion which is considered to represent the measured characteristic or behavior in question. Construct validity refers to the utility of the instrument to measure the unobservable trait or behavior of interest.

Construct validity seeks evidence for support of the interpretation and meanings deduced from scores (Messick, 1989).

Messick (1989) criticized this traditional framework of validity as he felt it permits researchers to pick and choose aspects of validity they consider important attributes, thus allowing researchers to ignore other necessary components of validity. He asserts that the traditional framework disregards score interpretation and use as well as disregards the social consequences of score interpretation. He argues that validity should be viewed as a uniform concept which encompasses multiple forms of evidence to justify test interpretation and use (Messick, 1995). He asserts that content validity and criterion-related validity are subsidiary elements of construct validity, and that these components alone are not sufficient for validation – validation must also consider the consequential basis of validity (intended and unintended outcomes of test interpretation). Messick posits that it is score interpretation which should undergird all score-based inferences, that evidence and rationales supporting the trustworthiness of the interpretation of scores are subsumed under construct validity.

Messick's unitary validity concept highlights six interdependent and complementary aspects of construct validity: content, substantive, structural, generalizability, external, and consequential (Messick, 1995). Content validity considers the representativeness, relevance, and technical quality of the sample of tasks included on a measurement. The dimensions of a construct domain should be defined in relationship to the knowledge, skills, attitudes, motives, and other attributes that describe the construct. Substantive validity refers to both the appropriateness of the sampling domain processes and the consistency and regularity of responses to the domain processes. Structural aspects of validity refer to the relationship of scores to "structural" relations innate to the construct. This aspect to validity considers theories for guiding selection of content items, as well as the scoring criteria for these items. Generalizability validity refers to limitations inherent in score meaning. Researchers must consider how interpretations generalize across tasks, across subjects not sampled, across settings, etc. Generalizability refers to the representativeness of score interpretations within the specified construct. Convergent evidence and discriminant evidence are considered components of the external aspect of validity. External validity incorporates both forms of evidence to assess the extent to which scores relate to similar assessments of the same construct – this evidence either strengthens or discounts hypotheses of interest. Finally, the consequential aspect of validity considers the value implications of score interpretations, especially with respect to issues of bias, fairness, and social justice. The major aim for assessing consequential validity is to assess the intended and unintended consequences of interpretation use in hopes of preventing negative impacts on persons or groups which might result from test invalidity.

Validation is an evaluation argument that is continually evolving and ongoing. Validation encompasses issues of interpretability, relevance, utility of scores, value implication of scores, and social consequences of score interpretations (Messick, 1989). In Messick's model, substantive validity refers to the consistencies in domain responses. The validity of a total score is predicted by the strength with which the comprised items measure the same thing, and is generally referred to as the reliability of a scale or measure. Reliability is a necessary condition for validity, but validity is not guaranteed even if a test or scale is reliable.

Reliability

Reliability is defined by the Merriam-Webster's Collegiate Dictionary as "the extent to which an experiment, test, or measuring procedure yields the same results on repeated trials". Research-based measurement relies on accurate and dependable scores; hence, the reliability of measurement is a major concern for research in the socio-behavioral sciences (Cronbach, 1951). Reliability is contingent on the measurement tools used to collect research data. Mislevy (1994) categorized reliability indices into four senses: a) true-score reliability, b) differential likelihood, c) reproducibility, or d) credibility.

True-score reliability, as reflected in classical test theory, is often used in the estimation of reliability for many self-report constructs. Classical test theory is the earliest development of a theory to capture the reliability of an examinee's responses and is credited to Charles Spearman. True score reliability relates to the consistency in an examinees scores by considering errors of measurement. An examinee's observed score for a measurement is assumed to be a function of the person's true score plus some error of measurement (Crocker & Algina, 1986; Lord & Novick, 1968; Pedhazur & Schmelkin, 1991). Multiple approaches in methods to assess true score reliabilities have been proposed and will be discussed in more detail.

A different approach to reliability, differential likelihood, is most commonly used for assessing ability levels between examinees or groups of examinees, and is increasingly being used in high-stakes testing. This reliability approach is related to item response theory which takes into account characteristics of test items and how examinees respond to the items. A unique feature of item response theory is that test developers can account for the difficulty and discrimination of selected test items and can adapt test items according to how examinees respond to the test. (Crocker & Algina, 1986; Yen & Fitzpatrick, 2006)

Reliability as a reflection of reproducibility in measurement relates to assessing the probability of agreement when using observational measures. Computations for this sense of reliability involve estimating the proportion of agreement among raters, decision-consistency coefficients, and generalizability coefficients (Mislevy, 1994). Interrater agreement is perhaps the most common approach for measuring the extent to which multiple observers agree in their coding, but Pedhazur and Schmelkin (1991) warn that interrater agreement addresses a particular source of error but is not an index of reliability. They advise that generalizability theory is more appropriate for assessing interobserver reliability since it distinguishes among different sources of error.

Reliability as credibility refers to an overall judgment for the extent one judges information to be trustworthy (Moss, 2004). Perhaps, credibility is most strongly associated with qualitative research which employs methods of triangulation for confirming that independent measures provide some level of agreement (or at least do not contradict) research findings. The reliability of the data relates to the extent different sources of data converge (Miles & Huberman, 1994).

True-score reliability continues to be the predominant theory as the basis for estimating reliability of self-report measures. Theoretically, true-score reliability considers the decomposition of observed scores into true scores and error scores and assumes 1) that the observed score consists of a true component score and an error component (two unknowns); 2) the mean error score is zero (meaning if an infinite number of measurements were collected, the errors would cancel to zero); 3) the true score is equivalent to the average of all observed scores (infinites measures); 4) the expected correlation between true scores and error scores is zero; 5) the average of observed scores across individuals is assumed to equal the average of the true scores; and 6) the observed score variance is supposed to equal the true score variances plus the error score variances (Crocker & Algina, 1986; Lord & Novick, 1968; Pedhazur & Schmelkin, 1991).

Variations of the true-score model which relax these assumptions have been proposed. These more lenient models include measures of tau-equivalency, essentially tau-equivalency, or congenericity. Under the assumption of tau-equivalency measures are assumed to have identical true scores without assuming homogeneity of variances; whereas essentially-tau equivalency assumes that the true scores and variances differ for each measure, but these differences in true scores are only by an additive constant. The assumptions for how scores vary between measures are relaxed further for congeneric models which only assume that the true scores on different measures are assessing the same phenomenon and are highly correlated to each other (Feldt & Brennan, 1989; Pedhazur & Schmelkin, 1991).

Formulations and approaches to characterizing true-score reliability require a) data from two equivalent forms of a test (parallel or alternative test forms); b) data from a repeated test or measure (test-retest); or c) data resulting in the subdivision of a test into two or more equivalent fractions from a single administration (internal consistency). The primary objective for administering parallel forms of a test is to control for error which may result in the sampling of items. Parallel test forms take into account the variation resulting from the sample of tasks, increasing the accuracy of generalizing from the specific score on a set of tasks to the broader domain those tasks represent. Forms of a test should be constructed according to the same specifications but with different/separate samples from the defined behavior domain. Characteristics include questions of the same difficulty, of the same sort, and of the same type (Thorndike, 2005).

Reliability estimated from parallel measures is typically found by computing the correlation between the two measures. Reliability estimates can capture all sources of variation that can affect scores including a) variation arising within the measurement procedure itself; b) changes in the person from day to day; c) changes in the specific sample of tasks; and d) changes in the individual's speed of work (Thorndike, 2005). Reliability estimates computed from parallel forms are generally more conservative than other approaches of assessing reliability. However, the usefulness in using parallel forms is limited as assumptions are restrictive, researchers are limited in their ability to construct equivalent forms, and the cost of multiple administrations can be high (Pedhazur & Schmelkin, 1991).

The model of test-retest reliability is similar to the conception of parallel forms – subjects are given the same test at different points in time and the reliability coefficient is computed as the correlation between the two measures. This reliability coefficient can be impacted by carry over effects, thus researchers must consider the nature of the measure {are constructs of interest assumed to remain constant?} and timing {sufficient time should elapse between the administration of measures so that responses are not impacted by the previous responses} (Thorndike, 2005). Reliability estimations using this approach can also be costly which has lead many practitioners to consider other approaches to reliability estimation.

Administering a single test form is often preferable and economically feasible for many researchers. An alternative model using a single administration of a test is to estimate the split-half reliability of scores. The items on a measure are split into two halves, equivalent in nature, and the scores on these two halves are correlated to provide an estimate of reliability. The most widely used split-half reliability is the Spearman-Brown formula. The Spearman-Brown formula assumes that parts included on a measure are strictly parallel. Making a decision for exactly how to split a measure into two equivalent halves can be problematic; therefore, various formulations of alternative internal consistency estimations have been introduced. The most common reported index of internal consistency reliability is Cronbach's alpha. Cronbach's alpha is computed by splitting a scale into as many parts as it has items, calculating the correlation between each subpair, and taking the average of all correlations (Cronbach, 1951). This estimate assumes that items comprising the measure must be at least essentially tau equivalent.

The number of reliability indexes that can be used to calculate an estimate of reliability continues to expand. Limitations for using measures of internal consistency include a) the coefficients only give evidence on the precision with which we can appraise a person at a specific moment; b) items on a single-administration of a test may lack true independence; c) items might measure multiple traits; and d) consistency of performance can be greatly affected if a test is timed. However, some research inferences can be accurately inferred from measurements taken during a single administration, and the cost and feasibility for collecting data at one time point is very attractive to researchers (Thorndike, 2005).

Reliability of scores of a measurement instrument is a unitless index that falls between zero and one. Reliability estimates at zero represent complete random error in measurement while coefficients near one indicate no random errors of measurement (Guilford, 1954; Pedhazur & Schmelkin, 1991). Mathematically, reliability coefficients can be negative. Negative coefficients are possible when items are negatively correlated with other items in the scale {plausible when reverse scoring negatively stated items when most items are positively stated} or when variability of the individual exceeds their shared variance {can occur when items tap a variety of different constructs} (Streiner, 2003). Values near zero and negative values indicate serious problems in the construction of the scale.

The level of acceptability for any reliability coefficient is generally determined by the type of decisions made on the basis of scores and the possible consequences of these decisions. On a continuum it is generally agreed that lower levels of reliability are acceptable in the initial stages of research, higher reliabilities are required to determine differences between groups, and high reliabilities needed for inferential decisions (Nunnally, 1978; Pedhazur & Schmelkin, 1991; Thorndike, 1951). Debate over what values represent these lower, higher, and high reliabilities is ongoing. Perhaps Kelley (1927) introduced the first generally accepted guidelines for minimum reliabilities needed according to research purposes. These guidelines were developed using the assumption that for a test to be useful it must permit discrimination of a difference as small as .26 times the standard deviation of a grade group with chances five to one of being correct. Guidelines stated that a) coefficients around .50 were acceptable to evaluate level of group accomplishment; b) coefficients around .90 were acceptable to evaluate differences in level of group accomplishment on two or more performances; c) coefficients around .94 were acceptable to evaluate level of individual accomplishment; and d) coefficients around .98 were acceptable to evaluate differences in level of individual accomplishment in two ore more performances. Several other major works have been cited to justify levels of acceptable reliability (e.g., Nunnally, 1967, 1978; Nunnally & Bernstein, 1994; Thorndike, 1951). In the words of Thorndike:

> It must be recognized, however, that these values are arbitrary, being derived from the above assumptions as to what it would be reasonable to expect a test to do in the way of discrimination between individuals and groups. How low a reliability one is willing to accept in any given case depends upon the practical values which are involved in that particular case. If some action must necessarily be taken and only unreliable measures are available as a basis for action, one may have to make the best of an unsatisfactory situation and use the most reliable of the available measures even if it has a reliability coefficient of .40 or .50. (p. 609)

Reliability coefficients provide a basis for comparing the precision of measurement across tests. Tests have a number of different acceptable coefficients of reliability, depending on what major sources of measurement error are to be considered in the calculation of a reliability coefficient. Variability of a group on a measured trait can affect reliability estimates achieved across different tests – as the variability in a group increases, the reliability coefficient for the test also increases. Heterogeneous samples tend to yield higher reliability coefficients than homogeneous samples because it is easer to differentiate between placements of individuals. An additional factor includes the level of the group on the trait being measured since the accuracy of a test differs according to the difficulty level of items on the test for different groups. The length of the test can impact reliability estimates. As the length of the test is increased, chance errors of measurement tend to cancel. Even the operations used for estimating reliability can affect estimates since different procedures treat different sources of variation in different ways (Pedhazur & Schmelkin, 1991) Speaking of the reliability of a test offers evidence for the amount of measurement error that could be expected when using the instrument (Nunnally, 1978). Thorndike (2005) suggests when the dimensions of multiple tests are similar (e.g. the quality of test items, the traits measured by those items, length, and the nature of the examinees) tests that yielded higher reliability coefficients should be preferred.

Validity Review of the Motivated Strategies for Learning Questionnaire

The Motivated Strategies for Learning Questionnaire (MSLQ) was developed in 1991 through research conducted by the National Center for Research to Improve Postsecondary Teaching and Learning. The instrument assesses student motivation and learning strategies use for post-secondary students (Pintrich, Smith, Garcia & McKeachie, 1991). The MSLQ is a self-report instrument that includes 81 items partitioned into a motivation section and a learning strategies section. The motivation section is comprised of three components: a value component which includes scales of intrinsic goal orientation, extrinsic goal orientation, and task value; an expectancy component which includes scales for control of learning beliefs and self-efficacy for learning and performance; and an affective component which includes a scale for test anxiety. The learning strategies section includes two components: a cognitive and metacognitive strategies component which includes scales for rehearsal, elaboration, organization, critical thinking, and metacognitive self-regulation; and a resource management strategies component which includes scales for time and study environment, effort regulation, peer learning, and help seeking (Pintrich et al., 1991; Pintrich, Smith, Garcia, & McKeachie, 1993). Developers of the instrument claim, "The MSLQ seems to represent a useful, reliable, and valid means for assessing college students' motivation and use of learning strategies," (Pintrich et al., 1993, p. 812). A review of the validity of the MSLQ is discussed using Messick's (1995) six dimensions of construct validity.

Content Validity

Brualdi (1999) states, "A key issue for the content aspect of validity is determining the knowledge, skills, and other attributes to be revealed by the assessment tasks," (p. 1). The theoretical framework for the MSLQ is presented in McKeachie, Pintrich, Lin, and Smith (1986). The focus of the authors is to explore effective problem-solving skills that facilitate students' motivation and use of self-regulated learning strategies. Their theoretical framework integrates ideas from a general cognitive approach and a social-cognitive approach to learning. Developers of the instrument believe that students are active agents in the learning process and that their academic success is mediated by both motivational and cognitive characteristics of the student. Content validity for the MSLQ is presented in terms of the theoretical rationales for the fifteen scales included on the instrument, with a description of the items used to represent each of the fifteen constructs.

The first section of the MSLQ includes constructs related to student motivation. Motivation to succeed in academic tasks is commonly referred to as achievement motivation and is considered the driving force in which learners achieve their academic goals. Motivation involves the mental process students use to activate, sustain, and maintain behavior (Pintrich & Schunk, 2002). Theoretical frameworks for achievement motivation attempt to explain why learners choose to engage in achievement tasks, expend effort to achieve these tasks, and persist on such tasks when faced with difficulty (Dweck, 1999; Nicholls, 1984).

Ford (1992) proposed Motivational Systems Theory which asserts achievement motivation is impacted through interrelations of personal goals, personal agency beliefs, and emotions. Authors of the MSLQ partitioned aspects of motivation into similar components. The motivational section of the MSLQ includes components of value, expectancy, and affect, which are further divided into scales of intrinsic and extrinsic goal orientations, task value, control of learning beliefs, self-efficacy of learning and performance, and test anxiety. Authors of the instrument used expectancy-value models of achievement for the framework of these motivational scales (McKeachie et al., 1986). The general expectancy-value model is a derivation of Atkinson's (1964) model of achievement motivation. Expectancy-value models account for a learner's beliefs in their abilities and the value they assign to a given task which relates to the learners' overall motivation (Wigfield & Eccles, 2000). Students who perceive a task with positive value accompanied with perceptions of capability are more likely to engage and persist in the task, whereas students may expend very little effort to confront a task when they do not value the task or feel inadequate to perform the task successfully (Pintrich & Schunk, 2002).

The second section of the MSLQ aims to measure students learning strategy use based on cognitive theory. Cognitive theorists advocated that learning involves more complex mental associations which may not be reflected through overt behavior changes of the learner. Major principles for learning within this framework include an individual's a) prior knowledge, b) processes for organizing the information to be learned, and c) processes for perceiving, comprehending, and storing information (Gredler, 2005; McKeachie et al., 1986; Schunk, 2004; Shuell, 1986).

A significant element of the MSLQ is an emphasis on the efforts students exert in the learning process, referred to as the students' self-regulated learning. Self-regulated learning refers to the active role individuals play in acquiring and constructing knowledge, and is considered a key for conditioning students to be life-long learners. Self-regulation is a central component in social-cognitive theory and amalgamates aspects of motivational, cognitive, metacognitive, affective and environmental factors (Paris & Paris, 2001; Zimmerman, 1986). Zimmerman (1986) defines self-regulated learning as "the degree to which learners are metacognitively, motivationally, and behaviorally active participants in their own learning process" (p. 309).

Several models of self-regulated learning have been presented according to theoretical perspectives of different researchers, but most definitions tend to reveal considerable overlap (Pintrich, 2000b). The MSLQ is based on a self-regulation model that considers both motivational and cognitive processes. Pintrich (2000b) describes self-regulated learning as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment. These self regulatory activities can mediate the relationships between individuals and the context, and their overall achievement," (p. 453). Pintrich (2000b, 2004) presents four-phases to self-regulation: a) forethought, planning and activation, b) monitoring, c) control, and d) reaction and reflection. The forethought phase concerns student's knowledge, goals, planning, efficacy judgments, and task value beliefs. The monitoring phase considers student's meta-cognitive reflections on the learning process; while the control phase involves the student's selection and use of appropriate learning strategies. The reaction and reflection phase considers student's task evaluation of the learning process and possible adjustments. A description of motivation, cognitive, and self-regulation factors included on the MSLQ follows.

Intrinsic and Extrinsic Goals. Students' intentions for academic success are generally referred to as learning goals (Ames, 1992) and are considered motivators for guiding and directing academic behaviors. Personal goals for engaging in a learning task may include a desire to master the subject matter, satisfy a need for achievement, receive recognition or approval, avoid failure, obtain the knowledge and skill base necessary for the workforce, outperform peers, and/or avoid negativity from social networks (Dweck, 1986; Dweck & Leggett, 1988; Harackievicz, Barron, Pintrich, Elliot & Thrash, 2002; Pintrich, 2000a).

Goal theory generally emphasizes mastery and performance goals which have also been expressed as intrinsic and extrinsic goal orientations since learners may have an "intrinsic" desire to master the material or an "extrinsic" desire to meet some level of performance demands. Mastery and performance goals are also partitioned into approach and avoidance orientations (Elliot & Church, 1997; Linnenbrink & Pintrich, 2000, 2002; Pintrich, 2000a). However, at the time the MSLQ was developed, authors of the instrument used a theoretical framework that does not partition mastery and performance goals into these approach and avoidance dimensions. Instead items on the instrument are intended to assess the degree students are intrinsically and extrinsically motivated.

Students with intrinsic goals perceive learning tasks as opportunities to increase their knowledge of the subject matter (Dweck & Leggett, 1988) and possess real interest and desire to master understanding. These students tend to have a higher degree of interest for academic tasks, higher perception of task importance and utility, increased cognitive engagement, positive perceptions of academic efficacy, stronger focus for developing new skills, expend more effort when encountering challenges, and utilize successful learning strategies (Elliot & Church, 1997; Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997; Pintrich, 2000a; Wigfield & Eccles, 2000; Wolters, 2004).

Students with extrinsic goal orientations are focused on demonstrating their ability, outperforming others, getting good grades or other external benefits such as praise, proving their self worth to others, and/or to avoid negative consequences. Students with higher degrees of performance goal orientations are perceived to prefer less challenging tasks, utilize surface-level learning strategies more frequently, are less willing to seek help, lack strong efficacy beliefs, and give up when faced with adversity (Ames, 1992; Elliot & Church, 1997; Elliot & Harackiewicz, 1996; Middleton & Midgley, 1997; Pintrich, 2000a).

Research on goal orientations stresses that variations in orientations exist both across individuals and within individuals according to situational demands (Ames, 1992; Ames & Archer, 1988) Lin et al. (2003) assert, “it seems likely that intrinsic and extrinsic motivation, rather than being at opposite ends of a single dimension, may be much more complex in their relationships with one another and other variables affecting student achievement," (p. 253). Multiple classroom, task, and motivational features as well as individual characteristics may negatively or positively influence learning goals during the process to reach these goals (Brookhart & Durkin, 2003; Duncan & McKeachie, 2005). At the start of a course, students may exhibit high intrinsic goal orientations, but as students progress through their studies, they tend to become more extrinsically motivated as they are under pressure to pass exams and demonstrate success (Adcroft, 2010; Ditcher, 2001). Other studies also suggest that younger students are more intrinsically motivated to explore and understand the world around them, but their intrinsic interest fades throughout their progression in an academic environment which regulates the content students are required to learn (Lepper & Hodell, 1989).

Whether students with higher degrees of intrinsic or extrinsic motivation demonstrate greater academic achievement is uncertain. Harackiewicz et al. (2002) found that students with mastery goal orientations have higher levels of academic achievement; yet, numerous other studies suggest extrinsically motivated students perform better (as measured by grades, etc.) than intrinsically oriented students (Elliot & Church, 1997; Harackiewicz, Barron, Tauer, Carter & Elliot, 2000). Despite these findings, researchers believe that students who hold a stronger sense of intrinsic motivation will be more likely to persist in the presence of challenges and are better prepared for life-long learning (Elliot & Harackiewicz, 1996; Pintrich, 2000a; Wolters, 2004).

The authors of the MSLQ use four items to estimate intrinsic goal orientation and four items to estimate extrinsic goal orientation. Items included to represent intrinsic goal include:

- In a class like this, I prefer course material that really challenges me so I can learn new things;

- In a class like this, I prefer course material that arouses my curiosity, even if it is difficult to learn;
- The most satisfying thing for me in this course is trying to understand the content as thoroughly as possible; and
- When I have the opportunity in this class, I choose course assignments that I can learn from even if they don't guarantee a good grade.

The four items used to represent extrinsic goal orientation are:

- Getting a good grade in this class is the most satisfying thing for me right now;
- The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade;
- If I can, I want to get better grades in this class than most of the other students; and
- I want to do well in this class because it is important to show my ability to my family, friends, employer, or others.

Task Value. Task value is identified with an individual's appreciation for a task's relevance. The quality and quantity of effort an individual expends on an activity is codetermined by the perceived value of the task and one's expectancy of success which influences one's incentive to engage in different learning tasks (Raynor, 1981; Schunk, 1991). Eccles (1983) proposed that the overall value of any specific task is a function of the attainment (importance for doing well on a task), intrinsic interest (interest for engaging in the task), and utility (importance of the task for some future goal) values of the task. Developers of the MSLQ include six items intended to represent all three task value components. Items include:

- I think I will be able to use what I learn in this course in other courses;
- It is important for me to learn the course material in this class;

- I am very interested in the content area of this course;
- I think the course material in this class is useful for me to learn;
- I like the subject matter of this course; and
- Understanding the subject matter of this course is very important to me.

Control of Learning Beliefs. Control of learning beliefs is similar to Rotter's (1966) locus of control construct which relates to the amount of control one perceives to have over his or her learning ability. Persons with internal locus of control attribute successes and failures to their own behaviors while students with external locus of control attribute their successes or failures to external causes such as luck, chance, or others actions. McKeachie et al. (1986) present Weiner's (1986) attributional theory model which comprises three dimensions for students' control of learning beliefs: locus, stability, and controllability. The locus dimension refers to the locus of causality for a student's success or failure. The stability dimension refers to the persistence of such locus of causality beliefs over time, and the controllability dimension refers to the learner's ability to control causes of successes or failures. Eccles (1983) avers:

> According to these theorists [attribution theorists], it is not success or failure per se, but the causal attributions made for either of these outcomes that influence future expectancies. For example, if people attribute success to a stable factor such as ability, then they should expect continued success. If, on the other hand, they attribute success to an unstable factor such as effort or good luck, they should be uncertain about future outcomes. Similarly, attributing failure to stable factors should produce expectations of continued failure, while attributing failure to unstable factors should not. Consequently, individuals who attribute their success to an unstable factor such as task ease and their failure to a stable factor such as lack of

ability should have lower expectancies than do individuals exhibiting the reverse attributional pattern, even if their performance histories have been identical. (p. 86)

McKeachie et al. (1986) propose that perceptions of personal control serve as mediators that influence learning activities and achievement. Developers of the MSLQ use four items to represent students' internal control of learning beliefs. Items are intended to assess students' beliefs in the efforts they expend to influence their academic performance. These items are:

- If I study in appropriate ways, then I will be able to learn the course material in this course;
- It is my own fault if I don't learn the material in this course;
- If I try hard enough, then I will understand the course material; and
- If I don't understand the course material, it is because I didn't try hard enough.

Self-Efficacy of Learning and Performance. In general, self-efficacy is referred to as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances," (Bandura, 1986, p. 391). In the expectancy-motivation framework, self-efficacy is a prominent concept of learners' expectations and can be defined as their internal belief for executing and succeeding in academic tasks at designated success levels (Bandura 1986, 1989; Schunk, 1991). The strength of one's self-efficacy is determined by previous performance or accomplishments (successes and failures), attributions of prior accomplishments (ability, effort, difficulty, and luck), vicarious experiences or model similarity (observations and comparisons of others performance), forms of persuasion (praise and/or criticisms), and physiological indexes (anxiety, fatigue, illness, etc.) (Bandura, 1986; Pajares & Miller, 1994; Schunk, 1991).

Persons with high self-efficacy beliefs show confidence in their skills and abilities to do well and have been shown to participate more in learning activities. These students tend to expend greater effort and persistence and achieve higher levels of academic performance than students with low self-efficacy (Pintrich & DeGroot, 1990; Pintrich & Schunk, 2002; Schunk, 1991). Conversely, students with low self-efficacy have little confidence in their skills and abilities and are less likely to persist when faced with challenges (Pintrich & Schunk, 2002). Lack of self-efficacy has also been coupled with debilitating affect of high test anxiety (Bandura, 1986)

Self-efficacy is not considered a static trait, but is considered to vary across different performance domains (Bandura, 1997; Lent, Brown & Hackett, 1994). Level and degree of self-efficacy beliefs differ according to task-performing situations and personal perceptions (Pintrich & Schunk, 2002). Bandura (1977, 1986) surmised that perceived self-efficacy influences choice of activities, persistence, effort expenditure, and task accomplishments. Research of self-efficacy is distinguished from similar constructs due to the context- and task-specific nature of the construct (Pajares, 1996). Bandura (1997) postulates that students tend to generalize self-efficacy across learning domains only when activities share similar skill sets. He also claims self-efficacy judgments are most accurately estimated when judgments and actions are measured in close temporal proximity.

McKeachie (1990) asserts when students are efficacious, learning becomes more intrinsically satisfying:

> We now see more clearly that students come into classes not only with background knowledge that may facilitate or interfere with their learning but also with experiences or anxieties that interact with the teacher's behavior and the tasks of learning in affecting the students' explicit or implicit choice of skills and strategies for learning. When students' lack

> a sense of self-efficacy, learning becomes routine drudgery with little thought about the meaning and purpose of the task and little motivation for going beyond the goal of meeting the minimum demands of the course; when, on the other hand, students begin to feel competent to learn – to think about subject matter as it relates to other learning and experiences – learning becomes intrinsically satisfying. (p. 140)

The eight items used to comprise the scale for self-efficacy for learning and performance are intended to measure both efficacious appraisals of ability and performance expectations in a specific college course (Pintrich et al., 1991). Items include:

- I believe I will receive an excellent grade in this class;
- I'm certain I can understand the most difficult course material presented in the readings for this course;
- I'm confident I can understand the basic concepts taught in this course;
- I'm confident I can understand the most complex material presented by the instructor in this course;
- I'm confident I can do an excellent job on the assignments and tests in this course;
- I expect to do well in this class.
- I'm certain I can master the skills being taught in this class; and
- Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.

Test Anxiety. Test anxiety can be defined as an unpleasant feeling or emotional state students display when completing tests or other cognitive measures (Pintrich et al., 1991). Test anxious students are believed to have difficulty utilizing self-regulated learning strategies. VanZile-Tamsen and Livingston (1999) suggest test anxious students have difficulty focusing attention on key points, organizing course material, encoding information using more surface-level strategies, managing study time, and using external resources such as peers. Students with high degrees of test anxiety perform poorly on tests and other cognitive measures (Bandalos, Finney & Geske, 2003; Pajares & Miller, 1994; Zeidner, 1998). McKeachie, Lin, and Middleton (2004) differentiated students exhibiting high test anxiety into two groups – those who performed poorly on exams despite having good study skills (worry impacts the students capacity to effectively retrieve studied material) and those who performed poorly as a result of poor strategy use. Authors of the MSLQ use five items to represent the worry component of test anxiety and refer to students' negative thoughts that disrupt academic performance (Pintrich et al., 1991). The test anxiety scales does not include items to represent the emotionality component of test anxiety. Items included on the MSLQ are:

- When I take a test I think about how poorly I am doing compared with other students;
- When I take a test I think about items on other parts of the test I can't answer;
- When I take tests I think of the consequences of failing;
- I have an uneasy, upset feeling when I take an exam; and
- I feel my heart beating fast when I take an exam.

Cognitive Scales. Central to the cognitive perspective are information-processing theories, which focuses on the mechanisms used for encoding, storing, and retrieving information (Gredler, 2005; Schunk, 2004; Shuell, 1986). The multi-stage model of memory is considered a dominant perspective for the nature of human memory. This model asserts that information must first be attended to or perceived through one's senses to process selected information into short-term or working memory. Once in short-term or working memory the information can then be encoded into a meaningful form and moved into long-term memory. There are three basic types of long-term memory: episodic (personal), semantic (general knowledge), and procedural memory (adaptive reasoning). The final stage in memory is the retrieval of information, a process dependent on the encoding processes (Gredler, 2005).

Learners can engage in several learning strategies believed to influence the learner's encoding process (Weinstein & Mayer, 1986). McKeachie et al. (1986) used Sternberg's (1985) model to guide their research of information processing approaches to intelligence. Components of Sternberg's model consider performance of simple tasks, performance of more complex tasks, and how individuals relate prior knowledge to influence performance. Four scales (rehearsal, elaboration, organization, and critical thinking) are included on the MSLQ to assess general cognitive learning strategies.

Rehearsal is considered a simple learning strategy in which information is stored into working memory. Rehearsal strategies include methods for naming, repeating, and reciting material for learning (Weinstein & Mayer, 1986). The four items included to represent rehearsal on the MSLQ are:

- When I study for this class, I practice saying the material to myself over and over;
- When studying for this class, I read my class notes and the course readings over and over again;
- I memorize key words to remind me of important concepts in this class; and
- I make lists of important terms for this course and memorize the lists.

Elaboration strategies include paraphrasing or summarizing learning material to help learners build internal connections between one's prior knowledge and the new material (Weinstein & Mayer, 1986). This strategy is considered a higher order learning skill because the strategy allows learners to store learned information into long-term memory. Developers of the

MSLQ use six items to represent use of elaboration learning strategies. Items for the elaboration strategy scale include:

- When I study for this class, I pull together information from different sources, such as lectures, readings, and discussions;
- I try to relate ideas in this subject to those in other courses whenever possible;
- When reading for this class, I try to relate the material to what I already know;
- When I study for this course, I write brief summaries of the main ideas from the readings and the concepts from the lectures;
- I try to understand the material in this class by making connections between the readings and the concepts from the lectures; and
- I try to apply ideas from course readings in other class activities such as lecture and discussion.

Organization, also considered a higher order learning strategy includes methods of outlining, taking notes, mapping or connecting key ideas in learning material (Weinstein & Mayer, 1986). Authors of the instrument represent organization with four items. These items are:

- When I study the readings for this course, I outline the material to help me organize my thoughts;
- When I study for this course, I go through the readings and my class notes and try to find the most important ideas;
- I make simple charts, diagrams, or tables to help me organize course material; and
- When I study for this course, I go over my class notes and make an outline of important concepts.

Critical Thinking is a higher order learning strategy which involves applying learned information to knowledge of new situations. Critical thinking is self-guided and self-disciplined processes for conceptualizing, applying, synthesizing, and evaluating information to be learned (Scriven & Paul, n.d.). Five items are used on the MSLQ to represent critical thinking:

- I often find myself questioning things I hear or read in this course to decide if I find them convincing;
- When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence;
- I treat the course material as a starting point and try to develop my own ideas about it;
- I try to play around with ideas of my own related to what I am learning in this course; and
- Whenever I read or hear an assertion or conclusion in this class, I think about possible alternatives.

Metacognitive Self-Regulation. Metacognition is a concept central to cognitivism, and is considered a higher order thinking process which incorporates knowledge of cognition as well as the regulation of cognition. Knowledge of cognition refers to knowledge about tasks, strategies, instructional plans, and goals; while regulation of cognition refers to goal setting, planning, monitoring one's understanding, and evaluating progress towards the completion of the task (Flavell, 1979, 1992; Pintrich, 2002). The effectiveness of incorporating metacognitive strategies within the learning process is dependent on the learner's belief about personal agency (self-efficacy), achievement goal orientation (mastery, performance, and failure-avoidance), and the learner's topic or domain knowledge (expert or novice) (Pintrich, 2002; Pintrich & Schunk, 2002).

Authors of the MSLQ use twelve items to assess metacognitive self-regulation. These items focus on both the control and regulation of metacognition, and aspects measured include three processes of metacognition (planning, monitoring, and regulating). The planning aspect refers to activities such as goal setting and task analysis, which are intended to help students plan learning strategies and processing of information. Monitoring activities consist of tracking one's attention during learning tasks, quizzing oneself on learning material, and connecting material with prior knowledge. The regulating component refers to adjustments to cognitive activities to accomplish learning goals (McKeachie et al., 1986; Pintrich et al., 1991). The items representing the metacognitive self-regulation construct are

- During class time I often miss important points because I am thinking of other things (REVERSED);
- When reading for this course, I make up questions to help focus my reading;
- When I become confused about something I'm reading for this class, I go back and try to figure it out;
- If course materials are difficult to understand, I change the way I read the material;
- Before I study new course material thoroughly, I often skim it to see how it is organized;
- I ask myself questions to make sure I understand the material I have been studying in this class;
- I try to change the way I study in order to fit the course requirements and instructor's teaching style;
- I often find that I have been reading for class but don't know what it was all about (REVERSED);

- I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when studying;
- When studying for this course I try to determine which concepts I don't understand well;
- When I study for this class, I set goals for myself in order to direct my activities in each study period; and
- If I get confused taking notes in class, I make sure I sort it out afterwards.

Time and Study Management. Resource management strategies comprise different approaches to manage and control time, effort, study environment, and in seeking assistance from qualified persons. Time and study management involves choosing environments that are conducive to learning (i.e., free from distractions), and effectively scheduling, planning, and managing one's study time. Study environments that are conducive to effective studying mean choosing physical locations that are relatively free from visual, auditory, and other distractions. Time management entails that the learner has an awareness of deadlines and the length of time needed for task completion as well as prioritizes learning tasks (McKeachie et al., 1986; Pintrich et al., 1991; Zimmerman & Risemberg, 1997). Successful learners will be sensitive to the physical learning environment and their time management – making adjustments as necessary. Eight items are used to assess both time management and choice of proper study environments. Items include:

- I usually study in a place where I can concentrate on my course work;
- I make good use of my study time for this course;
- I find it hard to stick to a study schedule (REVERSED);
- I have a regular place set aside for studying;
- I make sure I keep up with the weekly readings and assignments for this course;

- I attend class regularly;
- I often find that I don't spend very much time on this course because of other activities (REVERSED); and
- I rarely find time to review my notes or readings before an exam (REVERSED).

Effort Regulation. The effort a student expends to reach his or her learning goals is termed effort regulation. Effort regulation is similar to volitional control which is defined as the "tendency to maintain focus and effort toward goals despite potential distractions," (Corno, 1994, p. 229). Effort regulation enhances the ability of the learner to handle setbacks and failures within the learning process by correctly allocating resources and appropriate effort for more successful learning in the future (Chen, 2002). The effort students expend on a learning task is influenced by the importance, usefulness, and value ascribed to the task (Pintrich & Schrauben, 1992). The four items used to assess effort regulation on the MLSQ refer to students' commitments to complete study goals when faced with difficulties or distractions (Pintrich et al., 1991). Items include:

- I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do (REVERSED);
- I work hard to do well in this class even if I don't like what we are doing;
- When course work is difficult, I give up or only study the easy parts (REVERSED); and
- Even when the course materials are dull and uninteresting, I manage to keep working until I finish.

Peer Learning. Peer learning involves using peers to collaboratively understand course material or information to be learned (Jones, Alexander & Estell, 2010). Eccles, Wigfield, and Schiefele (1998) found that a crucial part of the learning process for adolescents was academic peer support, and Allen (1993) expressed that in traditional undergraduate education, a student's peer group is the single most important source of influence on growth and development. Freeman, Alston, and Winborne (2008) state "When students grapple with material and tasks in collaboration with their peers they are pushed to consider alternate ideas and perspectives, be responsible to others, and engage in critical and divergent thinking and, therefore, be intellectually enriched," p. 227. Peer group membership tends to result in *homophily*, similarities among group members evidenced in their academic motivations and achievements (Jones et al., 2010).

Slavin (1996) identifies four theoretical perspectives believed to explain achievement effects of cooperative peer learning: motivational, social cohesion, cognitive, and developmental perspectives. The motivational perspective focuses on students' goal structures. In this perspective attainment of students' goals are dependent on a group's performance, thus students encourage and take steps to ensure other group members work hard to attain overall group success. Groups illustrate social cohesion when students help others learn because of their friendships together and desires for each member to be successful. The cognitive perspective stresses that cooperative peer learning increases students' achievements through interactions that elicit mental processing of information. Under the developmental perspective, Slavin references Vygotsky's zone of proximal development as a model in which children's growth is promoted because children are assumed to be in similar zones of development and observe modeled behaviors from more advanced students in the group. The MSLQ includes three items to measure peer collaboration. Items for the peer learning scale include:

- When studying for this course, I often try to explain the material to a classmate or a friend;

- I try to work with other students from this class to complete the course assignments; and
- When studying for this course, I often set aside time to discuss the course material with a group of students from the class.

Help Seeking. Research on help seeking examines the determinants in which students will reach out for assistance and who students seek help from. Help seeking can be an adaptive learning strategy that allows one to optimize learning. Students who utilize local resources such as instructors, peers, tutors, additional textbooks, and/or internet resources when encountering difficulties with a learning task can enhance their learning experiences (Ames, 1983; Karabenick, 1998; Zimmerman & Martinez-Pons, 1986).

Two major impetuses for seeking help have been proposed: instrumental and executive help seeking (Nelson-Le Gall, 1981). Students who seek help to decrease their subsequent need for assistance, i.e., seek just enough help to achieve independently or improve the quality of their performance are considered instrumental help seekers. This form of help-seeking is viewed favorably as a necessary and adaptive approach to learning. Karabenick (2003, 2004) found that students who adopt mastery goals are more likely to seek instrumental help-seeking. Students whose goals for seeking help are to minimize their effort expenditure for achievement are regarded as pursuing executive help-seeking goals. With expedient help-seeking, students seek help from others even when achievement could likely be attained independently, but getting help allows the learners to avoid work and eases and lowers the cost of goal attainment (Karabenick, 2004; Nelson-Le Gall, 1981, 1986). Research suggests that students with higher levels of self-efficacy appear more likely to seek help when they need it than students who are less efficacious. Students who judge their competencies negatively tend to avoid asking others for help (Ryan, Gheen & Midgley, 1998; Ryan & Pintrich, 1998).

Developers of the MSLQ use four items to assess students' willingness to seek help from others when students face difficulties with learning tasks (Pintrich et al., 1991). Help seeking items include:

- Even if I have trouble learning the material for this class, I try to do the work on my own, without help from anyone (REVERSED);
- I ask the instructor to clarify concepts I don't understand well;
- When I can't understand the material in this course, I ask another student in this class for help; and
- I try to identify students in this class whom I can ask for help if necessary.

Three of the four items relate to students' willingness to seek help from peers and their instructors, while one item relates to students' willingness to seek help in general.

Structural Validity

Structural validity considers the structural or interrelations between items on the instrument and the construct (Messick, 1995). The structural relations of MSLQ subscales for both the motivational and learning strategy components were confirmed using confirmatory factor analyses which test the theoretical model and its operationalization (Pintrich et al., 1991, 1993). Confirmatory factor analyses allow researchers to specify which items or indicators fall onto which factors or latent variables.

Many fit indices have been established to assess the fit between observed data and theoretical models (Hair et al., 2006; Meyers et al., 2006). Developers of the MSLQ used several common absolute fit indices including the χ^2/df ratio, the goodness-of-fit index (GFI – proportion of variance in the sample correlation / covariance accounted for by the predicted model), an adjusted goodness-of-fit index (AGFI – takes into account model complexity by adjusting the GFI by a ratio of the degrees of freedom used in a model to the total degrees of freedom available), and the root mean residual (RMR – measure of the differences between values predicted by a model and the values actually observed or estimated). Garcia and Pintrich (1996) indicated that a χ^2/df ratio below five is indicative of a good fit between the observed and reproduced correlation matrices - the χ^2/df ratio for the motivation components model was 3.49 and 2.26 for the learning strategies component model. GFI values range from zero to one with higher values indicating better fit. Conventionally GFI values of .90 or higher signify acceptable models. Conversely, lower RMR values represent a better fit of data to the expected model (Hair et. al., 2006; Meyers et al., 2006). For the motivation model, researchers reported a GFI of .77, AGFI of .73, and RMR of .07. A GFI of .78, AGFI of .75, and RMR of .08 were found for the learning strategies model (Garcia & Pintrich, 1996; Pintrich et al., 1993).

Validation studies of the structural validity of the MSLQ have also been explored by several other researchers. Hamilton and Akhter (2009) examine the validity of the motivational scales of the MSLQ using both exploratory factor analysis and confirmatory factor analysis with a sample of 327 students enrolled at a university in New Zealand. The results of exploratory and confirmatory factor analysis did not confirm the factor structure presented by the developers of the instrument. Hamilton and Akhter suggest that some items for both the test anxiety and self-efficacy scales need revision and recommend at least one item for the self-efficacy scale should be removed because the item is highly skewed and does not combine with other efficacy items.

In Jacobson's (2000) doctoral work to determine differences between traditional college students (age 18-22) and nontraditional students (age 23 and above) in their reported motivation and learning strategies use, Jacobson examines the factor structure of the MSLQ by conducting confirmatory factor analysis for the overall sample of 396 students. Jacobson also compared structural findings for both the traditional and nontraditional students. Jacobson claims, "Overall, the models show sound structures, and reasonable factor validity for the MSLQ scale with these sample subjects. The comparative fit index (CFI) along with the critical ratios indicate that this model shows 'good fit' with this sample of subjects. All critical ratios exceeded the 1.96 criteria and all CFI's were greater than .90," p. 49.

Davenport (2003) examines the factorial, structural, and predictive validity of the MSLQ for 226 undergraduate educational psychology students. Davenport compares the structural model presented by Pintrich et al. (1991) using aggregated MSLQ scale scores. Although Davenport found satisfactory fit of the sample data to the model proposed by Pintrich et al., he notes that slight modifications to the model produce a better fitting model. Davenport concludes:

> [It] appears that the MSLQ does provide accurate information about students' motivations attitudes and learning behaviors. The structural analysis showed that the interrelationships between the MSLQ factors are not random occurrences and that the descriptiveness of the information provided in the 15 scale scores is severely reduced when instructors and researchers limit their examinations of the MSLQ to a simple correlation matrix. However, it does appear that the MSLQ suffers from some problems common to many Likert-scale, self-report measures. Some items on the instrument may need to be revised. However, these problems did not appear to be significant. In spite of these difficulties, the MSLQ does appear to be a valid measure of motivation and learning strategy use and resource use. Although it may not demonstrate the stringent psychometric qualities of a diagnostic instrument (it was not intended to be such), it does appear to be appropriate for group research and instruction. (p. 191)

Generalizability Validity

Generalizability validity refers to the appropriateness of score interpretations across different tasks, different samples, different settings, and so forth (Messick, 1995). The MSLQ was designed to be used with post-secondary students, and data presented in the manual (Pintrich et al., 1991) was developed on a sample of 380 college students from the United States (356 students from a public 4-year university and 24 students attending a community college) within 37 classrooms covering 14 subject domains and 5 disciplines (natural science, humanities, social science, computer science, and foreign language). The generalizability validity of the MSLQ should be considered across multiple contexts and should be considered by any researcher considering the appropriateness for using the instrument.

Is the MSLQ appropriate to use with younger students, i.e., can researchers use the instrument to assess motivation and learning strategies use of elementary, junior high or high school students? Pintrich and DeGroot (1990) adapted the MSLQ instrument using a sample of seventh and eighth grade students. Developers of this junior high version of the MSLQ used factor analysis to guide scale construction. The abridged version of the instrument includes 56 items representing three motivation scales (self-efficacy, intrinsic value, and test anxiety) and two cognitive components (cognitive strategy use and self-regulation). Researchers who are unsure about the generalizability of the 81-item MSLQ instrument may determine that Pintrich and DeGroot's (1990) junior high version is more appropriate for their sample.

A second consideration for the generalizability of the instrument might consider if cultural differences would inhibit the generalizability of MSLQ scales for students outside the United States? For instance, Chinese cultures ascribe different educational practices and reflect different values to students' education (Ho, 1994). Several researchers assert that Asian cultures emphasize effort over ability to explain academic performance (Chen & Stevenson, 1995; Hau & Salili, 1991). Research by Stigler, Smith and Mao (1985) and Whang and Hancock (1994) found that Chinese students have lower efficacy judgments of their cognitive competence despite higher performances than their American counterparts. Stevenson and Chen (1993) posit that Chinese students may report lower self-efficacy beliefs because of higher parental standards for their performance. Such theoretical ascriptions of different learning groups could suggest that researchers may not find utility in using some of the fifteen MSLQ scales.

The specificity of item responses may provide some limitations to the generalizability of participants' responses. For example, researchers continue to debate about the generality and specificity of self-efficacy judgments. "Researchers have been debating how specific is too specific to lose all practical relevance of findings and how general is too general to transform percepts of efficacy into something akin to personality traits," (Bong, 1999, p. 1). The MSLQ is probably most appropriate for researchers wanting to gauge a students' efficacy beliefs for a specific subject.

Generalizability relates to the degree research findings can be inferred from a sample population to a more general population. Generalizability judgments should consider contextual, cultural, and other mediating factors that may influence subject's responses. Researchers who are trying to determine if an instrument is appropriate for their own research purposes should consider the theoretical framework and relevance of the instrument to answer research questions.

External Validity

The external aspect of construct validity assesses the extent to which the scores relate to similar assessments of the same construct – this evidence either strengthens or discounts hypotheses of interest (Brualdi, 1999; Messick. 1995). Messick (1995) states, "the meaning of the scores is substantiated externally by appraising the degree to which empirical relationships with other measures – or the lack thereof – are consistent with that meaning," p. 746. The authors examined the predictive validity of the MLSQ scales by correlating each of the six motivation subscales and nine learning strategies scales with students' final grades in designated MSLQ courses, see Table 2.1. For motivation scales, the strengths of the relationships between intrinsic goal, task value, and self-efficacy, and test anxiety with final grade were moderate. Weak relationships were found between extrinsic goal and control of learning beliefs with final grade. The relationship with final grade and test anxiety was negative, while relationships with all other motivation scales to final grade were positive. The directions of correlations for the motivational scales were as expected (Pintrich et al., 1991, 1993). For learning strategies scales, peer learning related negatively (and weakly) to final grade. The scales for rehearsal and help seeking demonstrated almost no relationship to final grade. Each of the remaining learning strategy scales positively (although moderately) related to final course grades. Most of the relationships of the learning strategies scales with final grade were consistent with theory, although Pintrich et al. (1993) do note that the negative relationship between peer learning with final grade and the negligible relationship of help seeking with final course grade were not expected.

Table 2-1

Descriptive Statistics for Motivation Scales of the MSLQ – Manual Data

Item	*M*	*SD*	Alpha	Correlation to Final Grade
Intrinsic Goal (IG)	5.03	1.09	.74	.25
Extrinsic Goal (EG)	5.03	1.23	.62	.02
Task Value (TV)	5.54	1.25	.90	.22
Control of Learning Beliefs (CB)	5.74	.98	.68	.13
Self-Efficacy (SE)	5.47	1.14	.93	.41
Test Anxiety (TA)	3.63	1.45	.80	-.27
Rehearsal (R)	4.53	1.35	.69	.05
Elaboration (E)	4.91	1.08	.76	.22
Organization (O)	4.14	1.33	.64	.17
Critical Thinking (CT)	4.16	1.28	.80	.15
Metacognitive Self-Regulation (MSR)	4.54	.90	.79	.30
Time & Study Management (TSM)	4.87	1.05	.76	.28
Effort Regulation (ER)	5.25	1.10	.69	.32
Peer Learning (PL)	2.89	1.53	.76	-.06
Help Seeking (HS)	3.84	1.23	.52	.02

Positive correlations were exhibited among intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, and self-efficacy for learning and performance scales. Test anxiety negatively correlated with intrinsic goal orientation, task value, control of learning beliefs, and self-efficacy for learning and performance scales, but was positively related to external goal orientation. All subscales for the learning strategies component were positively related together, correlations coefficients ranged from .05 to .70. The correlations between all fifteen scales of the MSLQ are presented in Table 2.2. Pintrich et al. (1993), claim:

> The six motivational subscales and the nine learning strategies subscales represent a coherent conceptual and empirically validated framework for assessing student motivation and use of learning strategies in the college classroom. The six motivational scales measure

three general components of college student motivation that seem to be distinct factors. In addition, the learning strategy scales represent an array of different cognitive, metacognitive, and resource management strategies that can be reliably distinguished from one another on both conceptual and empirical grounds. Finally, the subscales seem to show promising predictive validity. The motivational scales were related to academic performance in the expected directions. In the same fashion, the learning strategies scales were positively related to course grade. These significant, albeit modest relations with course grade are reasonable, given the many other factors that are related to college course grade that are not measured by the MSLQ (individual course grades themselves are not very reliable measures of performance or learning). The MSLQ seems to represent a useful, reliable, and valid means for assessing college students' motivation and use of learning strategies. (p. 812)

Table 2-2

Correlations Included in the MSLQ Manual Among Motivation and Learning Strategies Scales

	EG	TV	CB	SE	TA	R	E	O	CT	MSR	TSM	ER	PL	HS
IG	.15	.68	.29	.59	-.15	.10	.48	.27	.58	.50	.32	.43	.13	.10
EG		.18	.14	.15	.23	.23	.13	.09	.06	.07	.13	.11	.20	.08
TV			.30	.51	-.14	.12	.44	.19	.39	.45	.37	.47	.09	.16
CB				.44	-.10	.02	.22	.02	.18	.17	.00	.07	-.03	.00
SE					-.37	.10	.35	.21	.42	.46	.32	.44	.05	.08
TA						.11	-.13	-.05	-.11	-.24	-.17	-.21	.10	.08
R							.36	.49	.15	.39	.38	.26	.21	.18
E								.52	.57	.67	.44	.44	.19	.28
O									.31	.55	.44	.36	.23	.22
CT										.53	.25	.25	.25	.19
MSR											.58	.61	.15	.25
TSM												.70	.10	.21
ER													.05	.18
PL														.55

A large number of studies have explored how different theoretical constructs of the MSLQ relate to one another, relate to other factors, and relate to outcome achievement variables like grades or retention. The findings across these studies are numerous and varied. Each study could be used as convergent or discriminant evidence for scales on the MSLQ; however, with such an abundance of findings across studies using the instrument and such a vast array of research methods employed (different scales used, different research questions, different samples, significant versus nonsignificant findings), conclusive evidence for the external validity of any of the subscales becomes muddled and confused.

Consequential Validity

The consequential aspect of validity considers the value implications of score interpretations. Social consequences of a test can be either positive, like when scores lead to improved educational policies, or negative, such as when scores negatively bias students or groups of students (Messick, 1995). The MSLQ scales are modular and are generally used for research purposes. The authors did not include norms for the MSLQ as they posit that responses likely vary as a function of course subject. Developers of the instrument include a feedback form that provide students relative comparisons of their scores to class means and quartile information for nine of the fifteen scales (task value, self-effcacy for learning and performance, test anxiety, rehearsal, elaboration, organization, metacognition, time and study environment management, and effort regulation). Pintrich et al. (1991) included a feedback form so that users of the instrument can share with students how they compare to other students in the class. The form includes a place to report the students' score, the class mean score, and quartile scores that indicate where 25%, 50%, and 75% of participants scored above. Feedback is intended to help students determine their own strengths and weaknesses and can be considered a consequential impact for using the MSLQ. In addition to providing numerical summaries to the students, the developers describe each scale and provide additional feedback about students' scores. For instance, for the organization scale, developers provide the following suggestions to students:

> Outline course material and identify where the text and lecture overlap and don't overlap. This will give you a starting point in developing connections between ideas presented in two different contexts. Make charts, diagrams, or tables of the important concepts. Something like a flowchart or a tree diagram is usually very helpful in trying to understand how different ideas "go together". (p. 57)

The instrument has been used for numerous, diverse research purposes. For instance, researchers have used scales from the MSLQ to a) create new scales (e.g., Eom & Reiser, 2000; Hong & Aqui, 2004; Laszlo & Kupritz, 2003; Niemi, Nevgi & Virtanen, 2003; Nokelainen & Ruohotie, 2000; Schatt, 2011); b) to explore the validity of factors from other inventories by comparing MSLQ factors (e.g., Moak, 2002; Muis, Winne & Jamieson-Noel, 2007; Richardson, 2007; Ross, Blackburn, Salisbury-Glennon, Forbes & Miller, 2002; Yeager, 2010); c) to determine groups such as low and high self-regulated learners (e.g., Arend, 2009; Inpornvijit, 2009; McManus, 2000); d) to assess the effectiveness of program interventions (e.g., Arsal, 2010; Brookhart & Durkin, 2003; Covalt, 1997; Duijnhouwer, Prins & Stokking, 2010; Fisher, 1998; Freeman et al., 2008; Guvenc, 2010; Hammann, Berthelot, Saia & Crowley, 2000; Hancock, 2002; Higgins, 2000; Hofer & Yu, 2003; Kimber, 2009; Kramarski & Michalsky, 2009; Matthews, 2004; McWhorter & O'Connor, 2009; Milner, 2009; Moos & Honkomp, 2011; Paracha, Mohamad, Jehanzeb & Yoshie, 2009; Poole, 1999; Wilke, 2003); e) to compare how different groups respond to scales on the MSLQ (e.g., Barise, 2000; Bidjerano, 2005; Birenbaum & Rosenau, 2006; Blom & Severiens, 2008; Chang, 2009; Exner, 2010; Harlow, 2006; Lynch, 2010; Mullen, 2010; Salamonson, Everett, Koch, Wilson & Davidson, 2009; Sims & Sperling, 2001; Spitzer, 2000; VanZile-Tamsen & Livingston, 1999; Vogt, 2003); f) to explore the relationships of MSLQ factors with other MSLQ factors (e.g., Kaya, 2008; VanZile-Tamsen, 2001); g) to explore the relationships of MSLQ factors with other factors like boredom, procrastination, shame, instructional preference, personality, and self-concept (e.g., Cole & Denzine, 2004; Garcia-Ros, Perez & Talaye, 2008; Garner, 2009; Kesici, Baloglu & Deniz, 2011; Pekrun, Goetz, Daniels, Stupnisky & Perry, 2010; Polleys, 2001; Rey, 2010; Saglam, 2010; Turner & Shallert, 2001; Wolters, 2003); h) to explore the relationships of MSLQ factors with academic achievement (e.g., Campbell, 2001; Carroll & Garavalia, 2004; Chen, 2002; Cobb, 2007; Doljanac, 1995; Fergusson, 2004; Foust, 2008; Howey, 1999; Hsu, 1998; Huang, 2008; Kilic-Bebek, 2010; Kosnin, 2007; Little, 2008; Lynch & Dembo, 2004; Matuga, 2009; Monetti, 2002; Moore, 2007; Niemczyk & Savenye, 2001; Payne, 1992; Puzziferro, 2008; Schutz, Drogosz, White & Distenfano, 1998; Willems, 2001; Yukselturk & Bulut, 2005).

Although users of the instrument may find general differences in groups of learners and their self-report of motivation and learning strategies use, these differences are more likely to be ascribed to contextual and cultural differences between groups of learners. Researchers who have used the MSLQ do not appear to use results from the instrument to negatively bias different learner groups. Overall, it appears that consequential validity of the MSLQ appears acceptable for both intended and actual uses of the instrument.

Substantive Validity

A component of substantive validity refers to the observed consistencies in test responses (Messick, 1995). Reliability estimates provide evidence of the extent to which measurements are repeatable and stable (Crocker & Algina, 1986); therefore, evidence for the substantive validity of scales on the MSLQ can be provided through the reliability of scores obtained as different samples respond to items on an instrument. Authors of the instrument include estimates of Cronbach's alpha to measure the internal consistency of each of the fifteen MSLQ subscales. The alphas included in the MSLQ manual are provided in Table 2.1. (The table also includes additional descriptive information for the MSLQ scales.) Alphas ranged from a low of .52 for the help seeking scale to a high of .93 for the self-efficacy scale. Although the reliabilities reported for two of the subscales are equal to or above .90 and two subscales had alphas between .80 and .89, the majority (ten of the fifteen) of reliabilities reported are between .60 and .79. The developers of the instrument claim that the coefficients for the MSLQ scales are robust and demonstrate good internal consistency (Pintrich et al., 1993).

Reliabilities reported in test or instrument manuals provide evidence for the consistency of scores for the sample in which the test or instrument was developed. These reported coefficients provide a basis for researchers to compare the precision of measurement across tests or similar instruments (Thorndike, 2005). Vacha-Haase (1998) introduced reliability generalization (RG) as a meta-analytic technique to demonstrate the variation of reliability scores across different samples using the same measurement instrument.

Reliability Generalization. Concern about authors referring to the 'reliability of the test' or making statements like 'the test is reliable' promoted Vacha-Haase's development of reliability generalization. Vacha Haase (1998) posits, "such statements contribute to the endemic confusion and misunderstanding of the concept and features of score reliability," (p. 6). Such claims infer that reliability is an immutable quality of an instrument or test and ignores characteristics which can affect score reliabilities (Reese, Keiffer & Briggs, 2002; Thompson, 1994, 1995; Thompson & Vacha-Haase, 2000; Vacha-Haase, 1998).

Reviews of several journal articles and other manuscripts show that the majority of empirical studies fail to report psychometric data for their studies. Green, Chen, Helms, and Henze (2011) reviewed a random selection of empirical articles published in *Psychological Assessment* for years 1989, 1996, and 2006. Results indicate that reliability reporting had not improved much over time as only 21% of authors reported reliability data for their own samples in 1989, 10% in 1996, and 28% in 2006. Vacha-Haase, Ness, Nilsson, and Reetz (1999) reviewed practices regarding the reporting of reliability coefficients in three different psychological journals (*Journal of Counseling Psychology, Psychology & Aging, Professional Psychology: Research and Practice*) from 1990 to 1997, and found that less than 40% of authors provided reliability coefficients for the data being analyzed. In their review, 23% of authors presented reliabilities from previous studies, and 36% of authors made no mention of reliability. In Meier's and Davis' (1990) review of the *Journal of Counseling Psychology* (*JCP*) volumes 1967, 1977, and 1987, the authors note "The majority of the scales described in *JCP* volumes were not accompanied by reports of psychometric properties," (p. 114).

Another common practice of some researchers is to reference reliability scores from a manual or other previous research as sufficient evidence for the reliability of the measure. Vacha-Haase, Kogan, and Thompson (2000) coined this process reliability induction. A review of studies using the Dyadic Adjustment Scale found that slightly less than half (42%) of studies using the instrument claimed the instrument was reliable by reporting estimates from the manual or other previous study or by claiming the instrument was shown to be reliable without providing data to support the claim (Graham, Liu & Jeziorski, 2006). Roughly one-quarter of studies using the Alcohol Expectancy Questionnaire (AEQ) used score reliability information from the original AEQ study or another previous study (Keiffer, Cronin, & Fister, 2004).

Vacha-Haase illustrates the application of meta-analytic techniques to explore reliabilities for a test or instrument by examining reliabilities reported in studies that administered the Bem Sex Role Inventory (BSRI). The BSRI assesses masculine and feminine sex role orientation. Vacha-Haase located 628 articles which included an administration of the BSRI and categorized each article according to how reliability coefficients were reported. The majority of studies (65.8%) did not provide any information for reliability, 21% cited or referenced reliability as having been reported within the manual or other study, and only 13% of the studies (82 total) reported reliability coefficients for the data analyzed.

Of these 82 studies, 57 of the articles reported reliability in a meaningful manner. Reliability coefficients for 87 pairs of masculine (M) and feminine (F) psychological trait scales from these 57 articles were used in subsequent data analyses to examine a) typical score reliabilities for each scale; b) the amount of variability in reliability scores across studies; and c) potential sources of variation in reliability coefficients across studies. To explore potential sources that might affect score reliabilities across studies, several study-specific characteristics were coded including types of reliability coefficients computed, referent used, form length, sample size, gender, article type, language of administration, sample type, and response format.

Results of Vacha-Haase's initial RG study demonstrate that reliability coefficients do vary across studies. For the BSRI, the feminine scale exhibited more variability in reliability of scores than the masculine scale. Sample size and length of test form significantly related to variations in score reliabilities. Numerous RG studies have been conducted over the past decade. Several RG studies were severely limited due to the extreme numbers of cases in which authors failed to report reliability coefficients for their samples (e.g., Beretvas, Suizzo, Durham, & Yarnell, 2007; Caruso & Edwards, 2001; Deditius-Island & Caruso, 2002; Lane, White & Henson, 2002;

Shields & Caruso, 2004; Vacha-Haase, Kogan, Tani & Woodall 2001; Wheeler, Vassar, Worley & Barnes, 2011).

Reliability generalization studies provide persuasive evidence for the typical reliability of scores for a test or instrument and can suggest sample characteristics that may affect reliability scores. In Ross, Blackburn, and Forbes (2005) RG study of the Patterns of Adaptive Learning Survey, researchers found that reliability scores were generally more positive in studies citing the most recent version of the survey. Vassar's and Bradley's (2010) RG study of the Life Orientation Test found that reliabilities for the test were generally acceptable for early stages of research. The authors also found that standard deviation of scores, language, and adolescent respondents related to score reliabilities. In Kieffer and MacDonald's (2011) RG study of the Ways of Coping Questionnaire several sample characteristics (sample size, gender homogeneity, race, population subtype, and age) predicted reliability scores for different subscales. Reliability generalization studies of the motivational scales and learning strategies scales of the MSLQ can be used as further evidence for the reliability of the instrument.

CHAPTER III. MANUSCRIPT 1: MOTIVATED STRATEGIES FOR LEARNING QUESTIONNAIRE: A RELIABILITY GENERALIZATION STUDY OF THE MOTIVATIONAL SCALES

Research on motivation and the relationship motivation has to learning and academic success has received a great deal of attention over the past decades. For educational purposes, motivation research is generally referred to as achievement motivation and relates to a learner's motivation to succeed in academic tasks. Achievement motivation refers to the internal drives that influence a learner's academic interest, involvement, and engagement in the learning process. Motivation involves the mental process students use to activate, sustain, and maintain behavior (Pintrich, 2000a; Pintrich & Schunk, 2002).

The contemporary approach for understanding how motivation relates to academic success is framed within social-cognitive theory which emphasizes the triadic relationship between a learner's cognition, behavior, and environmental factors (Bandura, 1986; Schunk, 2004). Within social cognitive theory, learners are considered active agents in the learning process and the focus of current research relates to how individuals self-regulate the learning process and how this self-regulation influences learning and academic achievement.

Various measurement protocols including self-report questionnaires, structured interviews, teacher judgments, and observations of performance are used to assess aspects of self-regulated learning (Winne & Perry, 2000). Self-report questionnaires are often used to collect data of learners' responses since such instruments are generally cost effective and can be scored easily.

The Motivated Strategies for Learning Questionnaire (MSLQ) includes sections to assess multiple aspects of motivation and cognitive and metacognitive learning strategies use. The motivational section of the MSLQ assesses three separate components of motivation comprised of six factors. Components of the MSLQ include value (goal orientation and task values), expectancy (personal agency beliefs of control of learning and self-efficacy), and affect (test anxiety). These components for understanding motivation reflect motivational systems theory as motivational systems theory also suggests three psychological functions of motivation that includes personal goals, personal agency beliefs, and emotional arousal processes (Ford, 1992).

Motivated Strategies for Learning Questionnaire

The MSLQ assesses motivational orientations and learning strategies of college students in relation to specific college courses. The motivation section of the MSLQ includes 31 items that represent intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, self-efficacy for learning and performance, and test anxiety (Pintrich, Smith, Garcia & McKeachie, 1991). The instrument uses a Likert-type scale with end point anchors of 1 (*not at all true of me*) to 7 (*very true of me*). The instrument has been used for numerous research purposes as the MSLQ is economically feasible to administer as there is no fee required to purchase the inventory and the instrument is readily available within the public domain. In addition, the instrument is easily scored and interpretations for individual student profiles are provided.

Developers of the MSLQ use intrinsic and extrinsic goal orientations to assess students' motives for engaging in learning tasks. Intrinsic goal orientation refers to students engaging in learning tasks for challenge, curiosity, and mastery while extrinsic goal orientation refers to students engaging in learning tasks for reasons such as grades, rewards, performance, competition, and evaluation by others (Pintrich et al., 1991). Four items are used to assess the intrinsic goal and extrinsic goal orientation scales. The third value component, task value, is comprised of six items which refer to students' evaluations of how interesting, how important, and how useful a task is (Pintrich et al., 1991).

The MSLQ includes two expectancy scales, control of learning beliefs and self-efficacy for learning and performance. Control of learning beliefs includes four items that refer to students' beliefs that their efforts will result in positive outcomes. The scale assesses students' beliefs that their success is contingent on their own efforts as opposed to external sources. Self-efficacy includes eight items that measure both the students' performance expectations and their appraisal of their ability to master a task (Pintrich et al., 1991). Developers of the instrument use one affect scale, test anxiety. Five items are used to assess the worry component of test anxiety which refers to students' negative thoughts which could disrupt performance (Pintrich et al., 1991). Examples of items used to represent the six motivation scales are provided in Table 3.1.

Table 3-1

MSLQ Motivation Strategy Scale Sample Items

Learning Strategy Scale (α)	Example Items
Intrinsic Goal Orientation (IG) (α =.74)	"In a class like this, I prefer course material that really challenges me so I can learn new things." "When I have the opportunity in this class, I choose course assignments that I can learn from even if they don't guarantee a good grade."
Extrinsic Goal Orientation (EG) (α = .62)	"The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade." "If I can, I want to get better grades in this class than most of the other students."
Task Value (TV) (α = .90)	"I think I will be able to use what I learn in this course in other courses." "I think the course material in this class is useful for me to learn." "Understanding the subject matter of this course is very important to me."
Control of Learning Beliefs (CB) (α = .68)	"If I study in appropriate ways, then I will be able to learn the course material in this course." "It is my own fault if I don't learn the material in this course."
Self-Efficacy for Learning and Performance (SE) (α = .93)	"I'm confident I can understand the basic concepts taught in this course." "Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class."
Test Anxiety (TA) (α = .80)	"When I take a test I think about how poorly I am doing compared with other students." "I have an uneasy, upset feeling when I take an exam."

Reliability and Reliability Generalization

Score reliability and validity findings provide evidence for whether a scale or instrument provides acceptable psychometric properties and meaningful construct interpretations. During instrument development, researchers recognize the importance of reporting reliability and validity data to facilitate the decision of other researchers to use their instrument. Test reliability provides an index for the effectiveness of an instrument to measure constructs of interest and suggests the extent to which measurements from the constructs are repeatable (Nunnally, 1978; Thorndike, 2005).

Wilkinson and the APA Task Force on Statistical Inference (1999) advise researchers to always examine the reliability of the scores for their data:

> It is important to remember that a test is not reliable or unreliable. Reliability is a property of the scores on a test for a particular population of examinees…Thus, authors should provide reliability coefficients on the scores for the data being analyzed even when the focus of research is not psychometric. Interpreting the size of observed effects requires an assessment of the reliability of the scores. (p. 596)

The assertion that reliability is a property of test scores that should be reported for scores from the sample data analyzed has also been stressed by others (e.g., Dawis, 1987; Thompson, 1994; Thompson & Vacha-Haase, 2000). However, several reviews of practices for reporting psychometric information like reliability coefficients demonstrates that authors seldom consider the reliabilities of scores for their own research samples (Green, Chen, Helms & Henze, 2011; Meier & Davis, 1990; Vacha-Haase, Ness, Nilsson & Reetz, 1999; Willson, 1980). This poor reporting practice of ascribing reliability as a fixed quality of a test or ignoring it altogether disregards factors and conditions that can attribute to poor reliability and subsequently threaten validity of research findings.

Thorndike (2005) notes that multiple factors such as a) the variability of the group on the trait the test measures, b) the level of the group on the trait the test measures, c) the length of the test, and d) the operations used for estimating the reliability can each affect reliability scores. Score unreliability affects statistical power (Henson, 2001), results of statistical significance, and attenuates effect sizes (Baugh, 2002; Hunter & Schmidt, 1994; Thompson & Snyder, 1998).

Vacha-Haase (1998) introduced reliability generalization (RG) as a meta-analytic application for exploring and characterizing score reliabilities of instruments and tests. Since Vacha-Haase's introduction to RG, numerous studies have been conducted to explore the reliabilities of scales on a single instrument (e.g., Henson & Hwang, 2002; Ross, Blackburn & Forbes, 2005; Ryngala, Shields & Caruso, 2005) or to compare the reliabilities achieved across similar instruments (e.g., Henson, Kogan & Vacha-Haase, 2001; Shields, Campfield, Miller, Howell, Wallace & Weiss, 2008). Numerous RG studies affirm the poor reporting practices of researchers in failing to report reliabilities for their samples (e.g., Beretvas, Suizzo, Durham & Yarnell, 2007; Caruso & Edwards, 2001; Deditius-Island & Caruso; 2002; Shields & Caruso, 2004).

Purpose

Since reliability is a necessary condition of validity (Pedhazur & Schmelkin, 1991), it is imperative that researchers confirm the reliability of data achieved with the sample used within their research. The purpose of the present study is to examine the psychometric properties of each motivational subscale of the MSLQ. Although psychometric properties included in the manual for the MSLQ provide researchers point estimates of scale reliabilities, a reliability generalization study utilizes additional information to more broadly generalize the extent to which instruments may yield reliable scores. RG studies can also explore what factors contribute to variations in reliability coefficient estimates. Therefore, a reliability generalization study of the six motivational subscales was used to explore the typical score reliabilities of each scale and to examine sources of measurement error variance across studies.

Methods

Literature Search

Articles using the MSLQ instrument were searched using Academic Premier, ERIC, PsychINFO, PsychARTICLES, Education Research Complete, and Google. Keywords searched included the instrument name, *Motivated Strategies for Learning Questionnaire*, and the common abbreviation, *MSLQ*. Additional studies using the MSLQ instrument were also located through studies retrieved. In totality, the search yielded a list of 363 studies from journal articles, dissertations, conference presentations, and other reports. Of the initial articles found, a total of 84 were not included because a) studies were not written in English (12), b) studies were false hits (1), c) studies could not be located or obtained (35), d) the author of the article only referenced or described the theoretical bases of the MSLQ (27), or e) the MSLQ was used as the framework in developing other scales and instruments (9).

Forty-one articles referenced a junior high version of the MSLQ. Pintrich and Degroot's, (1990) version of the MSLQ instrument is an abridged version validated with secondary students. The shortened version includes 56 items representing three motivation scales and two cognitive components. Articles that used this abridged version are not included in the RG analyses of the MSLQ motivation scales. In addition, 13 articles that used versions of the MSLQ created during the development phase were also eliminated from further analyses.

For the remaining 225 empirical studies using the MSLQ, each study was reviewed and categorized according to the following criteria: a) did not report reliability for participant scores, b) reported reliability coefficients from the MSLQ manual or other study, c) provided reliability coefficients for participants in the study, or d) reported the range or referenced the reliability coefficients for participants in the study. In slightly more than half of these studies, authors reported reliability coefficients for the participants in their study (123 or 54.6%), although authors in 14 of these studies either referenced the reliability for their study or reported ranges of the coefficients. Authors in roughly one-quarter (62 or 27.6%) of the studies referred to reliabilities found in the manual or other studies, and authors in 40 (17.8%) of the studies made no reference to reliability.

Studies in which authors did not report reliabilities at the scale level were excluded from the reliability generalization study. Additionally, studies sharing the same author(s) were reviewed to compare if samples were independent of the other. For studies in which data was exactly replicated, only one instance of the data was used for the RG study. If it could not be verified that a study was replicated due to differences reported in sample sizes, study characteristics, and/or reliabilities, these studies were not excluded but retained for use in the RG study. For the 96 remaining studies, 94 (98%) of the studies used Cronbach's coefficient alpha to estimate reliability, while one study used split-half coefficients, and one study computed a test-retest coefficient with Pearson's *r*. Since the computation of reliability estimates measure different sources of variation in measurement errors, only studies reporting alpha coefficients are used. Of these 94 studies, authors in 76 studies reported alpha coefficients for at least one of the six motivational scales with some authors reporting multiple coefficients of reliability for their samples. Therefore, the total number of reliability coefficients used in analyses for the six motivational scales ranged from .62 for the test anxiety scale to .91 for the intrinsic goal scale.

Coding of Study Characteristics

In addition to recording reliabilities for the motivational scales, study- and sample-specific characteristics were also coded. Each study was examined, and study-specific characteristics were coded for type of article (journal, dissertation or thesis, conference proceeding, book chapter, and unpublished report) and number of anchor items used (5 or 7). Sample characteristics included: sample size, mean age of participants, academic population of sample (elementary, middle, junior high/high school, undergraduate, graduate, or other), specificity level of course subject (studies which ask participants to respond for a specific course subject versus studies which ask participants to generalize their responses), country of study (Amsterdam, Australia, Brazil, Canada, Germany, Greece, Hong Kong, Israel, Malaysia, Netherlands, New Zealand, Norway, Slovenia, South Pacific, Spain, Taiwan, Turkey, United Kingdom, and United States), translation of instrument (English, Chinese, Turkish, German, Dutch, Spanish, Portuguese, and Norwegian), sex (numbers and percentages of males and females), and race (numbers and percentages of Asians, Blacks, Hispanics, Native Americans, Whites, or Other).

Due to small instances of studies conducted in several of the countries included, country of study was recoded into a regional variable that includes Asia, Australia, Europe, Middle East, North America and South America. However, due to the small number of samples for certain regions, the regional variable was then recoded into a dichotomous variable that compares studies completed in the United States to studies completed in other countries. Translation of instrument was also recoded for instruments administered in English and instruments administered in other languages. Academic population was regrouped to compare samples of elementary through secondary level students to samples of postsecondary students. Percentages of Native Americans and persons of other race were not included in analyses because information for these categories was insufficient. Only the percentage of females was used in analyses for comparative purposes. Descriptions of the final variables used in analyses are provided in Table 3.2.

Table 3-2

Coding Strategies for Study- and Sample-Specific Predictor Variables

Variable	Coding Classification (range)	Number of Sample Cases
Journal	0 = Dissertation (55) 1 = Journal article (56)	111
Likert-type scale	0 = 5 point Likert-type scale (13) 1 = 7 point Likert-scale (103)	116
N (Sample size)	Continuous (21 to 2,005; median = 150)	116
Mean age	Continuous (10.4 to 40.4; median = 20.9)	33
Academic population	0 = Elementary through secondary students (14) 1 = Postsecondary students (97)	111
Country	0 = studies conducted in other countries (48) 1 = studies conducted in the United States (68)	116
Translation	0 = Other language (18) 1 = English (85)	103
Course specificity	0 = Courses in general (8) 1 = Specific course (101)	109
Percentage of female participants	Continuous (0 to 100) - Calculated as the proportion of females to total sample size	84
Percentage of Asian participants	Continuous (0 to 100) - Calculated as the proportion of Asian participants to total sample size	35
Percentage of Black participants	Continuous (0 to 100) - Calculated as the proportion of Black participants to total sample size	31
Percentage of Hispanic participants	Continuous (0 to 100) - Calculated as the proportion of Hispanic participants to total sample size	33
Percentage of White participants	Continuous (0 to 100) - Calculated as the proportion of White participants to total sample size	40

Analyses

Reliability generalization adapts meta-analytic techniques to explore the typical reliability coefficient for a scale or instrument and to investigate sample characteristics that contribute to variation between reliability coefficients across studies. Typical reliabilities for scales on the motivational section of the MSLQ can be summarized through measures of central tendency such as mean, median, and mode, and accompanied by measures of dispersion such as standard deviation and interquartile range. Likewise, a number of statistical approaches can be employed to characterize the relationships between reliability estimates and different coded variables.

In Hunter's and Schmidt's (1977) early work with validity generalization studies, the researchers advocated a transformation of correlation and reliability coefficients using Fisher's *z*, as Fisher's *z* was believed to correct for departures in normality for correlational coefficients and thus provided more accurate estimates. However, Hunter and Schmidt (1990) modified their position in using Fisher's *z* transformation after evidence indicated that the transformation is positively biased, i.e., it gives larger weights to larger correlations than to smaller ones. Still several RG studies utilized this meta-analytic technique of transforming reliability coefficients using Fisher's *z* transformation (e.g., Beretvas et al., 2008; Campbell, Pulos, Hogan & Murry, 2005; Graham & Christianson 2009; Graham, Liu & Jeziorski, 2006; Lopez-Pina, Sanchez-Meca & Rosa-Alcazar, 2009; Miller, Byrne, Rutherford & Hansen, 2009; O'rourke, 2004; Wallace & Wheeler 2002). Several RG studies compared findings using original reliabilities and Fisher's *z* transformation and noted negligible differences between findings (Caruso & Edwards 2001; Ryngala et al., 2005; Shields & Caruso, 2004). Furthermore, promoters of reliability generalization have also argued that the transformation is unnecessary as coefficient alpha is a squared metric statistic (Henson & Thompson, 2002; Thompson & Vacha-Haase, 2000).

Feldt and Charter (2006) used Monte Carlo procedures to compare different conceptual approaches for averaging internal consistency coefficients and found that analyzing the original reliability coefficients is sufficient. Researchers compared six approaches for calculating averages for coefficients of internal consistency including the use of original values, weighted averages, and Fisher's *z* transformation and findings were similar across all approaches. Since there appears to be little difference in using a transformation of the reliability coefficients versus the original data coefficients, all data analyses for reliabilities of the MSLQ motivational scales will be conducted using original data coefficients.

The decision about which statistical techniques to employ for data analysis is restricted due to significant amounts of data not reported across empirical studies that used the MSLQ. Rampant missing data is a common and substantial limitation in many RG studies conducted to date (Beretvas et al., 2007; Caruso & Edwards, 2001; Deditius-Island & Caruso; 2002; Shields & Caruso, 2004). To make the best use for data that was available in their study of the Coopersmith Self-Esteem Inventory, Lane, White, and Henson (2002) opted to run individual statistical tests for each predictor variable which does amplify the experimentwise error rate. However, to minimize loss of information, the researchers sacrificed considerations of statistical significance and opted to utilize effect sizes as an indication for understanding characteristics that attributed to variations in reliability for their RG study. A similar approach is used to explore reliabilities of the MSLQ motivational scales. Regression analysis is used for continuous scored predictors and analysis of variance (ANOVA) conducted for discrete (nominal) variables. Effect size indices to approximate the proportion of variance in the dependent variable accounted for by the independent variables included in the analyses are η^2 (eta squared) and ω^2 (omega squared) for categorical predictors and R^2 (coefficient of determination) and Adjusted R^2 for continuous predictors. Omega squared (ω^2) and Adjusted R^2 are provided as each statistic theoretically corrects for the influence of sampling error. Means for each categorical group are included for discrete variables while beta coefficients and structure coefficients are also provided for each continuous predictor.

Results

In total there are 117 sample cases in which reliability was reported for at least one of the six motivational scales. Sample sizes range from 21 participants to 2,005 participants with a median sample size of 150 participants. Average participant ages range from 10.4 years to 40.4 years with a median age of 20.9 years, with most reliabilities reported for college student samples (83%). The majority of studies asked students to respond to questionnaire items in response to a specific course (101 or 93%). A recoded regional variable for countries included 70 cases for North America, 21 cases for Europe, 14 for the Middle East, 8 cases for Asia, 2 for Australia, and 1 for South America. However, due to small sample sizes for many of these regions, country was recoded into a variable to compare the United States (58%) to all other countries (42%). The majority of samples use an English version of the MSLQ (72.6%). Fifty six (48%) of the reliabilities are reported in journal publications, 55 (47%) from dissertations or theses, and 6 (5%) from conference proceedings. A 7-point Likert-type scale was used in 88% of the cases and a 5-point Likert-type scale was used in the remaining cases.

The descriptive statistics for the MSLQ motivational scales are presented in Tables 3.3 and 3.4. The tables includes the total number of coefficients found for each scale (*n*), the five-number summary, means and standard deviations. Figure 3.1 displays box and whiskers plots for each of the six motivational scales and indicates instances of mild and extreme outlying reliabilities. Outliers are observations with "*a unique combination of characteristics identifiable as distinctly different from other observations*" (Hair, Black, Babin, Anderson & Tatham, 2006 p. 73). Although outliers can be indicative of extraordinary events or observations, they can also frequently arise from procedural errors (data entry errors or mistakes in coding). Without having access to the original data sets to explore causes of the mild and extreme outliers, all observations were converted to *z*-scores, and values which exceeded three standard deviations from the mean were deleted prior to additional analyses. See Table 3.3 for the number of extreme observations associated with each scale.

Table 3-3

Descriptive Statistics for Reliabilities from the MSLQ Motivational Scales

Scales	*N*	Min	25th	50th	75th	Max	Mean	*SD*	Manual	Extreme Outliers
IG	91	.31	.66	.71	.76	.90	.70	0.09	.74	2
EG	85	.40	.62	.67	.74	.87	.67	0.09	.62	2
TV	74	.68	.80	.87	.91	.96	.85	0.07	.90	0
CB	66	.22	.55	.67	.73	.92	.65	0.13	.68	1
SE	86	.33	.86	.90	.92	.96	.87	0.10	.93	2
TA	62	.56	.73	.76	.81	.92	.76	0.07	.80	0

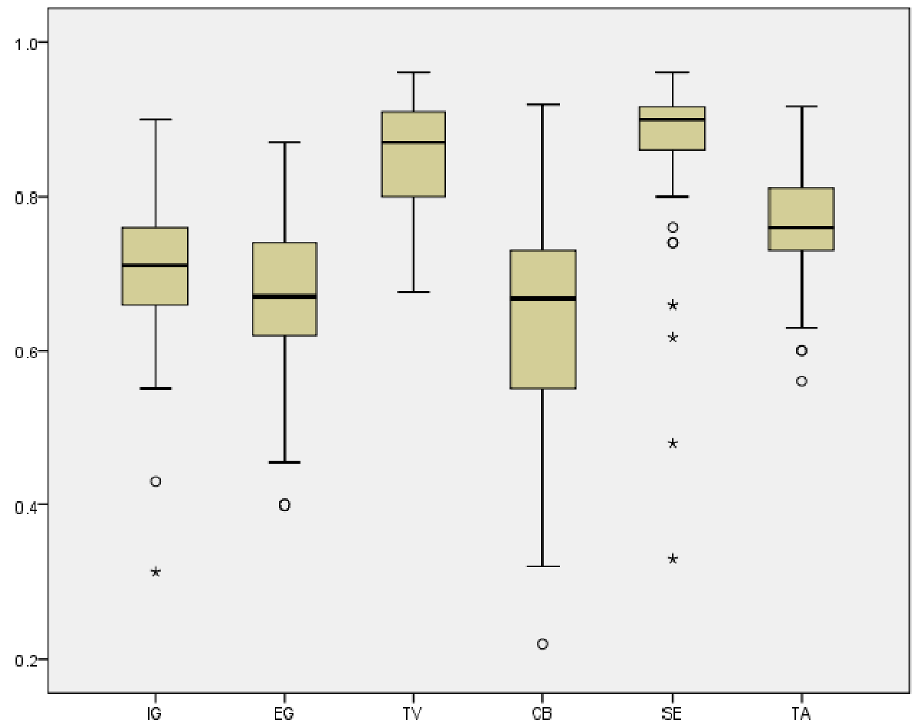

Figure 3-1. Box-and-whiskers plot of each MSLQ motivational scales

Table 3.4 displays typical magnitudes and variability for each of the six subscales after outlier deletion. In general, reliability coefficients for these scales were fair (mean values ranged from .65 for control of learning beliefs to .87 for self-efficacy). Mean reliabilities for the self-efficacy and task value scales exceed .80 and appear fairly robust. The mean values for reliability coefficients slightly differ from the reliability coefficients presented in the MSLQ manual. However, from a practical standpoint, these minor fluctuations in values would likely have relatively little influence on someone's interpretation of whether or not the scale would be assumed to produce adequately reliable results when used in future studies.

Table 3-4

Descriptive Statistics for Reliabilities from the MSLQ Motivational Scales after Omission of Outliers

Scale	*N*	Min	25th	50th	75th	Max	Mean	*SD*	Manual
IG	89	.55	.66	.71	.76	.90	.71	0.07	.74
EG	83	.46	.63	.67	.74	.87	.68	0.08	.62
TV	74	.68	.80	.87	.91	.96	.85	0.07	.90
CB	65	.32	.56	.67	.73	.92	.65	0.12	.68
SE	84	.62	.86	.90	.92	.96	.88	0.06	.93
TA	62	.56	.73	.76	.81	.92	.76	0.07	.80

Visual representations for the distribution of alpha coefficients obtained for each of the six motivational subscales are displayed in Figures 3.2-3.7. Figures include normal probability plots which compare the cumulative distribution of actual data values with the cumulative distribution of a normal curve. Departures from the straight diagonal line indicate departures from the normal distribution. Histograms were inlaid within each normal probability plot to further portray the distribution of reliability coefficients. In addition to visual representations, Shapiro-Wilk tests for normality were conducted for each scale. Significant departures from normality were evidenced for task value $\{W(74) = 0.94, p < .01\}$ and self-efficacy $\{W(84) = 0.79, p < .01\}$. Both the task value and self-efficacy scales exhibit negatively skewed distributions. Despite the deviations from normality for these scales, transformations were not completed as other RG studies have found negligible differences in results using transformed scales versus original data (Caruso & Edwards 2001; Ryngala et al., 2005; Shields & Caruso, 2004).

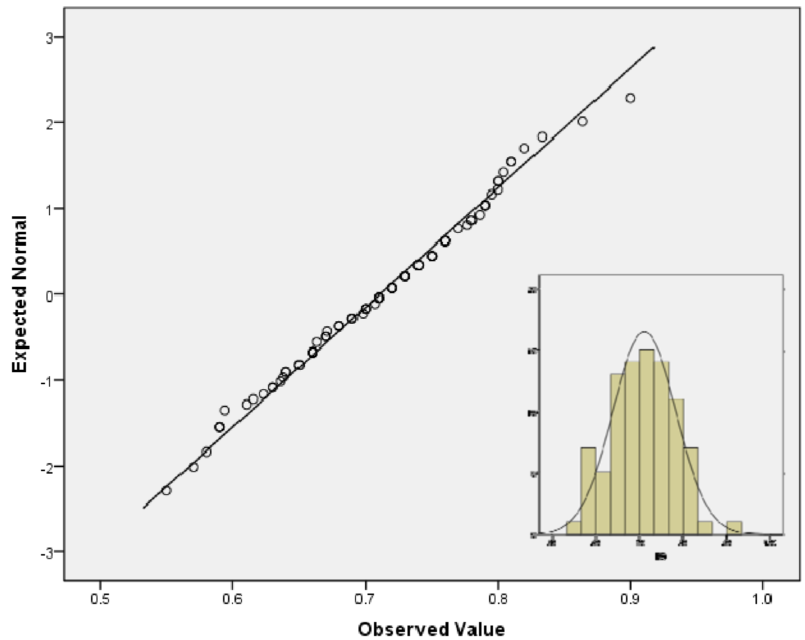

Figure 3-2. Normal probability plot and histogram for distribution of α for IG.

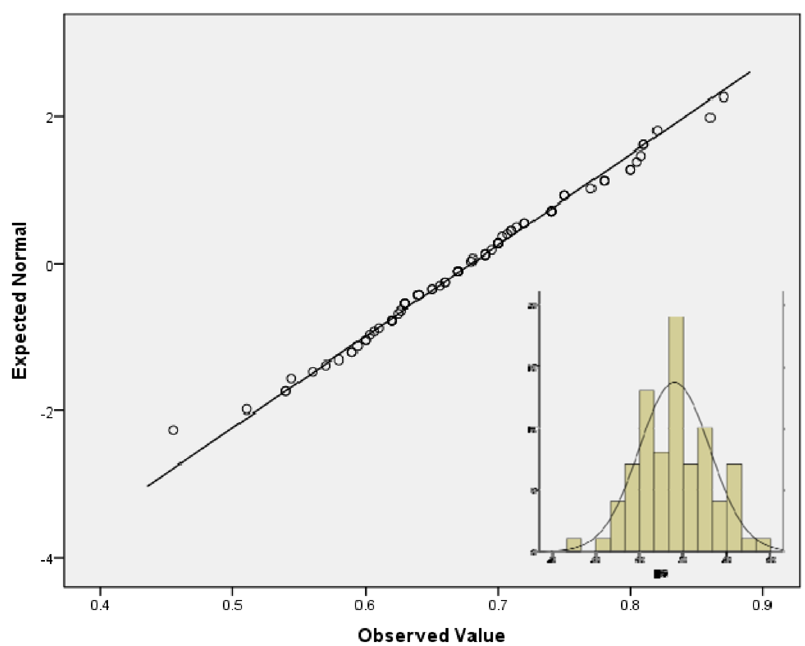

Figure 3-3. Normal probability plot and histogram for distribution of α for EG.

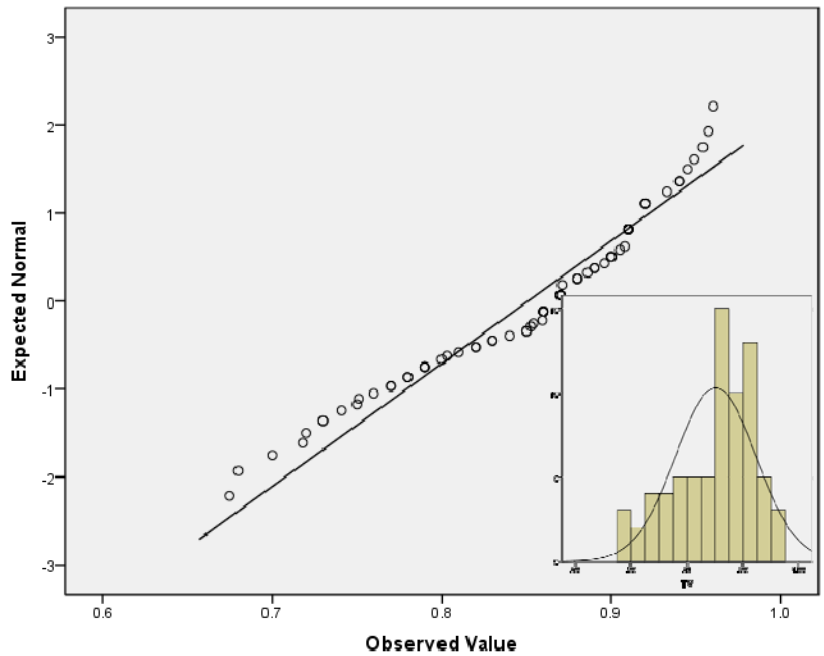

Figure 3-4. Normal probability plot and histogram for distribution of α for TV.

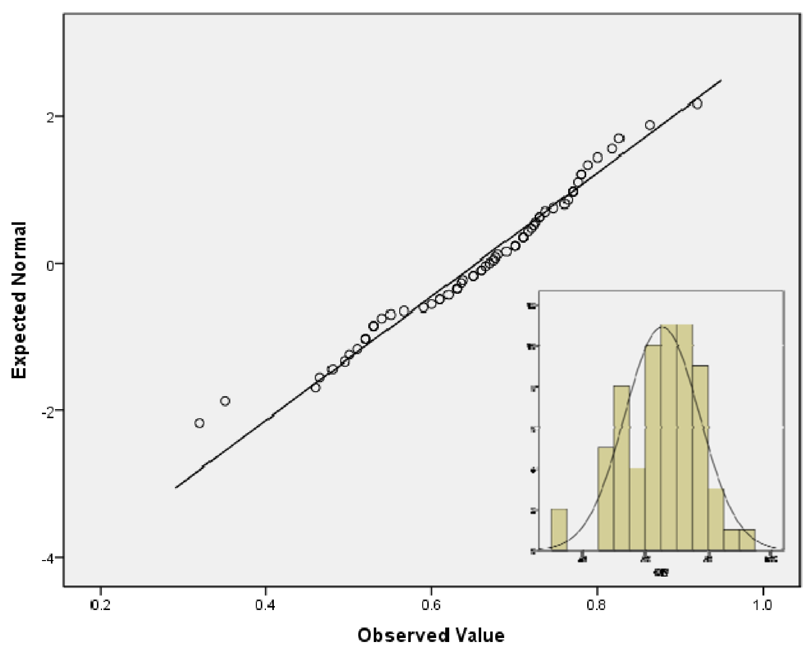

Figure 3-5. Normal probability plot and histogram for distribution of α for CB.

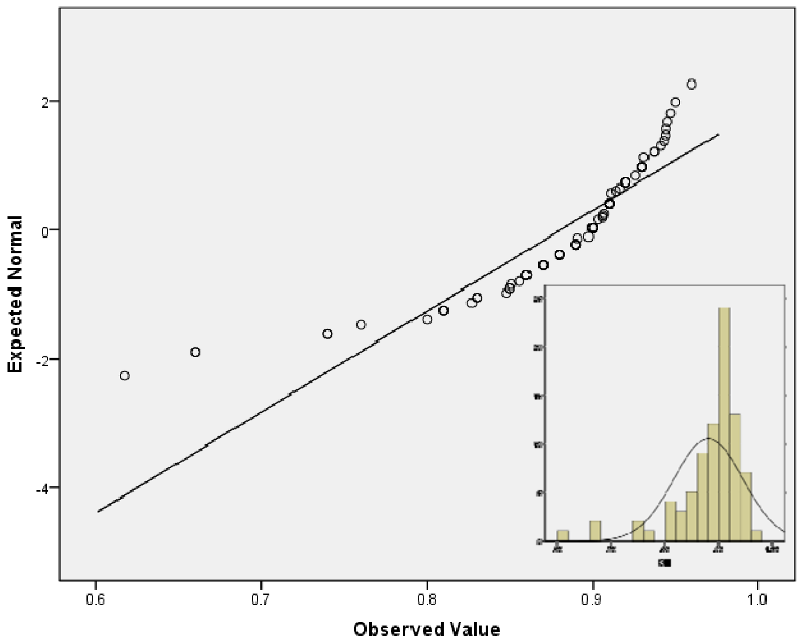

Figure 3-6. Normal probability plot and histogram for distribution of α for SE.

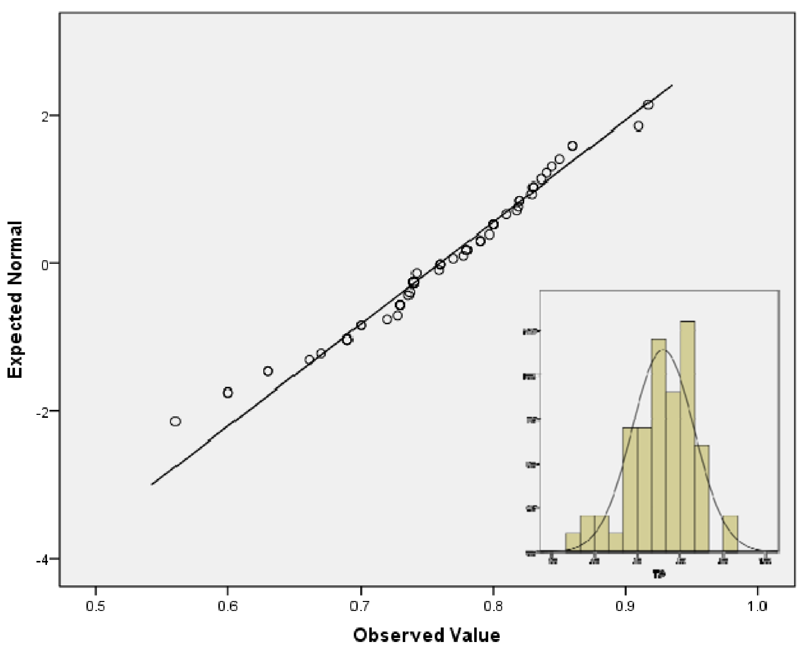

Figure 3-7. Normal probability plot and histogram for distribution of α for TA.

The results of ANOVA analyses used with categorical predictor variables are displayed in Table 3.5 with means for each categorical variable included in Table 3.6. Due to insufficient data, statistical significance is not considered and only η^2 and ω^2 are provided to portray the magnitude of effects for the predictors with MSLQ motivational scales reliabilities. Negligible to small differences in reliabilities were found for article type, number of Likert-type items used, and level of course specificity (effect sizes ranged from .00 to .08). Similarly, small differences in mean reliabilities were evident between academic populations and each motivational scale, although mean reliabilities did tend to be higher for college-aged samples versus younger samples for the task value, self-efficacy, and test anxiety scales.

Small effect sizes were found for five of the six MSLQ motivational scales when comparing mean reliabilities for samples from the United States versus samples from other countries. Country of origin accounted for approximately 20% of the variance in reliability estimates for the test anxiety scale with reliabilities being more positive with samples from the United States (M = .78) versus other countries (M = .71). Similarly, a slightly larger effect (η^2 = .29; ω^2 = .27) was also found for translation of the instrument and test anxiety. Mean reliabilities for the test anxiety scale were more positive for samples using English versions of the instrument (M = .78) than for samples using a translated version of the instrument (M = .69). Although minute, effects were also evident with translation of the instrument and intrinsic goal orientation (η^2 = .13; ω^2 = .12) and with task value (η^2 = .14; ω^2 = .13). Mean reliabilities tended to be higher with an English translation of the instrument for both scales.

Table 3-5

Effect Sizes of Differences Between Score Reliabilities for Categorical Predictors of MSLQ Motivational Scales

	IG			EG			TV			CB			SE			TA		
Variable	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η^2	ω^2	*n*
Article	0.03	0.02	83	0.02	0.00	76	0.04	0.02	68	0.06	0.05	59	0.02	0.01	78	0.03	0.02	56
Likert-Type Scale	0.00	-0.01	88	0.04	0.02	82	0.01	0.00	73	0.02	0.01	64	0.08	0.07	82	0.01	0.00	61
Academic Pop.	0.00	-0.01	84	0.01	0.00	78	0.08	0.06	67	0.01	-0.01	59	0.08	0.06	77	0.09	0.07	57
Course Specificity	0.05	0.04	81	0.00	-0.01	75	0.00	-0.01	65	0.00	-0.02	57	0.00	-0.01	75	0.00	-0.02	55
Geographic Region	0.05	0.04	88	0.00	-0.01	81	0.04	0.03	72	0.04	0.03	63	0.02	0.01	82	0.20	0.19	61
Translation	0.13	0.12	78	0.01	0.00	72	0.14	0.13	63	0.03	0.01	54	0.03	0.02	71	0.29	0.27	53

Table 3-6

Descriptive Statistics for Score Reliabilities and Cell Frequencies for Categorical Predictors

	IG			EG			TV			CB			SE			TA		
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Article																		
0 = Other	.72	0.07	42	.67	0.09	37	.86	0.07	42	.68	0.12	35	.89	0.07	47	.77	0.07	35
1 = Journal	.70	0.07	42	.69	0.07	40	.83	0.07	27	.62	0.11	25	.87	0.05	32	.75	0.07	22
Likert-Type Scale																		
0 - 5 point scale	.73	0.03	5	.63	0.08	6	.83	0.1	8	.71	0.13	6	.82	0.09	8	.78	0.07	7
1 - 7 point scale	.71	0.07	84	.68	0.08	77	.85	0.07	66	.65	0.12	59	.89	0.06	75	.76	0.07	55
Academic Population																		
0 = Secondary	.72	0.06	8	.71	0.11	8	.80	0.07	9	.62	0.07	8	.83	0.07	9	.70	0.09	7
1 = Postsecondary	.71	0.07	77	.68	0.08	71	.86	0.07	59	.66	0.13	52	.89	0.06	69	.77	0.07	51
Course Specificity																		
0 = General	.64	0.04	3	.70	0.05	4	.84	0.05	3	.67	0.18	3	.87	0.05	7	.76	0.06	3
1 = Specific	.72	0.07	79	.68	0.09	72	.85	0.08	63	.66	0.12	55	.88	0.06	69	.76	0.08	53
Geographic Region																		
0 = Other	.69	0.07	39	.69	0.08	38	.83	0.06	24	.62	0.11	21	.87	0.04	28	.71	0.07	19
1 = United States	.72	0.07	50	.67	0.08	44	.86	0.07	49	.67	0.12	43	.89	0.07	55	.78	0.06	43
Translation																		
0 = Other	.66	0.06	13	.66	0.06	13	.80	0.06	15	.62	0.1	13	.87	0.03	16	.69	0.07	12
1 = English	.72	0.06	66	.68	0.08	60	.87	0.07	49	.67	0.13	42	.90	0.06	56	.78	0.06	42

The results of regression analyses used to explore the relationship of continuous variables with reliabilities are shown in Table 3.6. Statistical significance was not included. Beta coefficients (b), structure coefficients (r_s), R^2 and adjusted R^2 are provided to demonstrate the relationships between each predictor variable with the six motivational scales.

The mean age of participants appears to have the strongest influence (albeit a small to moderate relationship) on reliability scores for the motivational scale, intrinsic goal orientation. The correlation between the two variables is r_s = -.27 with an R^2 = .07. However, with only 13 cases providing mean ages and reliabilities, there is insufficient data to confirm such a relationship is valid. The relationships between the remaining continuous variables and reliabilities for intrinsic goal orientation appear negligible.

The percentage of female (R^2 = .07), Hispanic (R^2 = .12), and White (R^2 = .13) participants appear to have some minor effect on the reliability scores for extrinsic goal orientation. Both the percentage of female (r_s = -.27) and percentage of White (r_s = -.35) participants relate negatively to extrinsic goal scores; hence, as the percentage for these two groups increase, reliabilities tend to decrease. The percentage of Hispanic (r_s = .35) participants positively relates to extrinsic goal scores.

Sample size, mean age of participants, percentage of Asian participants, and the percentage of Hispanic participants show a small to moderate linear relationship with reliability scores for task value, although the characteristics account for less than 10% of the variance in reliability estimates (see Table 3.6). These relationships were all negative, except for the relationship between task value and mean age. The sample size for mean age and task value reliabilities is small. Caution should be taken when interpreting the relationship between these two variables.

The percentage of Asian participants explains roughly 15% of the variance in reliability estimates for control of learning beliefs. The percentage of Asian participants is negatively related to control of learning belief reliability scores (r_s = -.38). Other continuous predictors appear to have little effect on the reliability of scores attained for the control of learning beliefs scale.

The percentage of Hispanic participants explains 37% of the variation in score reliabilities for the self-efficacy scale with a correlation of -.61. The percentage of Asian participants account for a smaller proportion of variance with self-efficacy scores (R^2 = .12) but also demonstrate a negative relationship with self-efficacy scores (r_s = -.35). The remaining continuous variables appear to have little effect of the reliability of self-efficacy scores. The percentage of Asian participants (R^2 = .15) and percentage of female participants (R^2 = .09) explain the greatest variance within reliabilities for the test anxiety scale as the remaining variables each accounted for less than 3% of the variance in test anxiety reliability scores.

Table 3-7

Regression Statistics for Continuous Predictors of MSLQ Motivational Scales

MSLQ Scale	Statistic	Sample Size	Mean Age	% Female	% Asian	% Black	% Hispanic	% White
	b	.000	-.003	.000	.000	.000	.000	.000
	r_s	-.110	-.266	-.151	-.169	-.070	.043	.034
IG	R^2	.012	.071	.023	.028	.005	.002	.001
	Adj R^2	.001	-.007	.006	-.014	-.038	-.038	-.031
	n	88	13	59	24	24	26	32
	b	.000	.001	-.001	.000	.000	.001	-.001
	r_s	-.022	.061	-.269	.040	.139	.348	-.354
EG	R^2	.000	.004	.072	.002	.019	.121	.126
	Adj R^2	-.012	-.096	.055	-.040	-.027	.083	.094
	n	82	11	56	25	22	24	29
	b	.000	.002	.000	.000	.000	-.001	.000
	r_s	-.213	.259	.059	-.302	.081	-.229	-.033
TV	R^2	.045	.067	.004	.091	.007	.053	.001
	Adj R^2	.032	-.005	-.013	.055	-.037	.011	-.032
	n	72	14	61	26	24	24	31
	b	.000	-.004	-.001	-.001	-.001	-.001	.000
	r_s	-.125	-.229	-.201	-.379	-.142	-.203	-.041
CB	R^2	.016	.052	.040	.144	.020	.041	.002
	Adj R^2	.000	-.053	.022	.105	-.029	-.004	-.035
	n	64	10	55	23	21	22	28
	b	.000	.001	.000	.000	.000	-.001	.001
	r_s	-.078	.111	.012	-.352	.066	-.607	.272
SE	R^2	.006	.012	.000	.124	.004	.368	.074
	Adj R^2	-.006	-.040	-.015	.093	-.034	.345	.047
	n	82	20	68	29	27	28	35
	b	.000	.000	.001	-.001	.000	.000	.000
	r_s	-.062	.015	.296	-.385	-.068	-.167	.103
TA	R^2	.004	.000	.087	.148	.005	.028	.011
	Adj R^2	-.013	-.143	.070	.111	-.039	-.016	-.023
	n	61	8	53	24	24	23	30

Discussion

The purpose for conducting a reliability generalization study on MSLQ motivational scales was a) to identify typical score reliabilities for each of the six motivational scales, and b) to determine sample characteristics that contribute to score variability across studies. Only 54% of empirical studies reviewed had authors report reliability estimates for the data in hand. Authors failed to report any type of reliability information about scales on the MSLQ in 18% of reviewed studies and referenced reliability from the manual or another previous study in 28% of reviewed studies. The large amount of missing reliabilities from studies that do not provide data for their samples is a limitation of the reliability generalization study.

For authors who reported alpha, reliability generalization methods were used to investigate the typical reliabilities and variability of reliability estimates for the six motivational scales of the MSLQ. Although what constitutes a level of acceptable reliability scores is debatable, results suggest that a number of researchers would consider reliabilities for the six motivational scales acceptable for research purposes (Kelley, 1927, Nunnally, 1967, 1978; Nunnally & Bernstein, 1994; Thorndike 1951). Reliabilities for the self-efficacy and task values scales were generally high, and mean reliabilities were at or above .85. The smallest reliabilities were evident for the control of learning and the extrinsic goal scales. Mean reliabilities for these scales were between .60 and .70.

A second aim of the reliability generalization study was to investigate factors that might affect reliabilities of the six motivational scales. Reliability generalization research considers how study characteristics such as type of article and number of anchor items used as well as how numerous sample-related characteristics (e.g., sample size, percentage of female participants, race, etc) relate to score reliabilities of the scales. Researchers can use the results of reliability generalization analyses to regard which conditions may relate to higher or lower score reliabilities (Thompson, 1999; Vacha-Haase, 1998).

Study-specific characteristics for article type and number of anchor items used were examined to explore if differences were evident in reliability scores of the MSLQ motivational scales. No significant differences exist across the six motivational scales for either of these study-specific factors. A comparison of reliabilities for samples who were asked to respond to motivational items with respect to a specific subject course versus samples providing general responses to motivational items also revealed negligible differences. However, there may not be enough cases in which students were asked to provide general appraisals to motivational items to determine if reliabilities would be lower for such general appraisals versus reflections for a specific course subject.

The MSLQ was developed to assess motivation of college student populations, but has also been used with younger student populations. There appears to be some minor differences in reliability scores for task value, self-efficacy and test anxiety scales. For each of these scales, reliabilities were more positive for postsecondary samples. Mean age was also considered to examine reliabilities between younger and older students. However, due to the insufficient number of cases providing ages for their samples, results are inconclusive.

To determine if reliability scores differ for samples not from the United States, country was coded into a dichotomous variable to compare all studies conducted in the United States to studies conducted outside of the United States. Although, comparing reliability scores for each country would offer stronger evidence for where cultural differences may exist, sample sizes in many of the sample countries restricted analyses to this broader comparison. A moderate effect is evident for the test anxiety scale with lower reliability scores for samples outside the United

States. Similarly, reliability scores for the test anxiety scale were more positive for English translations of the instrument. Small effects were also evident for scales of intrinsic goal orientation and task value in which mean reliabilities were higher for English translations of these motivational scales.

Finally, the percentage of female participants and the percentage of participants composing different racial groups were also used to explore variations in reliability coefficients. Relationships between these sample characteristics and reliabilities for intrinsic goal orientation and task value are rather minute. The percentage of Asian participants has a small, negative effect on reliability scores for test anxiety, control of learning beliefs, and self-efficacy. The percentage of Hispanic participants were slightly, positively related to reliabilities for extrinsic goal orientation. A more pronounced, but negative relationship for the percentage of Hispanic participants and reliabilities scores for self-efficacy is also evident. A small, negative effect is evident between extrinsic goal reliabilities and the percentage of White participants. However, race was simplistically operationalized as the percentage of participants described as Asian, Black, Hispanic, Native American, White, or Other (where other may also include not reported or combination of different minority races which were not specified). Many of the studies failed to provide descriptions regarding the racial identity of their participants, thus the affects of ethnicity on reliability scores warrants further investigation to determine why reliabilities would be higher or lower for different scales and groups.

Limitations

A leading limitation of many reliability generalization studies completed to date is the failure of researchers to include reliability estimates for samples in their studies (e.g., Beretvas et al., 2007; Caruso & Edwards, 2001; Deditius-Island & Caruso; 2002; Shields &

Caruso, 2004). Only slightly more than half of researchers who used the MSLQ actually reported reliability estimates for participants in their sample. The remaining authors either failed to mention reliability or referenced reliability from the manual or other study. In addition to the lack of reliability reporting in the available literature, it is also plausible that numerous researchers using the MSLQ instrument did not or could not publish their findings due to insignificant results. Rosenthal (1979) termed the tendency to publish positive research that demonstrates patterns of statistical significance while omitting studies in which researchers reached negative conclusions, or an absence of statistical significance the ‘file drawer problem’.

In addition to missing data on scale reliabilities, a number of studies did not include adequate descriptions of sample characteristics. With such a pervasive amount of data missing for many sample characteristics, data analyses were restricted to individual tests conducted on each predictor variable with each set of reliability coefficients. Results of statistical significance were not considered as the large number of tests inflates experimentwise error rates. Although this method provides evidence to the relationships between variables, more advanced statistical techniques can explore the relative predictive importance of each variable in combination with the other.

Conclusion

Reliability generalization is an important meta-analytic technique to examine score reliability. The results of the reliability generalization study on the motivational scales of the MSLQ demonstrate that across a variety of samples, the six scales tended to yield acceptable reliability estimates. Mean reliability scores exceeded .70 for four of the six scales and mean scores for extrinsic goal orientation and control of learning beliefs were just slightly less.

Aspects of both study-specific and sample-specific characteristics did affect the measurement of different motivation scales. Researchers who are concerned for the effects these factors can have on scale measurements, may wish to modify or add additional scale items to increase reliability of scores for their response populations. The reliability generalization study of the motivation scales of the MSLQ highlights that reliability is not an immutable property of the scale, but differs across samples. This study provides further evidence for the need of social science researchers to always examine the reliabilities of the scores their own data.

CHAPTER IV. MANUSCRIPT 2: MOTIVATED STRATEGIES FOR LEARNING QUESTIONNAIRE: A RELIABILITY GENERALIZATION STUDY OF THE LEARNING STRATEGIES SCALES

The efficiency with which the Motivated Strategies for Learning Questionnaire (MSLQ) can detect differences in learning strategy use is contingent upon the degree of the reliability of response scores. Within the measurement context, reliability can broadly be considered the extent to which measurements are repeatable or stable (Crocker & Algina, 1986) and provide evidence for the amount of random error captured within a measurement (Guilford, 1954, Pedhazur & Schmelkin, 1991). Manuals for tests and other instruments include descriptors of sample reliabilities computed during development of the test or instrument. Nunnally (1978) and Thorndike (2005) recommend that researchers consider the reliability of a test as an index of the effectiveness of the test, communicating an extent to which the results obtained from a measurement are repeatable.

A common reporting practice within social-behavioral and educational research is to disregard the reliability of scores for the sample for which data is collected. Too often, researchers who use instruments or tests to measure personal attributes, reference the reliability found during the development of the test or instrument or reliability found for another previous study, and not the reliability produced for their research sample (Vacha-Haase, Kogan & Thompson, 2000). Other researchers neglect to provide any information regarding the reliability of scores either for their sample or for other instances in which the instrument or test was used

(Green, Chen, Helms & Henze, 2011; Meier & Davis, 1990; Vacha-Haase, Ness, Nilsson, & Reetz, 19999; Willson, 1980). This infrequent reporting of reliability in research suggests that researchers may assume reliability is an inherent attribute of the test or instrument. However, reliability coefficients can be affected by heterogeneity of the group for the trait being measured, level of the group on the trait being measured, length of the test or instrument, and the operations used for estimating the reliability (Feldt & Brennan, 1993; Thorndike, 2005).

Poor reliability can indicate that response items for a test or instrument do not accurately reflect the construct being measured or can indicate unique qualities about the sample of interest (Dawis, 1987). Score inconsistency attenuates effect sizes and can misrepresent comparisons among groups of individuals (Baugh, 2002; Hunter & Schmidt, 1994; Thompson & Snyder, 1998) as well as affects statistical power (Henson, 2001). Wilkinson and the APA Task Force on Statistical Inference (1999) emphasize that researchers should consider the reliabilities for scores in their samples.

Vacha-Haase (1998) proposed reliability generalization (RG) as a possible and important tool for characterizing variability in reliability scores and examining study features that relate to these variations in score quality. Since Vacha-Haase's introduction to reliability generalization, numerous RG studies have been conducted (e.g., Beretvas, Suizzo, Durham & Yarnell, 2007; Caruso, 2000; Ross, Blackburn & Forbes, 2005).

MSLQ Learning Strategies Scales

The MSLQ is composed of two sections, and the second section includes items designed to assess the learning strategies usage of college students in relation to specific college courses. The learning strategies section of the MSLQ includes 50 items partitioned into two components: a component for cognitive and metacognitive strategies which includes scales for rehearsal, elaboration, organization, critical thinking, and metacognitive self-regulation; and a component for resource management strategies which includes scales for time and study management, effort regulation, peer learning, and help seeking (Pintrich, Smith, Garcia, McKeachie, 1991; 1993).

The learning strategies scales were based on a general cognitive model of learning and informational processing. Learning strategies can be defined as “behaviors and thoughts that a learner engages in during learning and which are intended to influence the learner’s encoding process,” (Weinstein & Mayer, 1986, p. 315). Developers of the instrument include four items to measure basic rehearsal strategies which allow learners to process information of simple tasks into working memory. Rehearsal strategies include reciting or naming items from a list. Elaboration strategies help students to process information into long-term memory by using tactics such as paraphrasing or summarizing material. These tactics can build internal connections between the new information and prior knowledge. Six items are used to measure elaboration strategies. Four items are used to assess organization strategies like outlining material or concept mapping. Such strategies are considered higher order learning strategies and can help learners select the appropriate information to be learned (Pintrich et al., 1991; Weinstien & Mayer, 1986). Critical thinking involves processes for applying previous knowledge to new situations to effectively problem-solve. Developers of the MSLQ use five items to assess students’ critical thinking (Pintrich et al., 1991).

Metacognition is considered a higher order thinking process that refers to the awareness, knowledge and control of cognition (Pintrich et al., 1991). Flavell (1992) describes knowledge of cognition as the knowledge of learning tasks, strategies, instructional plans and goals; and the regulation of cognition as the learner’s goal setting, planning, monitoring of understanding, and evaluating of progress towards the completion of the task. Developers of the MSLQ use twelve items that focus solely on the control and self-regulation aspects of metacognition. All three aspects regarding the control of cognition (goal setting, planning, monitoring one's understanding, and evaluating progress towards the completion of the task) are combined to represent the metacognitive self-regulation scale.

Developers of the MSLQ also included components to assess students' use of resource management strategies, tools learners can use to enhance learning. Eight items are used to represent time and study management, which refers to both the management of one's study time and the selection of study settings. Effort regulation, represented by four items, refers to the effort a student expends to reach learning goals. Peer learning is a strategy of collaborating with one's peers, which can help learners clarify course material, while help seeking involves seeking the support of others (peers, instructors, etc.) to help students navigate the learning material. Developers use three items to assess students' engagement with peer learning and four items with help seeking (Pintrich et al., 1991). Examples of different items used to represent all nine learning strategies scales are provided in Table 4.1. The reliabilities reported in the manual are also provided in the table.

Table 4-1

MSLQ Learning Strategy Scale Sample Items

Learning Strategy Scale (α)	Example Items
Rehearsal (R) (α = .69)	"When I study for this class, I practice saying the material to myself over and over." "I make lists of important terms for this course and memorize the lists."
Elaboration (E) (α = .76)	"I try to understand the material in this class by making connections between the readings and the concepts from the lectures." "I try to apply ideas from course readings in other class activities such as lecture and discussion."
Organization (O) (α = .64)	"When I study the readings for this course, I outline the material to help me organize my thoughts." "I make simple charts, diagrams, or tables to help me organize course material."
Critical Thinking (CT) (α = .80)	"When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence." "Whenever I read or hear an assertion or conclusion in this class, I think about possible alternatives."
Meta-Cognitive Self-Regulation (MSR) (α = .79)	"When I study for this class, I set goals for myself in order to direct my activities in each study period." "When studying for this course I try to determine which concepts I don't understand well." "I try to change the way I study in order to fit the course requirements and instructor's teaching style."
Time and Study Environment (TSM) (α = .76)	"I make good use of my study time for this course." "I have a regular place set aside for studying." "I attend class regularly."
Effort Regulation (ER) (α = .69)	"I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do." "Even when the course materials are dull and uninteresting, I manage to keep working until I finish."
Peer Learning (PL) (α = .76)	"When studying for this course, I often try to explain the material to a classmate or a friend." "I try to work with other students from this class to complete the course assignments."
Help Seeking (HS) (α = .52)	"I ask the instructor to clarify concepts I don't understand well." "I try to identify students in this class whom I can ask for help if necessary."

Purpose

Researchers are continually advancing understanding for how different constructs may relate to academic success; consequently, it is important to assess the accuracy and precision of available measurement tools used to establish frameworks and theories encompassing such research constructs. The purpose of the present study is to examine the psychometric properties of the MSLQ learning strategies scales. Although psychometric properties included in the manual for the MSLQ provide researchers point estimates of scale reliabilities, a reliability generalization study utilizes additional information to more broadly generalize the extent to which instruments may yield reliable scores and what factors contribute to variations in reliability coefficient estimates (Vacha-Haase, Henson & Caruso 2002). Therefore, a reliability generalization study of the nine learning strategies scales is used to explore the typical score reliabilities of each scale and to examine sources of measurement error variance across studies.

Method

Sample of Articles Using MSLQ Learning Strategies Scales

To locate articles using the MSLQ instrument, the Academic Premier, ERIC, PsychINFO, PsychARTICLES, Education Research Complete, and Google databases were searched using the instrument name, *Motivated Strategies for Learning Questionnaire*, and the acronym, *MSLQ*. Additional studies using the MSLQ instrument were located through studies retrieved. A total of 363 studies were identified. A number of studies were excluded because articles were either false hits, in non-English languages, or not obtainable. Researchers that did not collect original data, used Pintrich and DeGroot's (1990) Junior High version, or used a previous version of the instrument were also eliminated from the study. A remaining 225 studies were furthered reviewed.

Estimates of Score Reliability

Each of the 225 articles was examined and divided into four categories: a) articles that did not report reliability for participant scores or give any information about score reliability (n = 40, 17.8%), b) articles that used the MSLQ but reported score reliability from the MSLQ manual or other previous study (n = 62, 27.6%), c) articles that used the MSLQ and provided reliability information for participants in the study (n = 109, 48.4%), or d) articles that used the MSLQ and provided reliability information for participants in the study in such a way that the data could not be used {e.g., authors reported the range or referenced the reliability coefficients for participants in the study (n = 14, 6.2%)}. Of the 109 studies reporting reliabilities for their sample, 15 studies were excluded from the reliability generalization study either because data was replicated within another study, data was not reported at the scale level, or because researchers reported a reliability coefficient different from Cronbach's alpha.

The present reliability generalization study was based on 70 studies in which reliability estimates for at least one of the nine learning strategy scales on the MSLQ were reported. Since some of the articles reported reliability coefficients for multiple samples, the total number of reliability coefficients varied from a total of 53 coefficients for the Help Seeking scale to 79 coefficients for the Metacognitive Self-Regulation scale. All estimates of reliability included in the RG study of the MSLQ learning strategies scales were measures of internal consistency, Cronbach's α.

Coding of Study Characteristics

To assess sources of variability in reliability estimates, study- and sample-characteristics were coded from each study which included reliability coefficients for their sample. The sample characteristics were coded at the same level as the reported reliability estimates. If a study provided reliability coefficients for a subset of the sample, only characteristics reported at the subset level were included. Study characteristics coded from each study include a) type of article (coded as journal, dissertation or thesis, conference proceeding, book chapter, and unpublished report) and b) number of anchor items used (5 or 7). The following sample characteristics from each study were also recorded: a) sample size (recorded as a continuous variable representing total number of participants in the sample); b) age of participants (recorded as a continuous variable representing the mean age of sample participants); c) academic population (recorded as a nominal variable denoting elementary, middle school, junior high/high school, undergraduate, graduate and other populations); d) course (whether participants responded to survey items for a specific course or responded to items in a general sense); e) sex (recorded as the number and percentage of males and females within each sample); f) race (recorded as the number and percentage of Asian, Black, Hispanic, Native American, White, and participants of Other races in each sample); g) country of study (the country where the survey was administered); and h) translation (translation of the instrument administered to the sample).

Studies varied widely in the amount of information provided to describe sample characteristics of their research study; thus, some cases were collapsed into broader category groupings due to missing data and/or small sample sizes. Changes and descriptions for final variables used for analyses are reflected in Table 4.2. Information about the variable type, description of range reported, and the number of studies in which data was included for each variable is also presented in the table.

Table 4-2

Coding Strategies for Study- and Sample-Specific Predictor Variables

Variable	Coding Classification (range)	Number of Sample Cases
Journal	0 = Dissertation (49) or conference paper (4) 1 = Journal Article (42)	95
Likert-type scale	0 = 5 point Likert-type scale (14) 1 = 7 point Likert-scale (77)	91
N (Sample size)	Continuous (21 to 2,005; median = 161)	95
Mean age	Continuous (14.04 to 33; median = 21.4)	33
Academic population	0 = Elementary through secondary students (13) 1 = Postsecondary students (77)	90
Country	0 = studies conducted in other countries (28) 1 = studies conducted in the United States (66)	94
Translation	0 = Other language (18) 1 = English (69)	87
Course specificity	0 = Courses in general (8) 1 = Specific course (80)	88
Percentage of female participants	Continuous (0 to 100) - Calculated as the proportion of females to total sample size	83
Percentage of Asian participants	Continuous (0 to 100) - Calculated as the proportion of Asian participants to total sample size	37
Percentage of Black participants	Continuous (0 to 100) - Calculated as the proportion of Black participants to total sample size	35
Percentage of Hispanic participants	Continuous (0 to 100) - Calculated as the proportion of Hispanic participants to total sample size	33
Percentage of White participants	Continuous (0 to 100) - Calculated as the proportion of White participants to total sample size	41

Analyses

Henson and Thompson (2002) assert "RG [reliability generalization] is not conceived as a monolithic method, there are a variety of ways in which an RG study could be conducted and what variables could be considered in the analyses" (p. 124). Researchers have employed a number of descriptive and inferential statistical strategies to explore the variance within and between reliability coefficients for their study. Due to the pervasive amount of missing data with study characteristics recorded from studies using the MSLQ, an approach suggested by Lane, White, and Henson (2002) to use univariate analysis of variance (ANOVA) for nominal study characteristics and simple regression for continuous study characteristics was used to assess variations in reliability scores for the nine learning strategies scales. Lane et al., remark:

> We chose to run analyses for each predictor variable separately as against one analysis with simultaneous entry of the predictors. Although the latter option is preferred, after listwise deletion only a few cases remained and several predictors had to be omitted completely, resulting in considerable loss of data. Of course, this decision results in numerous analyses within the data set and inflation of experimentwise error. However, our focus was on the effects obtained from the analysis rather than on statistical significance of those effects. Accordingly, *p* values are not reported for the analyses. The preservation of data and resultant stability of the effect sizes obtained was critical to evaluating the relationships between study features and reliabilities across the studies. (pp. 691-692)

Similarly, results for the RG study of the MSLQ learning strategies scales are focused on effect sizes. Eta squared (η^2) and omega squared (ω^2) effects were used with nominal predictor variables. Coefficient of determination (R^2) and Adjusted R^2 effects were used with continuous predictor variables.

Reliability coefficients, similar to correlation coefficients, tend to be asymmetrically skewed. Hunter and Schmidt (1977) proposed that researchers conducting validity generalization studies with correlation coefficients should correct for departures in normality of the coefficients by transforming coefficients using Fisher's *z* transformation. However, Hunter and Schmidt (1990) reversed their position for using Fisher's *z* transformation claiming evidence suggests that the transformation is positively biased, i.e., larger weights are given to larger correlations than to smaller correlations. Still, several authors of RG studies have utilized this meta-analytic technique (e.g., Beretvas et al., 2008; Campbell, Pulos, Hogan & Murry, 2005; Graham, & Christianson 2009; Graham, Liu, & Jeziorski, 2006; Lopez-Pina, Sanchez-Meca & Rosa-Alcazar, 2009; Miller, Byrne, Rutherford & Hansen, 2009; O'rourke, 2004; Wallace & Wheeler 2002). Results of several RG studies that compare findings for using the *r*-to-*z* transformations and non-transformed alpha coefficients found negligible differences between findings and each approach (Caruso & Edwards 2001; Ryngala, Shields, & Caruso, 2005; and Shields and Caruso, 2004).

A Monte Carlo study conducted by Feldt and Charter (2006) compared results for averaging coefficients of internal consistency using six approaches including use of original data values, weighted averages, and Fisher's *z* transformation. Feldt and Charter concluded similar findings across all six approaches; therefore, it seems unnecessary to apply a transformation to the reliabilities of the learning strategies. Analyses conducted to typify and explore variations for reliabilities produced for the learning strategies scales are conducted using non-transformed alpha coefficients.

Results

Table 4.3 provides descriptive statistics for the typical magnitude and variability within reliability coefficients across studies in which researchers used the learning strategies scales of the MSLQ. A summary of scores is also portrayed in the box and whisker plots seen in Figure 4.1. Extremely low reliability values had been reported for peer learning (min = .03), effort regulation (min = .09), rehearsal, and critical thinking (min = .16 for both). In fact, outliers for reliability coefficients are evident in eight of the nine learning strategy scales, see Table 4.3. Due to the influential nature of outliers in distorting how different characteristics relate to variables of interest (Hair, Black, Babin, Anderson & Tatham, 2006), and since the causes of mild and extreme outliers can not be explored without access to original data sets, outlying cases exceeding three standard deviations from the mean were deleted from the dataset.

Table 4-3

Descriptive Statistics for Reliabilities from the MSLQ Learning Strategies Scales

Scale	*N*	Min	Lower Quartile	Median	Upper Quartile	Max	Mean	*SD*	Manual	Extreme Outliers
R	59	.16	.63	.69	.73	.84	.67	0.10	.69	1
E	65	.43	.72	.76	.80	.88	.75	0.08	.76	2
O	60	.32	.64	.69	.73	.86	.68	0.09	.64	2
CT	58	.16	.75	.80	.83	.91	.78	0.10	.80	1
MSR	79	.32	.74	.78	.82	.91	.77	0.09	.79	2
TSM	60	.46	.67	.75	.78	.85	.72	0.08	.76	1
ER	63	.09	.57	.65	.69	.82	.62	0.13	.69	1
PL	55	.03	.62	.70	.74	.87	.67	0.13	.76	1
HS	53	.41	.56	.60	.66	.82	.61	0.09	.52	0

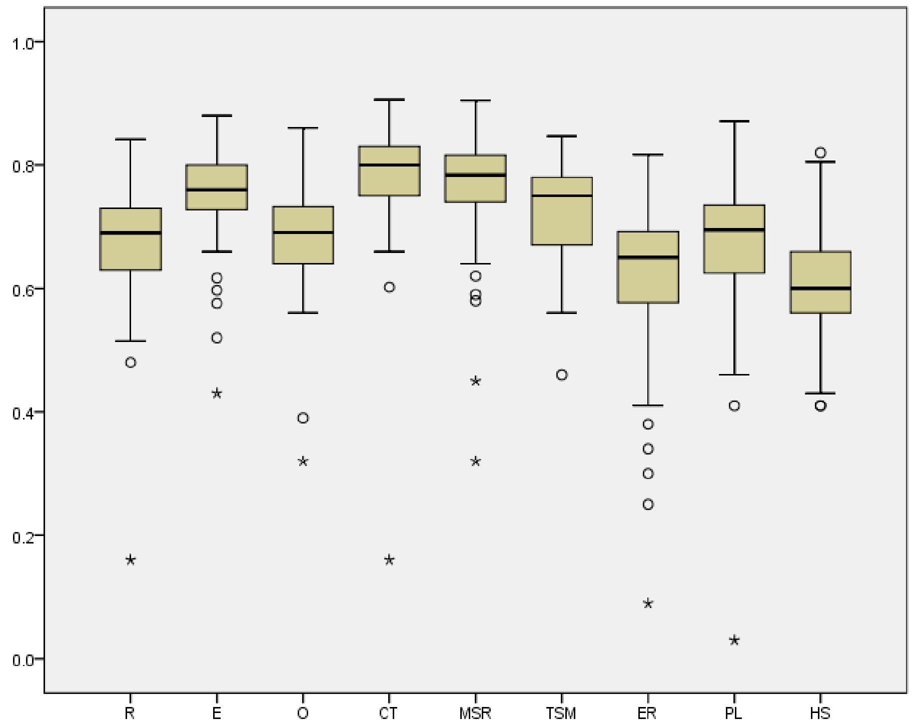

Figure 4-1. Box-and-whiskers plot of each MSLQ Learning Strategies scales

A descriptive summary of the reliabilities for each learning strategy scale after outliers were removed is provided in Table 4.4. In general, mean reliability coefficients of the MSLQ learning strategies scales were at or above .70. Only the effort regulation and help-seeking scales had average reliability coefficients closer to .60. Standard deviations of reliability coefficients ranged from 0.08 to 0.13. The reliabilities reported within the manual for each scale are also provided in Table 4.4. Overall, mean reliability scores found across studies were similar to reliability coefficients reported in the MSLQ manual. However, larger differences in reliabilities reported are evident for both the peer learning (α_{manual} = .76, α_{RG} = .68) and help-seeking (α_{manual} = .52, α_{RG} = .61) scales. Using an alpha level of .01, statistical comparisons using one-sample *t*-tests revealed statistically significant differences between the mean reliability coefficients found across studies using both the peer learning ($t(57) = -5.06$, $p < .001$) and help-seeking ($t(57)$ =

7.81, $p < .001$) scales and those reliability coefficients provided in the MSLQ manual. Hence, researchers might expect the typical reliability for using the peer learning scale to be slightly lower than the coefficient reported in the manual, and the typical reliability for the help seeking scale to be slightly higher than the coefficient reported in the manual.

Table 4-4

Descriptive Statistics for Reliabilities from the MSLQ Learning Strategies Scales after Omission of Outliers

Scale	*N*	Min	Lower Quartile	Median	Upper Quartile	Max	Mean	*SD*	Manual
R	58	.48	.63	.69	.73	.84	.68	0.07	.69
E	63	.58	.73	.76	.80	.88	.76	0.06	.76
O	58	.56	.64	.70	.73	.86	.70	0.07	.64
CT	57	.60	.76	.80	.83	.91	.79	0.06	.80
MSR	77	.58	.74	.79	.82	.91	.78	0.07	.79
TSM	59	.56	.67	.75	.78	.85	.73	0.08	.76
ER	62	.25	.58	.65	.70	.82	.62	0.12	.69
PL	54	.41	.63	.70	.74	.87	.68	0.09	.76
HS	53	.41	.56	.60	.66	.82	.61	0.09	.52

Visual representations of the distributions of reliabilities for the nine learning strategies scales are portrayed in Figures 4.2 to 4.10. Figures include inlaid histograms and normal probability plots. Normal probability plots compare the cumulative distribution of actual data values with the cumulative distribution of a normal probability plot. Departures from the straight diagonal line indicate departures from the normal distribution. In addition to visual representations, Shapiro-Wilk tests for normality were conducted for each scale. The distributions for five of the nine learning strategies scales do not appear to significantly depart from normal. Using an alpha level of .05, Shapiro-Wilk normality tests confirm that reliabilities for the critical thinking {$W(57) = 0.96$, $p = .05$}, metacognitive self-regulation {$W(77) = 0.96$, $p = .03$}, time and study management {$W(59) = 0.95$, $p = .01$}, and effort regulation {$W(62) = 0.90$, $p < .00$} scales are not normally distributed. Distributions for these scales are negatively skewed.

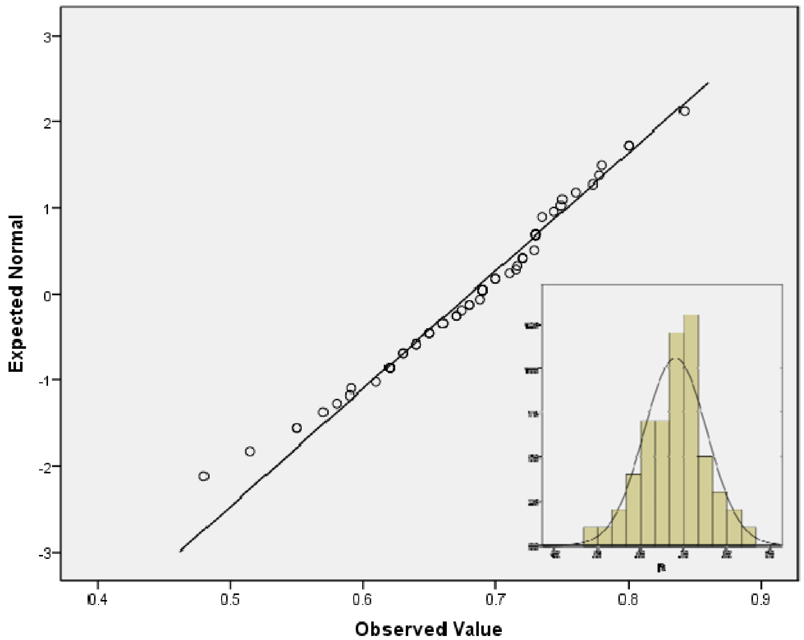

Figure 4-2. Normal probability plot and histogram for distribution of α for R.

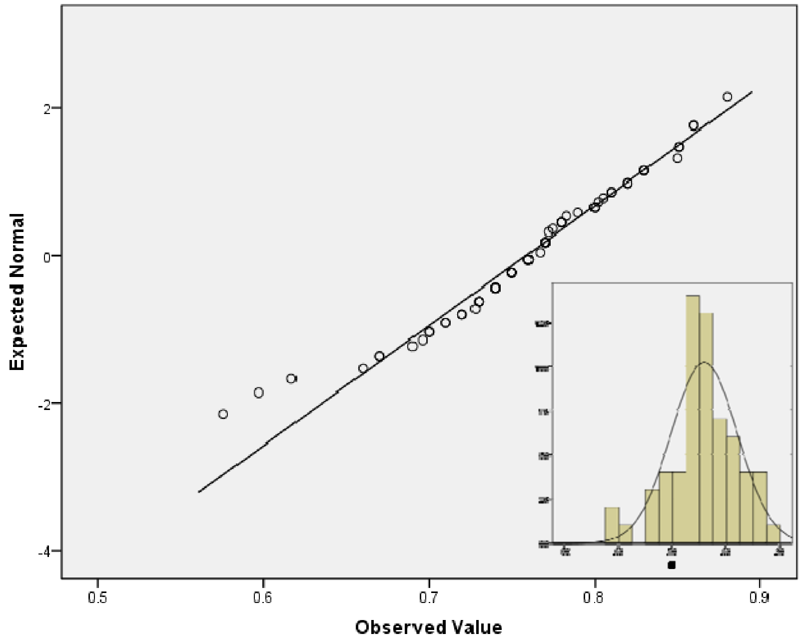

Figure 4-3. Normal probability plot and histogram for distribution of α for E.

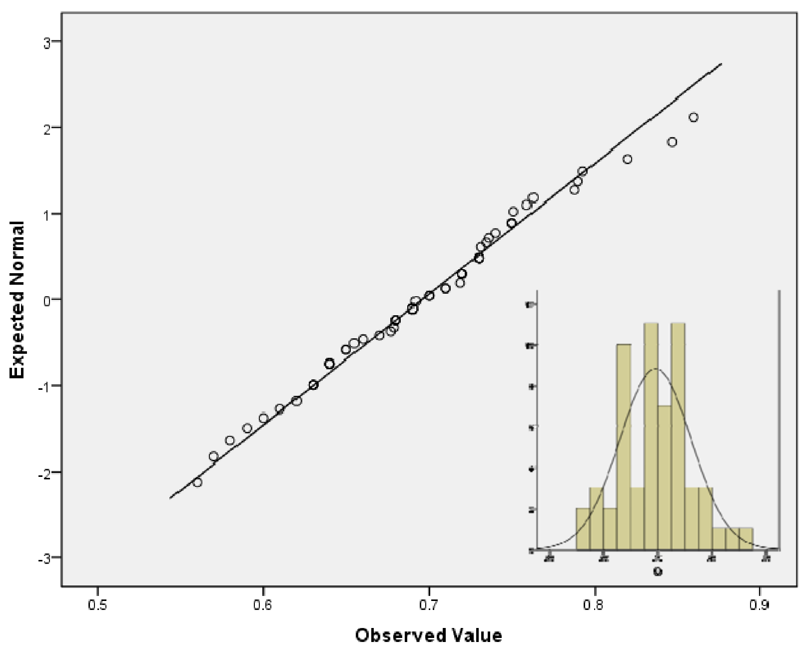

Figure 4-4. Normal probability plot and histogram for distribution of α for O.

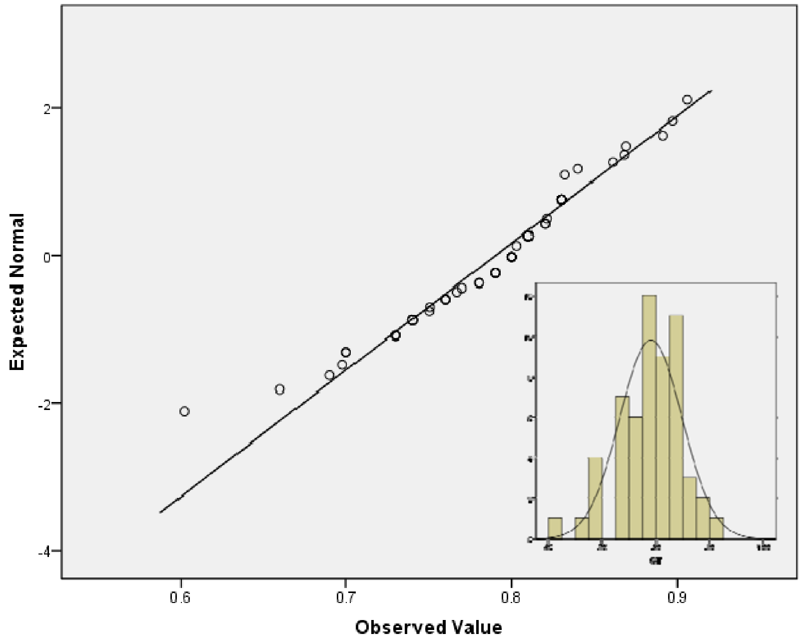

Figure 4-5. Normal probability plot and histogram for distribution of α for CT.

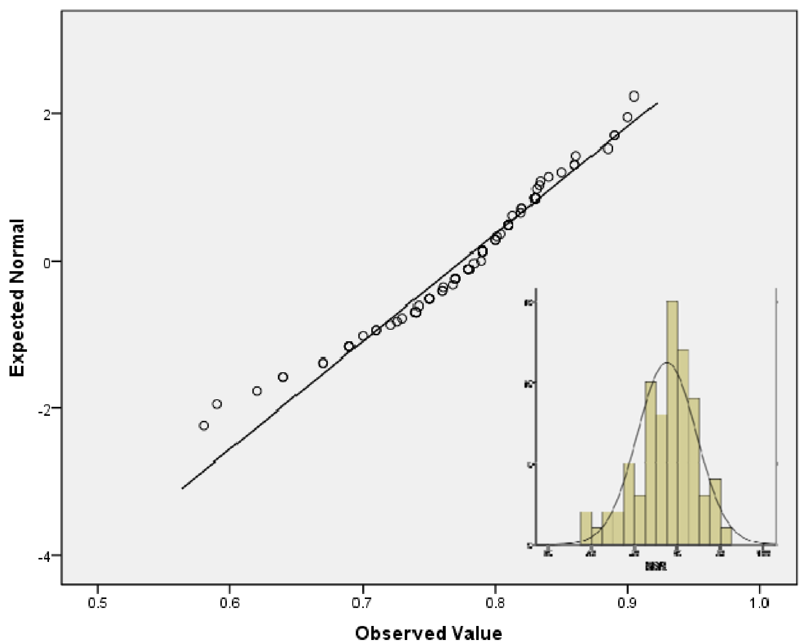

Figure 4-6. Normal probability plot and histogram for distribution of α for MSR.

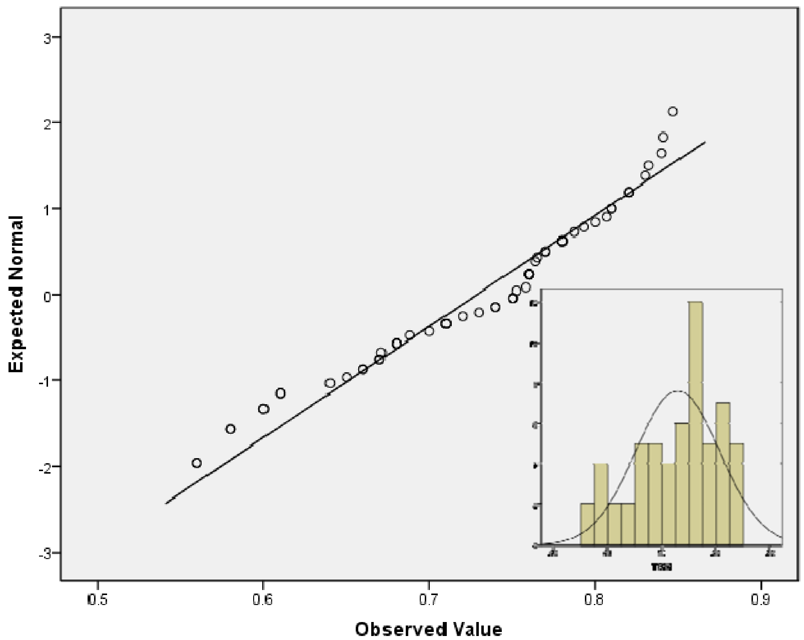

Figure 4-7. Normal probability plot and histogram for distribution of α for TSM.

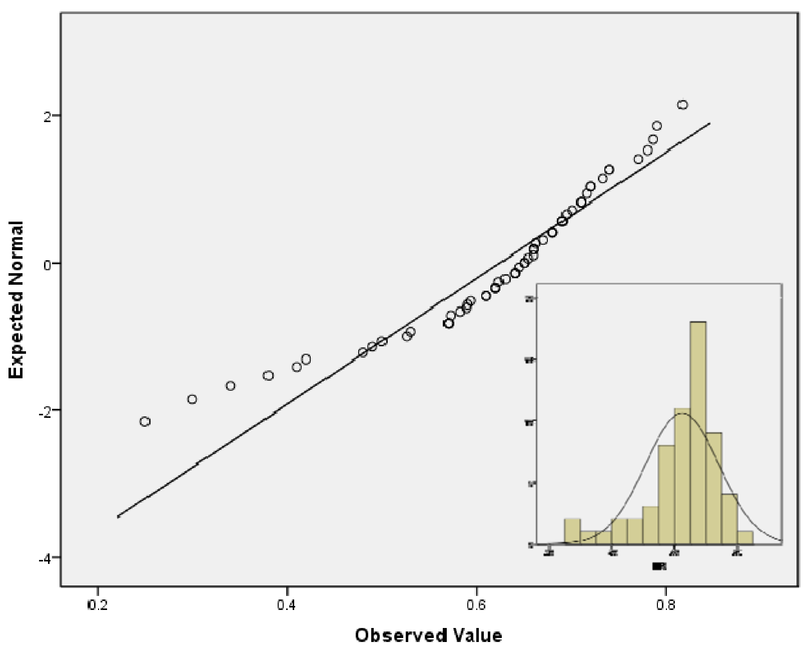

Figure 4-8. Normal probability plot and histogram for distribution of α for ER.

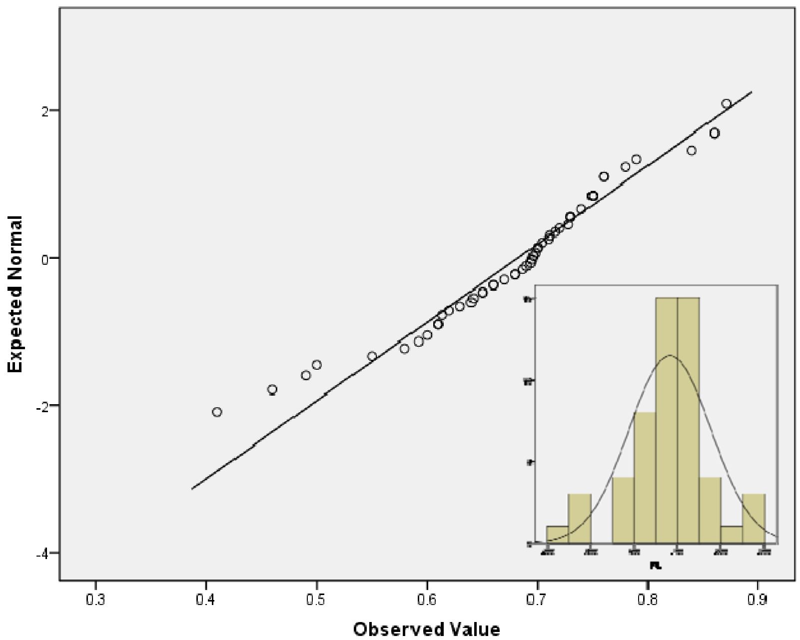

Figure 4-9. Normal probability plot and histogram for distribution of α for PL.

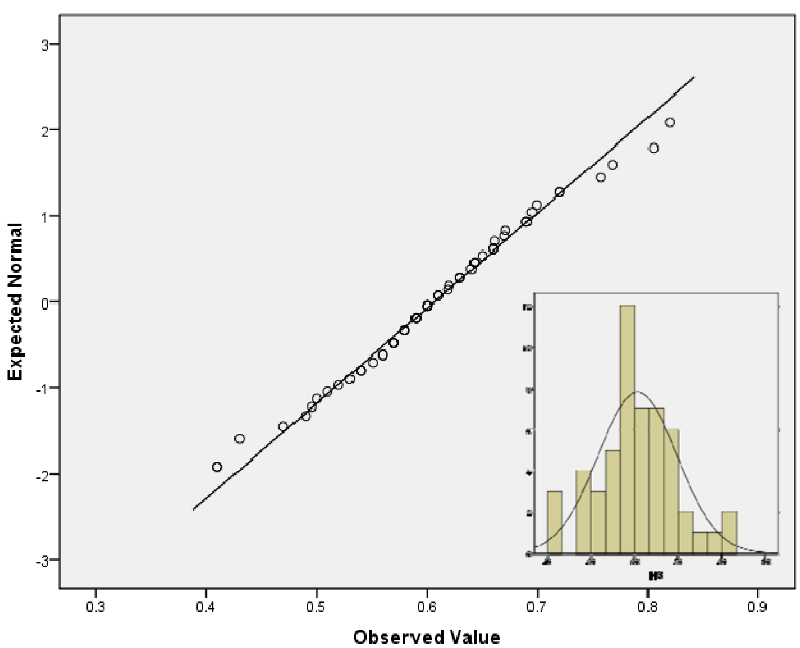

Figure 4-10. Normal probability plot and histogram for distribution of α for HS.

Effect sizes (both η^2 and ω^2) to denote the magnitude of effects of categorical predictors with alpha reliability coefficients for the nine learning strategies scales are provided in Table 4.5. Means and standard deviations by nominal categories for each of the nine learning strategies scales are included in Table 4.6. Differences between mean alphas are generally minute to negligible between categorical predictors and the five learning strategies scales representing the cognitive and metacognitive self-regulation section of the MSLQ. The rehearsal scale had several variables, article type (η^2 = .11), country (η^2 = .14), and translation of the instrument (η^2 = .11), account for slightly more than 10% of the variance in reliability scores. Alpha means for this scale were slightly higher for reliabilities reported in non-journal articles, samples within the United States, and for samples using the English translation of the instrument. Both study-characteristics (article type and number of Likert-type items used) accounted for slightly more than 10% of the variation in alpha scores for the critical thinking scale. For the critical thinking scale samples from non-journal articles and samples using 7-point anchor items tended to have more positive reliabilities. Country explains about 11% of the variance in reliabilities for the metacognitive self-regulation scale with samples from the United States reporting slightly higher alphas than samples from other countries.

Stronger effects are noticeable with three of the four scales included in the resource management component of the learning strategies section of the MSLQ. Translation of the scale, geographic region, and article type account for more than 20% of the variance in alpha scores for the time and study management scale (η^2 = .28, .26, and .20, respectively) and the peer learning scale (η^2 = .26, .31, and .22, respectively). Alpha means for both of these scales were more positive in non-journal articles, samples within the United States, and for samples using the English translation of the instrument. A similar trend is noticeable for the help seeking scale, although effect sizes were not as pronounced for country ($\eta^2 = .22$), translation ($\eta^2 = .12$), and article-type ($\eta^2 = .09$).

Table 4-5

Effect Sizes of Differences Between Score Reliabilities for Categorical Predictors of MSLQ Learning Strategies Scales

	R			E			O			CT			MSR		
Variable	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*
Article	0.11	0.09	53	0.02	0.00	58	0.02	0.00	54	0.11	0.09	52	0.07	0.05	72
Likert-Type Scale	0.01	0.00	56	0.05	0.04	61	0.02	0.01	56	0.12	0.10	55	0.00	-0.01	74
Academic Population	0.02	0.00	53	0.01	-0.01	58	0.01	-0.01	53	0.00	-0.01	52	0.00	-0.01	69
Course Specificity	0.01	-0.01	51	0.00	-0.02	55	0.01	-0.01	50	0.00	-0.02	50	0.00	-0.01	69
Geographic Region	0.14	0.12	56	0.00	-0.01	61	0.03	0.02	56	0.02	0.00	55	0.11	0.09	76
Translation	0.11	0.09	52	0.00	-0.02	57	0.00	-0.02	52	0.05	0.03	5	0.00	-0.01	69

Table 4-5 (*continued*)

Effect Sizes of Differences Between Score Reliabilities for Categorical Predictors of MSLQ Learning Strategies Scales

	TSM			ER			PL			HS		
Variable	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*	η2	ω2	*n*
Article	0.20	0.18	54	0.08	0.06	57	0.22	0.20	49	0.09	0.07	49
Likert-Type Scale	0.07	0.06	58	0.00	-0.01	60	0.12	0.10	52	0.00	-0.02	51
Academic Population	0.07	0.05	54	0.00	-0.02	57	0.00	-0.02	49	0.00	-0.02	48
Course Specificity	0.00	-0.02	52	0.02	0.00	55	0.00	-0.02	47	0.04	0.02	46
Geographic Region	0.26	0.24	57	0.07	0.06	60	0.31	0.29	52	0.22	0.21	52
Translation	0.28	0.26	53	0.02	0.00	56	0.26	0.24	48	0.12	0.09	48

Table 4-6

Descriptive Statistics for Score Reliabilities and Cell Frequencies for Categorical Predictors

	R			E			O			CT			MSR		
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Article															
0 = Other	.70	0.06	31	.77	0.07	34	.70	0.07	36	.81	0.05	32	.79	0.06	43
1 = Journal	.65	0.08	23	.75	0.06	25	.68	0.06	19	.77	0.06	21	.75	0.08	30
Likert-Type Scale															
0 - 5 point scale	.66	0.09	8	.72	0.07	8	.66	0.04	6	.73	0.09	5	.77	0.06	7
1 - 7 point scale	.69	0.07	49	.76	0.06	54	.70	0.07	51	.80	0.05	51	.78	0.07	68
Academic Population															
0 = Elementary - Secondary	.71	0.05	6	.78	0.07	7	.72	0.06	6	.78	0.06	7	.80	0.06	11
1 = Postsecondary	.68	0.07	48	.76	0.06	52	.70	0.06	48	.79	0.06	46	.78	0.07	61
Course Specificity															
0 = Courses in general	.66	0.06	3	.76	0.02	4	.68	0.07	3	.78	0.03	3	.77	0.02	5
1 = Specific course	.68	0.08	49	.76	0.06	52	.70	0.07	48	.79	0.06	48	.78	0.06	65
Geographic Region															
0 = Other	.64	0.08	16	.76	0.05	17	.67	0.07	13	.78	0.05	14	.74	0.08	21
1 = United States	.70	0.06	41	.76	0.06	45	.70	0.07	44	.80	0.05	42	.79	0.06	56
Translation															
0 = Other	.64	0.09	11	.76	0.06	12	.70	0.06	8	.77	0.06	8	.78	0.05	12
1 = English	.70	0.06	42	.76	0.06	46	.70	0.07	45	.80	0.05	44	.79	0.06	58

Table 4-6 (*continued*)

Descriptive Statistics for Score Reliabilities and Cell Frequencies for Categorical Predictors

	TSM			ER			PL			HS		
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*
Article												
0 = Other	.75	0.06	36	.65	0.09	35	.71	0.06	29	.63	0.09	31
1 = Journal	.68	0.08	19	.59	0.12	23	.63	0.10	21	.57	0.09	19
Likert-Type Scale												
0 - 5 point scale	.67	0.05	7	.61	0.16	8	.77	0.06	6	.59	0.02	3
1 - 7 point scale	.74	0.08	52	.63	0.11	53	.67	0.09	47	.61	0.09	49
Academic Population												
0 = Elementary - Secondary	.66	0.04	4	.65	0.07	6	.69	0.12	6	.62	0.04	4
1 = Postsecondary	.74	0.07	51	.63	0.11	52	.68	0.09	44	.61	0.10	45
Course Specificity												
0 = Courses in general	.72	0.11	3	.57	0.15	3	.68	0.16	4	.52	0.04	2
1 = Specific course	.74	0.07	50	.64	0.11	53	.69	0.09	44	.61	0.09	45
Geographic Region												
0 = Other	.66	0.08	14	.58	0.13	18	.60	0.11	14	.54	0.08	15
1 = United States	.75	0.06	44	.65	0.11	43	.71	0.07	39	.63	0.08	38
Translation												
0 = Other	.64	0.07	7	.61	0.11	12	.58	0.14	8	.54	0.09	9
1 = English	.75	0.06	47	.65	0.11	45	.71	0.07	41	.62	0.09	40

Beta coefficients (b), structure coefficients (r_s), coefficients of determination (R^2), and adjusted coefficients of determination (R^2) for simple regressions conducted between each continuous variable and the nine learning strategies scales are displayed in Table 4.7. The percentage of female participants had little effect on the variation in reliability scores for the nine learning strategies scales. Likewise, sample size tended to have no effect on reliabilities of the learning strategies scales, except for the critical thinking scale (R^2 = .12) in which increases in sample size tended to negatively relate with reliabilities for the scale. The number of cases that include participants' mean age and reliabilities tended to be low, thus interpreting effects for this continuous variable should be done cautiously. For example, only 11 cases were used to explore the relationship between mean age and alpha scores for the critical thinking scale. A scatterplot is used to inspect the relationship between these two variables, see Figure 4.11. Mean age explains 21% of the variance in alpha scores for critical thinking, but from the small number of data points, it appears undue leverage in some of the lower reliability coefficients is artificially magnifying the relationship between these variables.

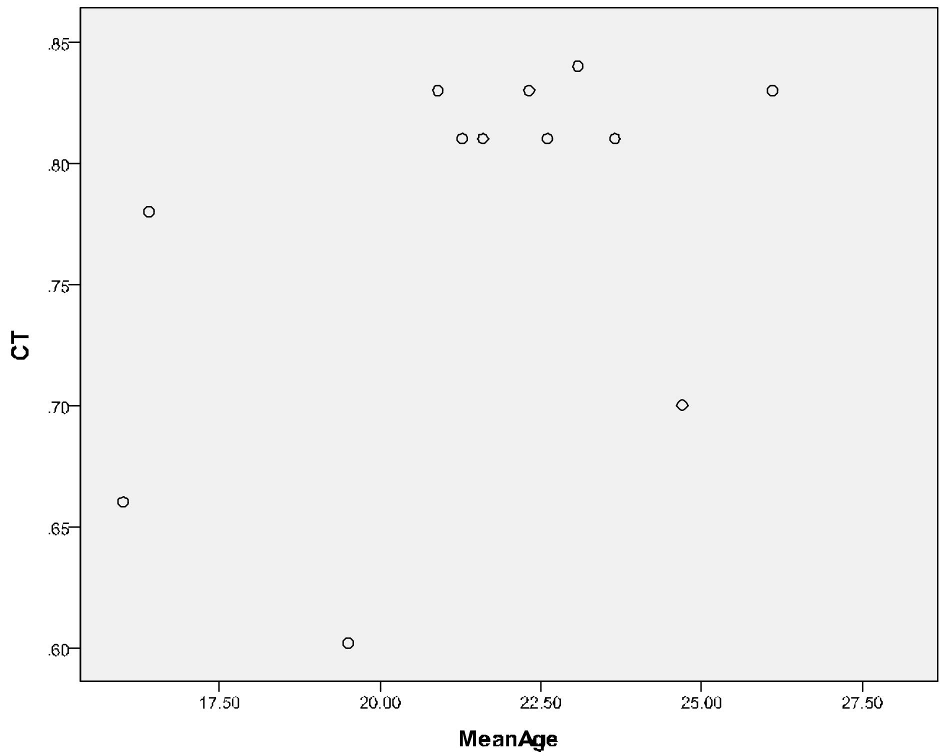

Figure 4-11. Scatterplot for mean age of participants with alpha scores for CT

The percentage of Asian participants moderately relates to reliabilities for the learning strategies scales, critical thinking (R^2 = .12, r_s = -.34), time and study management (R^2 = .38, r_s = -.61), effort regulation (R^2 = .21, r_s = -.46), peer learning (R^2 = .27, r_s = .52), and help seeking (R^2 = .10, r_s = -.32). The relationships between percentage of Asian participants and reliabilities were negative, except for peer learning. The percentage of Black participants moderately and positively relates to the rehearsal scale (R^2 = .21, r_s = .46). The percentage of Hispanic participants moderately relates to reliabilities for the learning strategies scales, elaboration (R^2 = .29, r_s = .54), metacognitive self-regulation (R^2 = .17, r_s = .41), time and study management (R^2 = .31, r_s = .56), peer learning (R^2 = .53, r_s = .73), and help seeking (R^2 = .10, r_s = -.32). The relationships between percentage of Hispanic participants and reliabilities were positive, except for help seeking. The percentage of White participants moderately and negatively relates to reliabilities for the rehearsal (R^2 = .28, r_s = -.53) and elaboration (R^2 = .14, r_s = -.38) scales and positively relates to reliabilities of the peer learning (R^2 = .18, r_s = .42) scale.

Table 4-7

Regression Statistics for Continuous Predictors of MSLQ Learning Strategies Scales

MSLQ Scale	Statistic	Sample Size	Mean Age	% Female	% Asian	% Black	% Hispanic	% White
	b	.000	.002	-.001	.000	.001	.001	-.001
	r_s	-.219	.070	-.226	-.120	.462	.167	-.532
R	R^2	.048	.005	.051	.014	.213	.028	.283
	Adj R^2	.031	-.086	.032	-.040	.174	-.029	.251
	n	57	12	51	19	21	18	23
	b	.000	.004	.000	.000	.001	.004	-.001
	r_s	-.074	.195	-.146	-.064	.276	.535	-.375
E	R^2	.005	.038	.021	.004	.076	.287	.141
	Adj R^2	-.011	-.036	.003	-.048	.032	.247	.103
	n	62	14	55	20	22	19	24
	b	.000	-.006	-.001	.000	.000	.003	-.001
	r_s	-.022	-.271	-.244	-.137	.151	.291	-.232
O	R^2	.000	.073	.059	.019	.023	.085	.054
	Adj R^2	-.017	-.019	.040	-.028	-.020	.039	.018
	n	57	11	50	22	24	21	27
	b	.000	.012	.000	-.001	.000	.001	.000
	r_s	-.339	.457	-.203	-.340	-.068	.156	.125
CT	R^2	.115	.209	.041	.116	.005	.024	.016
	Adj R^2	.099	.130	.022	.072	-.041	-.027	-.025
	n	56	11	50	21	23	20	25
	b	.000	.000	-.001	.000	.000	.003	-.001
	r_s	.156	.022	-.309	-.118	.056	.408	-.252
MSR	R^2	.024	.000	.095	.014	.003	.166	.063
	Adj R^2	.011	-.047	.081	-.021	-.034	.132	.033
	n	76	22	66	29	28	25	32

Table 4-7 (*continued*)

Regression Statistics for Continuous Predictors of MSLQ Learning Strategies Scales

MSLQ Scale	*Statistic*	Sample Size	Mean Age	% Female	% Asian	% Black	% Hispanic	% White
	b	.000	-.001	-.001	-.001	.000	.005	-.001
	r_s	-.179	-.052	-.239	-.618	.043	.555	-.230
TSM	R^2	.032	.003	.057	.382	.002	.308	.053
	Adj R^2	.015	-.074	.038	.357	-.037	.278	.019
	n	58	14	50	25	27	24	29
	b	.000	.001	-.001	-.001	-.001	.000	.000
	r_s	.039	.096	-.211	-.456	-.247	-.120	.064
ER	R^2	.002	.009	.045	.208	.061	.014	.004
	Adj R^2	-.015	-.067	.025	.172	.022	-.030	-.031
	n	61	14	51	23	25	23	29
	b	.000	.006	.000	-.001	.000	-.006	.001
	r_s	.062	.251	.119	-.516	-.084	-.725	.419
PL	R^2	.004	.063	.014	.266	.007	.525	.175
	Adj R^2	-.015	-.031	-.007	.233	-.036	.502	.141
	n	53	11	48	23	24	21	25
	b	.000	.002	.000	-.001	-.001	.003	.001
	r_s	-.140	.084	-.103	-.321	-.311	.218	.191
HS	R^2	.020	.007	.011	.103	.097	.048	.037
	Adj R^2	.000	-.103	-.011	.064	.057	-.003	-.002
	n	52	10	46	24	24	20	26

Discussion

Authors of empirical studies using the MSLQ reported coefficient alpha in just over half of the studies reviewed for analyses. Slightly less than a third of authors referenced reliabilities from the MSLQ manual or from another previous study, and roughly one-fifth failed to consider or mention reliability. For the 70 studies reporting reliability scores for MSLQ learning strategies scales, mean reliability estimates ranged from a low of .61 for the help seeking scale to a high of .79 for the critical thinking scale, with moderately low levels of variability. The average reliability levels for the nine learning strategies scales are generally considered acceptable for research purposes (Kelley, 1927; Nunnally & Bernstein, 1994; Thorndike, 1951). Mean scores for critical thinking, metacognitive self-regulation, and elaboration scales were more positive, while lower mean reliability scores were found for effort regulation and help seeking. Reliability scores for the peer learning scale were significantly lower than the $\alpha = .76$ reported in the manual, while reliability scores for the help seeking scale were significantly higher than the $\alpha = .52$ reported in the manual.

In addition, examinations between the relationships of several study-specific and sample-specific factors with variations in reliability scores were considered as part of a second aim for RG analyses. Study-specific characteristics for article type and number of anchor items used were examined to explore if differences were evident in reliability scores of the learning strategies scales. Small differences in reliability scores were found for seven of the nine learning strategies scales with higher reliabilities reported in non-journal articles. The use of 5-point Likert-type items versus 7-point Likert-type items did not seem to affect reliabilities for most of the learning strategies scales, although small effect sizes were evident for critical thinking and peer learning. For the critical thinking scale, higher mean reliabilities were found for samples using a 7-point Likert-type scale. Conversely, for the peer learning scale, higher mean reliabilities were found for samples using a 5-point Likert-type scale.

The MSLQ is designed to assess students' learning strategies use for a specific subject course and a large majority of researchers administering the survey asked their participants to respond to question items with respect to specific courses. Several researchers did appear to ask participants to provide 'general' appraisals of their study habits. Although, it was hypothesized that reliabilities would be stronger for subjects responding to scale items for a specific course, negligible differences are evident in reliability scores. The lack of any substantial finding could be attributed to the small number of cases in which participants provided general appraisals for their learning strategies use.

The MSLQ is intended to be used with college student populations, but several researchers have utilized the instrument with younger student populations. Score reliabilities for the younger students (elementary through secondary grade levels) were comparable to reliabilities of postsecondary students across all learning strategies scales. An examination of reliability scores with mean age proved inconclusive as the number of studies reporting ages for their samples was minimal. Thus, using results from the nominal academic population comparison, it seems acceptable to use the learning strategies section of the MSLQ with younger populations, although researchers should continue to consider the relevance of survey items with respect to their sample population and research interests.

Another consideration for conducting the RG study for learning strategies scales was to see if reliabilities would differ for sample groups not in the United States, which might suggest that constructs for different sample groups would need to be redefined for different cultures or that translations of the instrument might affect score reliabilities. For the RG study of learning strategies scales, country where the survey was administered was coded into a dichotomous variable to compare all studies conducted in the United States to studies conducted outside of the United States. Although, comparing reliability scores for each country would offer stronger evidence for where cultural differences may exist, sample sizes in many of the sample countries restricted analyses to this broader comparison. So for comparison of samples from the United States versus samples from other countries, small differences were found for the rehearsal and metacognitive self-regulation scales, while moderate differences were found for the peer learning, time and study management, and help seeking scales. Similar findings were also evident for translation of the instrument (except no differences were evident for the metacognitive self-regulation scale). Mean reliabilities were more positive for samples from the United States and for samples using an English translation of the instrument. This may suggest that not all scale items are appropriate for some cultures, or that translation of scale items need additional consideration.

Variables for the percentage of female participants and the percentage of different racial groups were also used to explore variations in reliability coefficients. Negative relationships were found between the percentage of Asian participants and reliability scores for critical thinking, time and study management, effort regulation, peer learning, and help-seeking. Positive relationships (although moderate) were found between the percentage of Black participants and reliability scores for rehearsal and elaboration scales. For the percentage of Hispanic participants, positive relationships were found among scores for elaboration, organization, metacognitive self-regulation, time and study management; and a negative relationship was evident with peer learning. A small negative effect was found for percentage of female participants' metacognitive self-regulation.

Limitations

A limitation to the reliability generalization study of learning strategies scales on the MSLQ is the lack of reliability estimates provided in studies using the instrument. In addition to the lack in reporting reliability estimates, the number of unpublished studies in which researchers may have used the instrument is unknown. Rosenthal (1979) suggests that statistically significant results are more likely to be published than studies demonstrating non-significant findings. This phenomena termed the 'file drawer problem' suggests that a number of additional studies omitted within the publication domain might actually depress result findings. The sample of studies available in which researchers used the MSLQ instrument may not be reflective of all studies that utilized the instrument, and thus findings may be more positive than would be true if additional 'file drawer' articles were available for review.

An additional limitation to the reliability generalization study was the underreporting of sample characteristics in studies that used the MSLQ instrument. This lack of reporting sample descriptors limited the statistical techniques that could be used to examine the relevant contribution of predictors in relation to other predictor variables. Instead, each predictor variable was examined with reliabilities scores from each scale, inflating experimentwise error rates. Results for statistical significance were disregarded, and only the magnitude and direction of effect sizes were used to explore findings.

Conclusion

The results of the reliability generalization study on the learning strategies scales of the MSLQ demonstrate moderate, but adequate reliabilities for all nine of the learning strategies scales. Mean reliability scores were between .60 and .80. Aspects of both study-specific and sample-specific characteristics did affect the measurement of different motivation scales. Specifically, differences were evident between reliability scores for five of the nine scales with country of study and between reliability scores for four of the nine scales for translation of the instrument. Negative findings were found between Asian learners and reliabilities for five of the nine scales suggesting that the learning strategies scales may not be appropriate across all cultures. In particular, learning strategy measures may need to be revised and tailored to fit different educational values and practices of the Asian culture.

Reliability generalization is an important meta-analytic technique to examine score reliability. Researchers should always consider the reliability of scores for their own samples as reliabilities are dependent on sample scores. Reliability generalization is an important tool that can be used to understand the expected reliabilities for a scale and also to understand sample characteristics that contribute to unreliability.

CHAPTER V. DISCUSSION

Validation is an ongoing process which is highly dependent on the nature and quality of accumulated evidence gathered for the construct under study (Pedhazur & Schmelkin, 1991). The purpose of this dissertation research was to review the validity of the Motivated Strategies for Learning Questionnaire (MSLQ) using Messick's concept of construct validity that incorporates content, substantive, structural, generalizability, external and consequential validity evidence (Messick, 1995). Although evidence for all six validity concepts is reviewed, the emphasis of this research was to further explore the reliability of MSLQ scales by conducting reliability generalization studies for motivation and learning strategies scales of the instrument.

The MSLQ is a self-report instrument designed to assess students' motivation orientations and use of learning strategies for college-aged students. The MSLQ was designed for post-secondary students, but has been used by researchers at the elementary, secondary, and post-secondary levels. The MSLQ is a situational specific instrument designed to be used for understanding students' motivation within a particular class of interest (Pintrich, Smith, Garcia, & McKeachie, 1991, 1993). The instrument is readily available within the public domain, economically feasible to administer, and can be easily scored. As such, the MSLQ has been used by many researchers for a plethora of research purposes since the instrument was developed in 1991.

The content aspect of validity can be determined by evaluating items used to measure a construct in light of the theoretical basis for the construct (Brualdi, 1999; Messick, 1995). The developers of the MSLQ outline the theoretical framework for the instrument in McKeachie, Pintrich, Lin, and Smith (1986), but also provide brief descriptions of each construct within the manual (Pintrich et al., 1991). The theoretical framework of the MSLQ integrates ideas from a general cognitive approach and a social-cognitive approach to learning.

Substantive validity refers to both the tasks providing appropriate sampling of domain processes and the accrual of empirical evidence of response consistencies (Messick, 1995). Developers of the instrument used 81 items to assess fifteen different motivation and learning strategies scales. The number of items representing each scale vary with three items used to represent the peer learning scale to twelve items used to represent metacognitive self-regulation. For test anxiety, developers only include items intended to measure student's worry. The emotionality component of test anxiety is not represented for this scale. Although metacognitive self-regulation includes three components to measure processes of metacognition (planning, monitoring, and regulating), developers combined the three processes into one overall measurement for metacognitive self-regulation.

The structural validity of the MSLQ was assessed using confirmatory factor analyses (Pintrich et al., 1991, 1993). Confirmatory factor analyses allow researchers to specify which items or indicators fall onto which factors or latent variables, and developers of the instrument claim fit indices indicate acceptable model fit. Validation studies completed by others have found some minor differences in the structural validity of MSLQ scores, but barring some minor revisions to the scales, these researchers tended to affirm the validity of interpretations based on the scale scores (Davenport, 2003; Hamilton & Akhter, 2009; Jacobson, 2000).

Evidence for the external validity of a construct is accrued by considering the empirical relationships between the measure and similar constructs or expected outcome variables

(Brualdi, 1999; Messick, 1995). The authors of the MSLQ provide correlations for each of the fifteen subscales with students' final grades in designated MSLQ courses. The authors claim the directions of correlations for the motivational scales were as expected and most of the relationships of the learning strategies scales with final grade were consistent with theory (Pintrich et al., 1991, 1993). Pintrich et al. (1993) do note that the negative relationship between peer learning with final grade and the negligible relationship of help seeking with final course grade were not expected.

In addition, the authors explored the correlational relationships among each of the fifteen constructs. Positive correlations were exhibited among intrinsic goal orientation, extrinsic goal orientation, task value, control of learning beliefs, and self-efficacy for learning and performance scales. Test anxiety negatively correlated with intrinsic goal orientation, task value, control of learning beliefs, and self-efficacy for learning and performance scales, but was positively related to external goal orientation. All subscales for the learning strategies component were positively related, with correlation coefficients ranging from .05 to .70.

Although a large number of studies have explored how different theoretical constructs of the MSLQ relate to one another, relate to other factors, and relate to outcome achievement variables like grades or retention, findings across such an abundant number of studies are varied. However, future research could include conducting a meta-analysis to analyze the strengths of the relationships between different variables measured by the MSLQ to gain a stronger theoretical perspective for the expected strength and magnitudes of such constructs. A meta-analytic study could lend further evidence to the external validity of the MSLQ.

Generalizability relates to the degree research findings can be inferred from a sample population to a more general population. Generalizability judgments should consider contextual, cultural, and other mediating factors that may influence subjects' responses. It remains unclear how appropriate the instrument would be for use with younger students. In addition, researchers have found differences between Asian cultures and Western cultures for the values ascribed to educational practices (Chen & Stevenson, 1995; Hau & Salili, 1991; Ho, 1994; Stigler, Smith & Mao, 1985; Whang & Hancock 1994). These differences could affect interpretation differences in scores and should be considered. Researchers who are trying to determine if the instrument is appropriate for their own research purposes should consider the theoretical framework and relevance across the context of their research purposes.

The consequential aspect of validity considers the value implications of score interpretations (Messick, 1995). The MSLQ has been used to collect data for numerous, diverse research purposes. Developers proposed that users of the instrument could provide students feedback for how their scores compare to class scores. The instrument has been used to explore the effectiveness of intervention programs designed to improve educational practices. Although users of the instrument may find general differences in groups of learners and their self-report of motivation and learning strategies use, these differences are more likely to be ascribed to contextual and cultural differences between groups of learners. Overall, it appears that consequential validity of the MSLQ appears acceptable, but should continue to be assessed according to the score interpretations of any researchers using the instrument.

In addition to reviewing the validity of the MSLQ, a major aim of the dissertation research was to explore the reliability of the MSLQ scales. Reliabilities reported in the MSLQ manual provide evidence for the consistency of scores for the sample in which the test or instrument was developed. However, further evidence for the reliabilities of each scale can be established using reliability generalization, a meta-analytic technique which explores the variation of reliability scores across different samples using the instrument (Vacha-Haase, 1998).

Reliability generalization is used to a) explore typical values of reliability coefficients for research employing a measurement instrument, b) explore the variation in reliability scores across samples, and c) investigate sample characteristics which help to explain variation in reliability scores across different research settings (Thompson, 1999; Vacha-Haase, 1998). The impetus of Vacha-Haase (1998) to introduce reliability generalization as a method for exploring the reliability of a measurement instrument was concern that poor reporting practices regarding reliability coefficients stemmed from "endemic confusion and misunderstanding of the concept and features of score reliability" (p. 6). Authors of several studies in which the psychometric reporting practices for published articles in select journals were reviewed, found that the majority of empirical studies failed to report psychometric data for their studies (Green, Chen, Helms & Henze, 2011; Meier & Davis, 1990; Vacha-Haase, Ness, Nilsson & Reetz, 19999; Willson, 1980).

Another common fallacy observed for reporting reliability of research is inducing the reliability of test scores. Vacha-Haase, Kogan, and Thompson (2000) coined the term reliability induction to "refer to the practice of explicitly referencing the reliability coefficients from prior reports as the sole warrant for presuming the score integrity of entirely new data" (p. 512). Graham, Liu and Jeziorski (2006) found that 42% of studies using the Dyadic Adjustment Scale either stated the instrument was shown to be reliable without providing data for their claim, or reported reliability coefficients from previous studies. Keiffer, Cronin, and Fister (2004) noted that roughly 25% of studies using the Alcohol Expectancy Questionnaire (AEQ) reported score reliability information from the original AEQ study or other prior studies.

Slightly more than half (123 of 225) of the authors who used the MSLQ for research purposes reported reliability scores for the participants in their studies. Just over a quarter (62 or 28%) of the authors referenced reliabilities from the manual or other previous study, and about 18% (40) of authors made no reference to reliability. Reliability is an important psychometric property of research as score unreliability affects statistical power (Henson, 2001), results of statistical significance, and attenuates effect sizes (Baugh, 2002; Hunter & Schmidt, 1994; Thompson & Snyder, 1998). Authors who failed to report reliabilities for their samples demonstrate a potential lack in understanding about reliability.

The majority of studies that reported reliability estimates for their research samples used Cronbach's alpha coefficient. Cronbach's alpha measures the consistency of an individual's performance across items on an instrument and is based on the variance in respondents' scores as well as the variances of each item (Cronbach, 1951). The distributions of alpha coefficients for nine of the fifteen MSLQ scales were approximately normal. Shapiro-Wilk normality tests were used to examine the distribution of alpha coefficients. At an alpha level of .05, task value, self-efficacy of learning and performance, critical thinking, metacognitive self-regulation, time and study management, and effort regulation scales demonstrate departures from normality. For each scale, the distributions of alpha were negatively skewed. Despite the departures from normality for these MSLQ variables, statistical analyses were conducted on the original reliability scores.

Descriptive statistics for reliabilities from the motivational and learning strategies scales of the MSLQ are displayed in Tables 3.4 and 4.4, respectively. Average reliability coefficients across the fifteen scales ranged from .61 to .88, levels that would be acceptable for most research purposes (Nunnally, 1967). Differences between average reliabilities found across studies and the reliabilities reported in the MSLQ manual were minor, except for the peer learning and help-seeking scales. For peer learning the manual reliability of .76 was higher than the mean of .68 found across 54 samples using the scale; conversely, the manual reliability of .52 for the help-seeking scale was lower than the mean of .61 found across 53 samples.

An additional aim of conducting reliability generalization studies for the motivational and learning strategies scales of the MSLQ is to investigate factors that might influence score reliabilities obtained when using the instrument. Previous reliability generalization studies have employed numerous statistical approaches to examine the relationships between study characteristics and reliability coefficients including Analysis of Variance (e.g., Caruso, 2000; Hellman, Muilenburg-Trevino & Worley, 2008), correlational analysis (e.g., Deditius-Island & Caruso, 2002; Henson, Kogan & Vacha-Haase, 2001; Shelds, Campfield, Miller, Howell, Wallace & Weiss, 2008); simple regression (e.g., Ross, Blackburn & Forbes, 2005), multiple regression (e.g., Bachner & O'rourke, 2007; Kieffer & MacDonald, 2011), and hierarchical regression (Vassar & Crosby, 2008). The decision for which statistical method to employ for analyzing variations in reliabilities of MSLQ scales was restricted as a large amount of sample characteristics were not provided across studies. Lane, White and Henson (2002) ran separate analyses for each predictor variable in their reliability generalization study of the Coopersmith Self-Esteem Inventory because of the pervasiveness of missing data across studies using the inventory. Since this approach inflates the experimentwise error rate in finding statistically significant results, Lane et al. focused solely on effect sizes and disregarded findings of statistical significance.

Univariate analysis of variance (ANOVA) for nominal study characteristics and simple regression for continuous study characteristics are used to explore the variability in reliability coefficients for both the motivational scales and learning strategies scales of the MSLQ.

Subsequently, η^2 and ω^2 effects were used with nominal predictor variables and R^2 and Adjusted R^2 effects were used with continuous predictor variables. For the variable, mean age of sample participants, a considerable majority of samples failed to provide means. Therefore, the present results are inconclusive concerning relationships between mean age and score reliabilities. A comparison of mean reliability scores for academic population across all fifteen MSLQ scales shows small differences between scores.

The relationships of two study-specific characteristics (type of article and number of Likert-type items) with score reliabilities were explored. Overall, the magnitude of effects for these characteristics with alphas for motivation and learning strategies scales are small to negligible. The strongest mean differences are evident for alphas found between article type for the time-and-study management (η^2 = .20) and peer learning (η^2 = .22) scales in which less positive alphas were reported in journal publications. For a small number of cases, participants were asked to provide general appraisals of their motivation and study habits. A comparison of scores for students generally appraising their beliefs to students who responded to the MSLQ with respect to a specific subject course revealed negligible differences between reliability scores. However, the lack in substantial findings could be due to the small number of cases in which participants provided general appraisals.

The MSLQ was developed with and intended to be used with college student populations, but several researchers have utilized the instrument with younger student populations. Score reliabilities for students in elementary through secondary grade levels were comparable to reliabilities of postsecondary students across all learning strategies scales, but minor differences were found between these two student populations and reliability scores for task value, self-efficacy and test anxiety scales. For each of these motivation scales, reliabilities were more positive for postsecondary samples. Mean age was also used to examine reliabilities between younger and older students. However, due to the insufficient number of cases providing ages for their samples, results are inconclusive across motivation and learning strategies scales.

Country where the survey was administered was coded into a dichotomous variable to determine if reliability scores differ for samples not from the United States. Moderate effects are evident for the test anxiety, time and study management, peer learning, and help seeking scales. Small differences were found for intrinsic goal orientation, task value, rehearsal, and metacognitive self-regulation scales. For these scales, reliabilities were generally lower for countries outside the United States. Similar findings were also evident for translation of the instrument (except no differences were evident for the metacognitive self-regulation scale), where mean reliabilities are higher for English translations of the instrument. This may suggest that not all scale items are appropriate for some cultures, or that translation of scale items need additional consideration.

Negative relationships were found between the percentage of Asian participants and reliability scores for task value (r_s = -.30), control of learning beliefs (r_s = -.38), self-efficacy (r_s = -.35) test anxiety (r_s = -.39), critical thinking (r_s = -.34), time-and-study management (r_s = -.62), effort regulation (r_s = -.46), peer learning (r_s = -.52), and help-seeking (r_s = -.32). Positive relationships (although moderate) were found between the percentage of Black participants and reliability scores for rehearsal (r_s = .46) and elaboration (r_s = .28) scales. For the percentage of Hispanic participants, positive relationships were found for extrinsic goal orientation (r_s = .35), elaboration (r_s = .54), organization (r_s = .29), metacognitive self-regulation (r_s = .41), time-and-study management (r_s = .56); and fairly strong, negative relationships were evident for self-efficacy (r_s = -.61) and peer learning (r_s = -.73). A small positive effect was found for percentage of female participants and test anxiety (r_s = .30) and negative effects for metacognitive self-regulation (r_s = -.31) and extrinsic goal orientation (r_s = -.27). Scale reliabilities tended to increase as the percentage of females in the sample increased, suggesting that these scales may be better suited to female populations.

Implications

Rowley (1976) states, "It needs to be established that an instrument is neither reliable or unreliable….A single instrument can produce scores which are reliable, and other scores which are unreliable" (p. 53). Dawis (1987) states, "Because reliability is a function of sample as well as of instrument, it should be evaluated on a sample from the intended target population – an obvious but sometimes overlooked point" (p. 486). More recently, Shields and Caruso (2003) note, "A particular test may produce scores that have a given reliability coefficient in one sample but a different coefficient in another. Consequently, the reliability coefficient provided in the test manual for a particular test is not to be taken as the reliability of the test; it cannot be disentangled from the sample in which it was derived" (p. 405). The reliability generalization studies of the MSLQ motivation and learning strategies scales further demonstrates how reliabilities vary by sample and are not immutable properties of test scores. The reporting practice for authors who used the MSLQ continues to demonstrate a potential lack in understanding of researchers about the necessity to consider reliability for their samples. Score unreliability affects statistical power (Henson, 2001), results of statistical significance, and attenuates effect sizes (Baugh, 2002; Hunter & Schmidt, 1994; Thompson & Snyder, 1998). Thus, researchers should always consider the reliability of scores for their samples.

Reliability generalization methodology can show factors that affect score reliabilities. With such a large number of researchers using the MSLQ to measure different motivation factors and learning strategies of learners, the reliability generalization studies provide further evidence of the typical reliabilities users may expect when using each motivation subscale. In addition, findings of the reliability generalization studies of the two MSLQ sections conclude that reliability coefficients do vary as a function of study-specific and sample characteristics.

Limitations

The results of the reliability generalization studies for both the motivation and learning strategies scales of the MSLQ should be considered in context of several limitations of the study. Although diligent attempts were made to retrieve all journal articles, conference proceedings, dissertations and theses, and other work in which researchers used the MSLQ instrument; a number of studies were not obtained for the study. In addition, it is likely that numbers of researchers have used the MSLQ instrument but have not published or made their results available within the public domain. Glass, McGaw and Smith (1981) claim that "the published stratum and unpublished stratum have opposite average effects, and a meta-analysis containing only published studies would be wholly unrepresentative of the population," (p. 64). Their assertion stems from Rosenthal's (1979) introduction of the 'file drawer problem' which postulates that only positive research in which authors conclude statistically significant results are published, while numerous research studies in which statistical significance is not concluded are 'filed away'. Thus, there is a plausible tendency that a large and valuable contribution of theoretical research is discarded or omitted due to negative findings. Implications for such a gap in available studies could be indicative of lower reliability scores which might impact overall reliability generalization findings.

Another limitation includes the failure of researchers to consider or include reliability data for the samples in their research. Only 54% of authors who used the MSLQ to collect data, reported reliability scores for their samples. Consequently, data was not available in almost half of the research articles in which authors used the instrument. Analyses were further limited by the lack of researchers to record sample descriptors. With large quantities of descriptors not included across studies, explorations of participant characteristics that might explain variability in reliability estimates across samples and studies is restricted. The prevalence in missing information limits the ability to generalize research findings.

Conclusions

Reliability generalization is an important meta-analytic technique to examine score reliabilities for a test or instrument. Overall, results of reliability generalization studies for both the motivation and learning strategies sections of the MSLQ demonstrate that the MSLQ can be used across a variety of different samples with reasonable confidence for obtaining generally reliable scores. However, aspects of both study-specific and sample-specific characteristics did affect the measurement of different motivation and learning strategies scales, highlighting the need for researchers to always consider the reliability of scores for their samples. Reliability generalization studies of both MSLQ sections suggest that reliability is not an immutable property of the scale, but differs across samples. This study provides further evidence for the need of social science researchers to always examine the reliabilities of the scores their own data.

REFERENCES

References marked with an asterisk (*) indicate studies included in the reliability generalization study of the motivation scales. References marked with a plus symbol (+) indicate studies included in the reliability generalization study of the learning strategies scales.

+Adcroft, A. (2010). The motivations to study of undergraduate students in management: The impact of degree programme and level of study. *International Journal of Management Education, 9*(1), 11-20.

Allen, A. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.

+Almegta, N. (1997). Relationship of self-efficacy, causal attribution, and emotions to female college students' academic self-evaluation (Doctoral dissertation, University of Arizona, 1996). *Dissertation Abstracts International, 58*, 192.

American Psychological Association. (1966). *Standards for educational and psychological tests and manuals.* Washington, DC: Author.

Ames, C. (1992). Classrooms, goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271.

Ames, C. & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260-267.

Ames, R. (1983). Help-seeking and achievement orientation, perspectives from attributional theory. In J. Fisher, A. Nadler, & B. DePaulo (Eds.), *New Directions in Helping* (pp. 165-186). New York: Academic Press.

Anastasi, A. (1954). *Psychological testing* (1st ed.). New York: Macmillian.

+Andreou, E. (2004). Bully/victim problems and their association with Machiavellianism and self-efficacy in Greek primary school children. *British Journal of Educational Psychology, 74*(2), 297-309.

+Andreou, E., & Metallidou, P. (2004).The relationship of academic and social cognition to behaviour in bullying situations among Greek primary school children. *Educational Psychology, 24*(1), 27-41.

Arend, B. (2009). Encouraging critical thinking in online threaded discussions. *Journal of Educators Online, 6*(1), 1-23.

+*Arsal, Z. (2010). The effects of diaries on self-regulation strategies of preservice science teachers. *International Journal of Environmental & Science Education, 5*(1), 85-103.

Atkinson, J. (1964). *An introduction to motivation*. New York: D. Van Nostrand.

Bachner, Y., & O'rourke, N. (2007). Reliability generalization of responses by care providers to the Zarit Burden Interview. *Aging and Mental Health, 11*(6), 678-685

Bandalos, D., Finney, S., & Geske, J. (2003). A model of statistics performance based on achievement goal theory. *Journal of Educational Psychology, 95*(3), 604-616.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1986). *Social foundations of thought and action: A scoial cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*(5), 729-735.

Bandura, A. (1994). Social cognitive theory and the exercise of control over HIV infection. In R. DiClemente & J. Pererson (Eds.), *Preventing AIDS: Theories and methods of behavioral interventions* (pp. 25-59). New York: Plenum.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1-26.

Barise, A. (2000). The effectiveness of case-based instruction vs. the lecture-discussion method in multicultural social work (Doctoral dissertation, McGill University, Canada, 2000). *Dissertation Abstracts International, 61*, 186.

+*Barker, J., & Olson, J. (1997) Medical students' learning strategies: Evaluation of first year changes. *Journal of the Mississippi Academy of Sciences, 42*(2), 96-100.

+Barlia, L. (1999). High school student's motivation to engage in conceptual change-learning in science (Doctoral dissertation, Ohio State University, 1999). *Dissertation Abstracts International, 60*(05), 232.

[+*]Bassili, J. (2008). Motivation and cognitive strategies in the choice to attend lectures or watch them online. *Journal of Distance Education, 22*(3), 129-148.

Baugh, F. (2002). Correcting effect sizes for score reliability: A reminder that measurement and substantive issues are linked inextricably. *Educational and Psychological Measurement, 62*(2), 254-263.

[+]Becker, D., & Dwyer, M. (1994). Using hypermedia to provide learner control. *Journal of Educational Multimedia and Hypermedia, 3*(2), 155-172.

[+*]Bembenutty, H., & Karabenick, S. (1997, March). *Academic delay of gratification in conditionally-admissible minority college students*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

[+*]Bembenutty, H., Karabenick, S., McKeachie, W., & Lin, Y. (1998, April). *Academic delay of gratification as a volitional strategy.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Benson, J. (1998). Review of the Motivated Strategies for Learning Questionnaire. In J. C. Impara & B. S. Plake (Eds), *The thirteenth mental measurements yearbook.* Buros Institute of Mental Measurements: Lincoln, NE.

Beretvas, S., Suizzo, M., Durham, J., & Yarnell, L. (2007). A reliability generalization study of scores on Rotter's and Nowicki-Strickland's Locus of Control Scales. *Educational and Psychological Measurement, 68*(1), 97-119.

[+*]Bhattacharyya, S. (2005). Dynamics of motivation and self-regulation among college pre-service teachers: A cross-cultural perspective (Doctoral dissertation, University of Memphis, 2004). *Dissertation Abstracts International, 65*, 115.

[*]Bidjerano, T. (2005, October). *Gender differences in self-regulated learning*. Paper presented at the Northeastern Educational Research Association, Kerhonkson, NY.

[+*]Birenbaum, M., & Rosenau, S. (2006). Assessment preferences, learning orientations, and learning strategies of pre-service and in-service teachers. *Journal of Education for Teaching, 32*(2), 213-225.

[+*]Black, K. (2009). Self-regulated learning as a predictor of persistence among first year students (Doctoral dissertation, University of Northern Colorado, 2009). *Dissertation Abstracts International, 70*, 122.

Blom, S., & Severiens, S. (2008). Engagement in self-regulated deep learning of successful immigrant and non-immigrant students in inner city schools. *European Journal of Psychology of Education, 23*(1), 41-58.

[+]Bogenmann, E. (2008). Motivational variables that influence the attendance of science trainees at research seminars (Doctoral dissertation, University of Southern California, 2008). *Dissertation Abstracts International, 69*, 144.

Bong, M. (1999, August). *Comparison between domain-, task-, and problem-specific academic self-efficacy judgments: Their generality and predictive utility for immediate and delayed academic performances.* Paper presented at the annual meeting of the American Psychological Association, Boston, MA.

Bredo, E. (1997). The social construction of learning. In G. Phye (Ed.), *Handbook of academic learning* (pp. 3-45). San Diego, CA: Acadmemic Press, Inc.

Brennan, R. (2001, Winter). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement, 38*(4), 295-317.

+Brookhart, S., & Durkin, D. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*(1), 27-54.

Brualdi, A. (1999). *Traditional and modern concepts of validity*. Washington DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED99CO0032)

+Cafarelli, A. (1999). Effects of the implementation of stage one of the Ursuline studies program on traditional-age female freshmen students' perceptions of self-esteem, self-efficacy, expectations for future success in college, problem-solving and critical thinking skills (Doctoral dissertation, Cleveland State University, 1998). *Dissertation Abstracts International, 60*, 378.

Campbell, J., Pulos, S., Hogan, M., & Murry, F. (2005). Reliability generalization of the Psychopathy Checklist Applied in Youthful Samples. *Educational and Psychological Measurement, 65*(4), 639-656.

+*Campbell, M. (2001). Motivational strategies, learning strategies and the academic performance of African-American students in a college business environment: A correlational study (Doctoral dissertation, Nova Southeastern University, 2001) *Dissertation Abstracts International, 62*, 167.

Carroll, C., & Garavalia, L. (2004). Factors contributing to the academic achievement of pharmacy students: Use of the goal-efficacy framework. *American Journal of Pharmaceutical Education, 68*(4), 1-8.

Caruso, J. (2000). Reliability generalization of the NEO Personality Scales. *Educational and Psychological Measurement, 60*(2), 236-254.

Caruso, J., & Edwards, S. (2001). Reliability generalization of the Junior Eysenck Personality Questionnaire. *Personality and Individual Differences, 31*, 173-184.

*Chalupa, M., Chen, C., & Charles, T. (2001). An analysis of college students' motivation and learning strategies in computer courses: A cognitive view. *The Delta Pi Epsilon Journal, 43*(4), 185-199.

+*Chan, J. (1999). Predictors of achievement using computer-assisted instruction: Self-efficacy for achievement and control of learning beliefs (Doctoral dissertation, University of Houston, 1999). *Dissertation Abstracts International, 60,* 112.

*Chang, W. (2009). The relationship of self-regulation and academic achievement in college students with and without attention deficit/hyperactivity disorder: A brain-behavior perspective (Doctoral dissertation, Colorado State University, 2009). *Dissertation Abstracts International, 69*, 218.

+*Cheang, K. (2009). Effect of learner-centered teaching on motivation and learning strategies in a third-year pharmacotherapy course. *American Journal of Pharmaceutical Education, 73*(3), 1-8.

*Chen, C. (2002). Self-regulated learning strategies and achievement in an introduction to information systems course. *Information Technology, Learning, and Performance Journal, 20*(1), 11-23.

Chen, C., & Stevenson, H. (1995). Motivation and mathematics achievement: A comparative study of Asian-American, Caucasian-American, and East-Asian high school students. *Child Development, 66*, 1215-1234.

*Chen, G, Gully, S., & Eden, D. (2004). General self-efficacy and self-esteem: Toward theoretical and empirical distinction between correlated self-evaluations. *Journal of Organizational Behavior, 25*, 375-395.

+Clark, A., Ernst, J., & Scales, A. (2009, June). *Results of a study using the Motivation Strategies for Learning Questionnaire (MSLQ) in an introductory engineering graphics course.* Paper presented at the annual meeting of the American Society for Engineering Education Southeast Section Conference, Austin, TX.

Clayton, T. (1965). *Teaching and Learning: A Psychological Perspective*. Prentice Hall, Inc. Englewood Cliffs, N.J.

Cobb, R., Jr. (2007). The relationship between self-regulated learning behaviors and academic performance in web-based courses (Doctoral dissertation, Virginia Polytechnic Institute and State University, 2003) *Dissertation Abstracts International, 67*, 146.

+Cole, J., & Denzine, G. (2004). "I'm not doing as well in this class as I'd like to": Exploring achievement motivation and personality. *College Student Motivation and Personality, 34*(2), 29-44.

Corno, L. (1994) Student volition and education: Outcomes, influences, and practices. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulation of learning and performance* (pp. 229-254). Hillsdale, NJ: Lawrence Erlbaum Associates.

+Covalt, K. (1997). The effects of guided imagery on college students' mathematics anxiety, mathematics self-efficacy, and mathematics performance (Doctoral dissertation, Northern Illinois University, 1996). *Dissertation Abstracts International, 57*, 181.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, and Winston, Inc.

Cronbach, L. (1949). *Essentials of psychological testing* (1st ed.). New York: Harper.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cronbach, L. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper.

+Cuevas, H. (2005). Transforming learning into a constructive cognitive and metacognitive activity: Use of a guided learner-generated instructional strategy within computer-based training (Doctoral dissertation, University of Central Florida, 2004). *Dissertation Abstracts International, 66*, 121.

+*Daniels, B. (2009). Motivation academic success and learning environments: Comparing high school face-to-face and online courses (Doctoral dissertation, George Mason University, 2009). *Dissertation Abstracts International, 69,* 127.

+*Davenport, M. (2003). Modeling motivation and learning strategy use in the classroom: An assessment of the factorial, structural, and predictive validity of the Motivated Strategies for Learning Questionnaire (Doctoral Dissertation, Auburn University, 1999). *Dissertation Abstracts International, 64*, 273.

Dawis, R. (1987). Scale construction. *Journal of Counseling Psychology, 34*(4), 481-489.

de la Harpe, B., & Radloff, A. (2000). Informed teachers and learners: The importance of assessing the characteristics needed for lifelong learning. *Studies in Continuing education, 22*(2), 169-182.

Deditius-Island, H., & Caruso, J. (2002). An examination of the reliability of scores from Zuckerman's Sensation Seeking Scales, Form V. *Educational and Psychological Measurement, 62*(4), 728-734.

Ditcher, A. (2001). Effective teaching and learning in higher education, with particular reference to the undergraduate education of professional engineers. *International Journal of Engineering Education, 17*(1), 24-29.

+*Doljanac, R. (1995). Using motivational factors and learning strategies to predict academic success (Doctoral dissertation, University of Michigan, 1994). *Dissertation Abstracts International, 56*, 96.

+Duijnhouwer, H., Prins, F., & Stokking, K. (2010). Progress feedback effects on students' writing mastery goal, self-efficacy beliefs, and performance. *Educational Research & Evaluation, 16*(1), 53-74.

Duncan, T., & McKeachie, W. (2005). The making of the Motivated Strategies for Learning Questionnaire. *Educational Psychologist, 40*(2), 117-128.

Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048.

Dweck, C. (1999). *Self-theories: The role in motivation, personality, and development*. Philadelphia: Taylor & Francis.

Dweck, C., & Leggett, E. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

[+]Ebrahim, S. (1998). The effects of locus of control, working memory, motivation, and verbal ability on foreign language learning (Doctoral dissertation, Pennsylvania State University, 1998). *Dissertation Abstracts International, 59*, 121.

Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives* (pp. 75-146). San Francisco: W. H. Freeman.

Eccles, J., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In N. Eisenberg (Ed.), W. Damon (Series Ed.) *Handbook of child psychology, Vol 3. Social, emotional, and personality development* (5th ed., pp. 1017-1095). New York: Wiley.

[+*]Eide, B. (1999). Integrating learning strategy instruction in an intermediate accounting course: Changes in accounting majors' learning strategy use and course performance (Doctoral dissertation, Virginia Commonweatlh University, 1998). *Dissertation Abstracts International, 59*, 159.

[+*]Eide, B., Schwartz, B., & Winter, K. (2004). Introducing and reinforcing learning strategies with accounting students. *Journal of Accounting and Finance, 12*(3), 60-69.

Elliot, A., & Church, M. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72*(1), 218-232.

Elliot, A., & Harackiewicz, J. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*(3), 461-475.

Eom, W., & Reiser, R. (2000). The effects of self-regulation and instructional control of performance and motivation in computer-based instruction. *International Journal of Instructional Media, 27*(3), 247-260.

Exner, S. (2010). An examination of the relationship among learning disability, attention deficit hyperactivity disorder, academic self-efficacy, effort, self-awareness and academic achievement in postsecondary students (Doctoral dissertation, Alfred University, 2010). *Dissertation Abstracts International, 71*, 98.

[+*]Feldon, D., Timmerman, B., & Stowe, K. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of Research in Science Teaching, 47*(10), 1165-1185.

Feldt, L., & Brennan, R. (1989). Reliability. In R.L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Feldt, L., & Charter, R. (2006). Averaging internal consistency reliability coefficients. *Educational and Psychological Measurement, 66*(2), 215-227.

Fergusson, J. (2004). A regression analysis of problem-based learning student variables (Doctoral dissertation, University of Nebraska Lincoln, 2003). Dissertation *Abstracts International, 64*, 173.

+*Fisher, J. (1998). Effects of supplemental instruction on undergraduate academic achievement, motivational orientations, and learning strategies (Doctoral dissertation, Auburn University, 1997). *Dissertation Abstracts International, 58*, 118.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-development inquiry. *American Psychologist, 34*(10), 906-911.

Flavell, J. (1992). Cognitive development: Past, present, and future. *Developmental Psychology, 28*(6), 998-1005.

Ford, M. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage Publications, Inc.

+*Foust, R. (2008). Learning strategies, motivation, and self-reported academic outcomes of students enrolled in web-based coursework (Doctoral dissertation, Wayne State University, 2008). *Dissertation Abstracts International, 69*, 861.

*Freedman, M. (1996). Self-efficacy and learning strategy use in a computer-based instructional setting (Doctoral dissertation, University of Southern California, 1995). *Dissertation Abstracts International, 57*, 105.

+Freeman, K., Alston, S., & Winborne, D. (2008). Do learning communities enhance the quality of students' learning motivation in STEM? *Journal of Negro Education, 77*(3), 227-240.

+Friedel, C. (2007). Dissimilar cognitive styles and their relationships with undergraduate student stress, motivation, and engagement (Doctoral dissertation, University of Florida, 2006). *Dissertation Abstracts International, 68*, 344.

*Fuller, R. (1999, July). *Do university students' conceptions of learning really influence their learning?* Paper presented at the HERSDA Annual International Conference, Melbourne: Australia.

Gagne, R. (1977). *The conditions of learning*. New York: Holt, Rinehart, and Winston, Inc.

Garcia, T., & Pintrich, P. (1996). Assessing students' motivation and learning strategies in the classroom context: The Motivated Strategies for Learning Questionnaire. In M. Birenbaum & F.Dochy (Eds.), *Alternatives in assessment of achievements, learning* *processes and prior knowledge* (pp. 319-339). New York: Kluwer Academic/Plenum Publishers.

Garcia-Ros, R., Perez, F., & Talaya, I. (2008). New university students' instructional preferences and how these relate to learning styles and motivational strategies. *Electronic Journal of Research in Educational Psychology, 6*(3), 547-570.

[+*]Garner, J. (2009). Conceptualizing the relations between executive functions and self-regulated learning. *Journal of Psychology, 143*(4), 405-426.

[+*]Gilles, L. (1994). *The assessment of student study skills and learning strategies to prepare teachers for academic advising tasks or the prevalidation of motivated strategies for learning questionnaire (MSLQ), learning and study strategies inventory (LASSI), and test of reactions and adaptation to college (TRAC).* (ERIC Document Reproduction Service No. ED372035)

Glass, G., McGaw, B., & Smith, M. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage Publications, Inc.

Graham, J., & Christiansen, K. (2009). The reliability of romantic love: A reliabilty generalization meta-analysis. *Personal Relationships, 16*, 49-66.

Graham, J., Liu, Y., & Jeziorski, J. (2006). The Dyadic Adjustment Scale: A reliability generalization meta-analysis. *Journal of Marriage and Family, 68*, 707-717.

Gredler, M. (Ed.). (2005). *Learning and instruction: Theory into practice* (5th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

[+]Green, A. (2001). The perceived motivation for academic achievement among African American college students (Doctoral Dissertation, Florida State University, 2000). *Dissertation Abstracts International, 61*, 134.

Green, C., Chen, C., Helms, J., & Henze, K. (2011). Recent reliability reporting practices in Psychological Assessment: Recognizng the people behind the data. *Psychological Assessment, 23*(3), 656-669.

Guilford, J. (1954). *Psychometric methods*. New York: McGraw-Hill.

Guvenc. H. (2010). The effects of cooperative learning journals on teacher candidates' self-regulated learning. *Educational Sciences: Theory & Practice, 10*(3), 1459-1487.

Hair, J., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

[+]Hamilton, R., & Akhter, S. (2009). Construct validity of the Motivated Strategies for Learning Questionnaire. *Psychological Reports, 104*(3), 711-722.

Hammann, D., Berthelot, J., Saia, J., & Crowley, E. (2000). Teachers' coaching of learning and its relation to students' strategic learning. *Journal of Educational Psychology, 92*(2), 342-348.

Hancock, D. (2002). Influencing graduate students' classroom achievement, homework habits and motivation to learn with verbal praise. *Educational Research, 44*(1), 83-95.

Harackievicz, J., & Barron, K., Pintrich, P., Elliot, A., & Thrash, T. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*, 638-645.

+*Harackiewicz, J., Barron, K., Tauer, J., Carter, S., & Elliot, A. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology, 92*(2), 316-330.

+*Harlow, J. (2006). Social integration, motivational orientation, and self-regulated learning strategies of online versus face-to-face theological seminary Biblical language students (Doctoral dissertation, University of North Carolina at Greensboro, 2006). *Dissertation Abstracts International, 67*, 135.

+*Harris, D. (1998). Effects of metacognitive skill training upon academic performance in accounting (Doctoral Dissertation, University of Texas Austin, 1998). *Dissertation Abstracts International, 59*, 155.

+*Harris, S., Edmundson, L., & Jacobson, R. (2006, November). *Motivational and learning strategies of community college students.* Paper presented at the meeting of the MidSouth Educational Research Association, Birmingham, AL.

Hau, K., & Salili, F. (1991). Structure and semantic differential placement of specific causes: Academic and causal attributions by Chinese students in Hong Kong. International *Journal of Psychology, 26*(2), 175-183.

Hellman, C., Muilenburg-Trevino, E., & Worley, J. (2008). The belief in a just world: An examination of reliability estimates across three measures. *Journal of Personality Assessment, 90*(4), 399-401.

Henson, R. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177-189.

Henson, R., & Hwang, D. (2002). Variability and prediction of measurement error in Kolb's Learning Style Inventory scores: A reliability generalization study. *Educational and Psychological Measurement, 62*(4), 712-727.

Henson, R., Kogan, L., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher Efficacy Scale and related instruments. *Educational and Psychological Measurement, 61*(3), 404-420.

Henson, R., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting "reliability generalization" studies. *Measurement and Evaluation in Counseling and Development, 35*, 113-126.

[+*]Higgins, B. (2000). *An analysis of the effects of integrated instruction of metacognitive and study skills upon the self-efficacy and achievement of male and female students.* Unpublished master's thesis, Miami University, Oxford, OH.

[+*]Highley, D. (1994). *The effects of a 'learning to learn' course on at risk students' motivation, self-regulated learning processes and academic achievement*. Unpublished doctoral dissertation, University of Southern California.

Ho, D. (1994). Cognitive socialization in Confucian heritage cultures. In P. Greenfield & R. Cocking (Eds.), *The development of the minority child: Culture in and out of context* (pp. 285-313). Hillsdale, NJ: Erlbaum.

[+*]Hofer, B., & Yu, S. (2003). Teaching self-regulated learning through a "Learning to Learn" course. *Teaching of Psychology, 30*(1), 30-33.

[*]Hollander, B. (2010). Religious affiliation and self-regulated learning strategies among modern orthodox Jewish female adolescents (Doctoral dissertation, Fordham University, 2009). *Dissertation Abstracts International, 71*, 80.

Hong, E., & Aqui, Y. (2004). Cognitive and motivational characteristics of adolescents gifted in mathematics: Comparisons among students with different types of giftedness. *Gifted Child Quarterly, 48*(3), 191-201.

[+]Hopstock, L. (2008). Motivation and adult learning: A survey among hospital personnel attending a CPR course. *Resuscitation, 76*(3), 425-430.

[+]Howey, S. (1999). The relationship between motivation and academic success of community college freshmen orientation students (Doctoral dissertation, Kansas State University, 1999). *Doctoral Dissertation Abstracts, 60*, 107.

[+*]Hsu, J. (1998). Value, expectancy, metacognition, resource management, and academic achievement: A structural model of self-regulated learning in a distance education context (Doctoral dissertation, University of Southern California, 1997). *Dissertation Abstracts International, 59*, 115.

[+*]Huang, S.(2008). Assessing motivation and learning strategies using the Motivated Strategies for Learning Questionnaire in a foreign language learning context. *Social Behavior & Personality, 36*(4), 529-534.

Hunter, J., & Schmidt, F. (1977). A critical analysis of the statistical and ethical implications of various definitions of test fairness. *Psychological Bulletin, 83*, 1053–1071.

Hunter, J., & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: SAGE Publications.

Hunter, J., & Schmidt, F. (1994). Correcting for sources of artificial variation across studies. In H. Cooper & L. Hedge (Eds.), *The Handbook of Research Synthesis* (pp. 323-336). New York: Russell Sage Foundation.

+Husman, J., Derryberry, W., Crowson, H., & Lomax, R. (2004). Instrumentality, task value, and intrinsic motivation: Making sense of their independent interdependence. *Contemporary Educational Psychology, 29*(1), 63-76.

+*Ingle, C. (2007). Predictors of critical thinking ability among college students (Doctoral dissertation, University of Kentucky, 2007). *Dissertation Abstracts International, 68*, 141.

Inpornvijit, K. (2009). Effects of feedback on self-regulated learning strategy use and academic performance (Doctoral dissertation, University of South Alabama, 2009). *Dissertation Abstracts International, 69*, 170.

+*Jacobson, R. (2000). Differences between traditional and nontraditional students on the Motivated Strategies for Learning Questionnaire (Doctoral dissertation, Auburn University, 2000). *Dissertation Abstracts International, 61*, 97.

+*Jacobson, R., & Harris, S. (2008). Does the type of campus influence self-regulated learning as measured by the Motivated Strategies for Learning Questionnaire (MSLQ)? *Education, 128*(3), 412-431.

+*Jans, S. (1997). *Improving adolescents' motivation through the use of creative teaching in the industrial arts*. Unpublished master's thesis, Saint Xavier University, Chicago, IL.

*Jones, M., Alexander, J., & Estell, D. (2010). Homophily among peer groups members' perceived self-regulated learning. *The Journal of Experimental Education, 78*, 378-394.

*Jones, M., Estell, D., & Alexander, J. (2008). Friends, classmates, and self-regulated learning: Discussions with peers inside and outside the classroom. *Metacognition and Learning, 3*(1), 1-15.

+Kadioglu, C., & Uzuntiryaki, E. (2008, March). *Motivational factors contributing to Turkish high school students' achievement in gases and chemical reactions*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Karabenick, S. (1998). Help seeking as a strategic resource. In S. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 1-11). Mahwah, NJ: Lawrence Erlbaum Associates.

Karabenick, S. (2003). Help seeking in large college classes: A person-centered approach. *Contemporary Educational Psychology, 28,* 37-58.

Karabenick, S. (2004). Perceived achievement goal structure and college student help seeking. *Journal of Educational Psychology, 96*(3), 569-581.

[*]Karadeniz, S., Buyukozturk, S., Akgun, O., Cakmak, E., & Demirel, F. (2008). The Turkish adaptation study of Motivated Strategies for Learning Questionnaire (MSLQ) for 12-18 year old children: Results of confirmatory factor analysis. *The Turkish Online Journal of Educational Technology, 7*(4), 108-117.

[+*]Kaya, S. (2008) The influences of student views related to mathematics and self-regulated learning on achievement of Algebra I students (Doctoral dissertation, Ohio State University, 2008). *Dissertation Abstracts International, 68*, 292.

Keiffer, K., Cronin, C., & Fister, M. (2004). Exploring variability and sources of measurement error in Alcohol Expectancy Questionnaire reliability coefficients: A meta-analytic reliability generalization study. *Journal of Studies on Alcohol*, 663-671.

Kelley, T. (1927). *Interpretation of educational measurements*. Yonkers, NY: World Book Co.

[*]Kesici, Ş., Baloğlu, M., & Deniz, M. (2011). Self-regulated learning strategies in relation with statistics anxiety. *Learning and Individual Differences, 21*(4), 472-477.

Kieffer, K., & MacDonald, G. (2011). Exploring factors that affect score reliability and variability in the Ways of Coping Questionnaire reliability coefficients: A meta-analytic reliability generalization study. *Journal of Individual Differences, 32*(1), 26-38.

[+*]Kilic-Bebek, E. (2010). Explaining math achievement: Personality, motivation, and trust (Doctoral dissertation, Cleveland State University, 2010). *Dissertation Abstracts International, 71*, 156.

[+*]Kimber, C. (2009). The effect of training in self-regulated learning on math anxiety and achievement among preservice elementary teachers in a freshman course in mathematics concepts (Doctoral dissertation, Temple University, 2009). *Dissertation Abstracts International, 70*, 171.

[+]Kirk-Kuwaye, M. (1995). Academic achievement motivation in college students: The effect of grade information on motivation (Doctoral dissertation, University of Hawaii, 1994). *Dissertation Abstracts International, 55*, 180.

[+]Koksal, M. (2009). Vocational high school students' sense of self-efficacy and test anxiety regarding biology learning. *Inonu University Journal of the Faculty of Education, 10*(1), 57-67.

[+*]Kosnin, A. (2007). Self-regulated learning and academic achievement in Malaysian undergraduates. *International Education Journal, 8*(1), 221-228.

[*]Kramarski, B., & Michalsky, T. (2009). Three metacognitive approaches to training pre-service teachers in different learning phases of technological pedagogical content knowledge. *Educational Research and Evaluation, 15*(5), 465-485.

[+*]Krov, K. (2010). Registered nurse persistence in baccalaureate education (Doctoral dissertation, Walden University, 2010). *Dissertation Abstracts International, 71*, 155.

Lane, G., White, A., & Henson, R. (2002). Expanding reliability generalization methods with KR-21 estimates an RG study of the Coopersmith Self-Esteem Inventory. *Educational and Psychological Measurement, 62*(4), 685-711.

[+*]Langley, S., II. (2008). Identifying self-regulatory factors that influence the academic achievement motivation of underprepared college students (Doctoral dissertation, University of Minnesota, 2007). *Dissertation Abstracts International, 68*, 80.

Laszlo, F., Jr., & Kupritz, V. (2003). The identification of online learning motives in use by undergraduate students. *Delta Pi Epsilon Journal, 45*(1), 63-72.

[+*]Lee, J., Yin, H., & Zhang, Z. (2010). Adaptation and analysis of Motivated Strategies for Learning Questionnaire in the Chinese setting. *International Journal of Testing, 10*(2), 149-165.

[+]Lee, L. (1998). Goal orientation, goal setting, and academic performance in college students: An integrated model of achievement motivation in school settings (Doctoral dissertation, University of Southern California, 1997). *Dissertation Abstracts International, 59*, 117.

Lefrancois, G. (1995). *Learning: Science and theory. Theories of human learning Kro's report* (3rd ed.). Pacific Grove, CA: Brooks/Cole Publishing Company.

Lent, R., Brown, S., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*, 79-122.

Lepper, M., & Hodell, M. (1989). Intrinsic motivation in the classroom. In C. Ames, & R. Ames (Eds.), *Research on motivation in education: Goals and cognitions* (Vol. 3, pp. 73-105). San Diego, CA: Academic Press, Inc.

[+]Lin, Y., Lin, G., & Laffey, J. (2008). Building a social and motivational framework for understanding satisfaction in online learning. *Journal of Educational Computing Research, 38*(1), 1-27.

Lin, Y., McKeachie, W., & Kim, Y. (2003). College student intrinsic and/or extrinsic motivation and learning. *Learning and Individual Differences, 13*, 251-258.

Linnenbrink, E., & Pintrich, P. (2000) Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In Sansone, C. & Harackiewicz, J. (Eds.). *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance* (pp. 195-227). San Diego, CA: Academic Press, Inc.

Linnenbrink, E., & Pintrich, P. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist, 37*(2), 69-78.

+*Little, A. (2008). An examination of motivational strategies and academic achievement in an online high school learning environment (Doctoral dissertation, George Mason University, 2008). *Dissertation Abstracts International, 69*, 139.

+*Liu, E., & Lin, C. (2010). The survey of Mathematics Motivated Strategies for Learning Questionnaire (MMSLQ) for grade 10-12 Taiwanese students. The Turkish Online *Journal of Educational Technology, 9*(2), 221-233.

+*Liu, M., & Hsiao, Y. (2001, June). *Middle school students as multimedia designers: A project-based learning approach.* Paper presented at the meeting of the National Education Computing Conference, Chicago, IL.

*Livingston, J. (2000). A qualitative study of university students' post-course strategy use (Doctoral dissertation, State University of New York Buffalo, 2000). *Dissertation Abstracts International, 61*, 241.

+*Longo, N. (2000). The relationship between identity style and self-regulated learning (Doctoral dissertation, University of Southern California, 1999). *Dissertation Abstracts International, 60*, 109.

Lopez-Pina, J., Sanchez-Meca, J., & Rosa-Alcazar, A. (2009). The Hamilton Rating Scale for Depression: A meta-analytic reliability generalization study. *International Journal of Clinical and Health Psychology, 9*(1), 143-159.

Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley Publishing Company, Inc.

Lynch, D. (2010). Motivational beliefs and learning strategies as predictors of academic performance in college physics. *College Student Journal, 44*(4), 920-927.

Lynch, R., & Dembo, M. (2004). The relationship between self-regulation and online learning in a blended learning context. *International Review of Research in Open and Distance Learning, 5*(2).

*Mack, A. (2008). Differences in academic performance and self-regulated learning based on level of student participation in supplemental instruction (Doctoral dissertation, University of Central Florida, 2007). *Dissertation Abstracts International, 69*, 233.

*Maslin, L. (1997). Self-regulated learning and science achievement in a community college (Doctoral dissertation, University of Southern California, 1997). *Dissertation Abstracts International, 58*, 127.

+Matthews, B. (2004). The effects of direct and problem-based learning instruction in an undergraduate introductory engineering graphics course. (Doctoral dissertation, North Carolina State University, 2004). *Dissertation Abstracts International, 65*(11), 197.

Matuga, J. (2009). Self-regulation, goal orientation, and academic achievement of secondary students in online university courses. *Educational Technology & Society, 12*(3), 4-11.

McKeachie, W. (1990). Learning, thinking, and Thorndike. *Educational Psychologist, 25*(2), 127-141.

[+*]McKeachie, W., Lin, Y., & Middleton, M. (2004). Two types of low test-anxious (low-worry) students. *Counseling & Clinical Psychology Journal, 1*(3), 141-152.

[+*]McKeachie, W., Lin, Y., & Strayer, J. (2002). Creationist vs. evolutionary beliefs: Effects on learning biology. *The American Biology Teacher, 64*(3), 189-192.

McKeachie, W., Pintrich, P., Lin, Y., & Smith, D. (1986). *Teaching and learning in the college classroom: A review of the research literature*. Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

[+*]McKenzie, K., & Gow, K. (2004). Exploring the first year academic achievement of school leavers and mature-age students through structural equation modeling. *Learning & Individual Differences, 14*(2), 107-123.

McManus, T. (2000). Individualizing instruction in a Web-based hypermedia learning environment: Nonlinearity, advance organizers, and self-regulated learners. *Journal of Interactive Learning Research, 11*(2), 219-251.

[+*]McWhorter, W. (2009). The effectiveness of using LEGO Mindstorms systems robotics activities to influence self-regulated learning in a university introductory computer programming course (Doctoral dissertation, University of North Texas, 2008). *Dissertation Abstracts International, 69*, 153.

McWhorter, W., & O'Connor, B. (2009, March). *Do LEGO Mindstorms motivate students in CS1?* Paper presented at the 40th ACM Technical Symposium on Computer Science Education.Chattanooga, TN.

Meier, S., & Davis, S. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology, 37*(1), 113-115.

Merriam-Webster's collegiate dictionary (11th ed.). (2003). Springfield, MA: Merriam-Webster.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*(9), 3-33.

Meyers, L., Gamst, G., & Guarino, A. (2006). *Applied Multivariate Research: Design and Interpretation.* Thousand Oaks, CA: Sage Publications.

Middleton, M., & Midgley, C. (1997). Avoiding the demonstration of lack of ability: An underexplored aspect of goal theory. *Journal of Educational Psychology, 89*(4), 710-718.

Miles, M., & Huberman, M. (1994) *An expanded sourcebook: Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.

Miller, B., Byrne, Z., Rutherford, M., & Hansen, A. (2009). Perceptions of Organizational Politics: A demonstration of the reliability generalization technique. *Journal of Managerial Issues, 21*(2), 280-300.

Miller, C., Woodson, J., Howell, R., & Shields, A. (2009). Assessing the reliability of scores produced by the Substance Abuse Subtle Screening Inventory. *Substance Use & Misuse, 44*, 1090-1100.

Milner, A. (2009). The effects of constructivist classroom contextual factors in a life science laboratory and a traditional science classroom on elementary students' motivation and learning strategies (Doctoral dissertation, University of Toledo, 2008). *Dissertation Abstracts International, 69*, 236.

+Miltiadou, M. (2001). Motivational constructs as predictors of success in the online classroom (Doctoral dissertation, Arizona State University, 2000). *Dissertation Abstracts International, 61*, 199.

Mislevy, R. (1994). *Can there be reliability without "reliability?"*. Princeton, NJ: Educational Testing Service.

+*Moak, C. (2002). The psychometric properties of the Learning and Study Strategies Inventory (Doctoral dissertation, Texas A&M University, 2002). *Dissertation Abstracts International, 63,* 231.

*Monetti, D. (2002). A multiple regression analysis of self-regulated learning, epistemology, and student achievement (Doctoral dissertation, Florida State University, 2001). *Dissertation Abstracts International, 62*, 170.

+*Moore, R. (2007). Course performance, locus of control, and academic motivation among developmental education students. *Research & Teaching in Developmental Education, 24*(1), 46-62.

+Moos, D., & Honkomp, B. (2011). Adventure learning: Motivating students in a Minnesota middle school. *Journal of Research on Technology in Education, 43*(3), 231-252.

Moss, P. (2004). The meaning and consequences of "reliability". *Journal of Educational and Behavioral Statistics, 29*(2), 245-249.

*Muis, K., Winne, P., & Jamieson-Noel, D. (2007). Using a multitrait-multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology, 77*(1), 177-195.

*Mullen, P. (2007). Use of self-regulating learning strategies by students in the second and third trimesters of an accelerated second-degree baccalaureate nursing program. *Journal of Nursing Education, 46*(9), 406-412.

[*]Mullen, P. (2010). Identification and comparison of academic self regulatory learning strategy use of students enrolled in traditional and accelerated baccalaureate degree nursing programs (Doctoral dissertation, Indiana University - Purdue University Indianapolis, 2010). *Dissertation Abstracts International, 70*, 155.

Nelson,-Le Gall, S. (1981). Help-seeking an understudied problem-solving skill in children. *Developmental Review, 1,* 224-246.

Nelson,-Le Gall, S. (1986). *Help-seeking behavior in learning* (LRDC-1986/3). Washington DC: National Institute of Education.

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

[+*]Nicholson, H. (2008). A comparison of the motivations and learning strategies employed by adult learners in industrial training programs (Doctoral dissertation, University of North Carolina, Greensboro, 2007). *Dissertation Abstracts International, 68*, 144.

[+*]Niemczyk, M., & Savenye, W. (2001, November). *The relationship of student motivation and self-regulated learning strategies to performance in an undergraduate computer literacy course.* Paper presented at the meeting of the National Convention of the Association for Educational Communications and Technology, Atlanta, GA.

Niemi, H., Nevgi, A., & Virtanen, P. (2003). Towards self-regulation in web-based learning. *Journal of Educational Media, 28*(1), 49-71.

Nokelainen, P., & Ruohotie, P. (2000, April). *Modern modeling of student motivation and self-regulated learning*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Nunnally, J. (1967). *Psychometric Theory*. New York: McGraw-Hill Book Company.

Nunnally, J. (1970). *Introduction to Psychological Measurement*. New York: McGraw-Hill Book Company.

Nunnally, J. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill Book Company.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill Book Company.

[+*]Olive-Taylor, R. (2008). The impact of an interdisciplinary seminar on first-year university students' development of personal epistemology and motivational and strategic components of self-regulated learning (Doctoral dissertation, University of North Carolina, Greensboro, 2007). *Dissertation Abstracts International, 69*, 171.

[+]Orhan, F. (2008). Self-regulation strategies used in a practicum course: A study of motivation and teaching self-efficacy. *H. U. Journal of Education, 35*, 251-262.

O'rourke, N. (2004). Reliability generalization of responses by care providers to the Center for Epidemiologic Studies-Depression Scale. *Educational and Psychological Measurement, 64*(6), 973-990.

*Ostavar, S., & Khayyer, M. (2004). Relations of motivational beliefs and self-regulated learning outcomes for Iranian college students. *Psychological Reports, 94*(3), 1202-1204.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543-578.

Pajares, F., & Miller, M. (1994). The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193-203.

Paracha, S., Mohamad, M., Jehanzeb, S., & Yoshie, O. (2009). Promoting autonomous computer assisted language learning. *Journal of Theoretical and Applied Information Technology, 5*(4), 493-498.

Paris, S., & Paris, A. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89-101.

*Payne, O. (1992, April). *The effects of learning strategies on a group of black secondary students' verbal and mathematics SAT scores*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

*Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R., & Perry, R. (2010). Boredom in achievement settings: Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531-549.

*Perry, R., Hladkyj, S., Pekrun, R., & Pelletier, S. (2001). Academic control and action control in the achievement of college students: A longitudinal field study. *Journal of Educational Psychology, 93*(4), 776-789.

+*Pfister, T. (2002). The effects of self-monitoring on academic procrastination, self-efficacy and achievement (Doctoral dissertation, Florida State University, 2002). *Dissertation Abstracts International, 63*, 172.

+Phan, H. (2007). An examination of reflective thinking, learning approaches, and self-efficacy beliefs at the University of the South Pacific: A path analysis approach. *Educational Psychology, 27*(6), 789-806.

Pintrich, P. (2000a). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544-555.

Pintrich, P. (2000b). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic Press.

Pintrich, P. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice, 41*(4), 219-225.

Pintrich, P. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Pintrich, P. & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in classroom academic tasks. In D. H. Schunk & J. L. Meese (Eds.), *Student perceptions in the classroom* (pp. 149-183). Hillsdale, NJ: Erlbaum.

Pintrich, P. & Schunk, D. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice-Hall.

+*Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Pintrich, P., Smith, D., Garcia, T., & McKeachie, W. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*. 801-813.

+*Polleys, M. (2001). A study of relationships between self-regulated learning, personality, and achievement (Doctoral dissertation, Auburn University, 2000). *Dissertation Abstracts International, 61*, 131.

+Poole, G. (1999). Assessment of factors influencing student motivation for higher academic achievement at Caney Creek High School in the Conroe independent school district (Doctoral dissertation, Texas A&M University, 1999). *Dissertation Abstracts International, 60*, 89.

+Poynter, B. (1999). A comparison of the effects of test-taking strategy instruction on goal orientation, self-efficacy, and test anxiety between secondary students with and without learning disabilities in different learning environments (Doctoral dissertation, Oklahoma State University, 1998). *Dissertation Abstracts International, 59*, 173.

*Puzziferro, M. (2008). Online technologies self-efficacy and self-regulated learning as predictors of final grade and satisfaction in college-level online courses. *American Journal of Distance Education, 22*(2), 72-89.

[*]Quible, Z. (2006). Analysis of the motivational orientation of and learning strategies used by students in a written business communication course. *Delta Pi Epsilon Journal, 48*(3), 168-190.

[+]Rasmussen, J. (2010). An investigation of scholar-baller and non-scholar baller division I football student athletes' academic, athletic, intrinsic motivation and athletic identity (Doctoral dissertation, University of Central Florida, 2010). *Dissertation Abstracts International, 71*, 861.

[+]Ray, M. (2003). The role of aptitude and achievement in developmental college students' motivational beliefs, perceptions of self-regulated learning, and accuracy of grade predictions (Doctoral dissertation, University of Missouri, Kansas City, 2003). *Dissertation Abstracts International, 64*, 88.

Raynor, J. (1981). Future orientation and achievement motivation: Toward a theory of personality functioning and change. In G. d'Ydewalle, & M. Lens (Eds.), *Cognition in human motivation and learning* (pp. 199-231). Leuven & Hillsdale, NJ: Leuven University Press & Erlbaum.

Reese, R., Kieffer, K., & Briggs, B. (2002). A reliability generalization study of select measures of Adult Attachment Style. *Educational and Psychological Measurement, 62*(4), 619-646.

[+]Rey, D. (2010). The relationship of gratitude and subjective well-being to self-efficacy and control of learning beliefs among college students (Doctoral dissertation, University of Southern California, 2010). *Dissertation Abstracts International, 70,* 135.

[+]Richardson, J. (2007). Motives, attitudes and approaches to studying in distance education. *Higher Education, 54*(3), 385-416.

Rosenthal, R. (1979). The 'file drawer' problem' and tolerance for null results. *Psychological Bulletin, 86*, 638-641.

Ross, M., Blackburn, M., & Forbes, S. (2005). Reliability generalization of the Patterns of Adaptive Learning Survey Goal Orientation Scales. *Educational and Psychological Measurement, 65*(3), 451-464.

Ross, M., Blackburn, M., Salisbury-Glennon, J., Forbes, S., & Miller, R. (2002, April). *Achievement goal orientation instruments: An empirical analysis of the commonalities and differences among scales*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

[+*]Rotgans, J., & Schmidt, H. (2009). Examination of the context-specific nature of self-regulated learning. *Educational Studies, 35*(3), 239-253.

Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and Applied, 80*(1), 1-28.

Rowley, G. (1976). The reliability of observational measures. *American Educational Research Journal, 13,* 51-59.

+Ruiz, E. (2007). Rhizomatous motivational aspects of Hispanic/Latino students in regular high school programs (Doctoral dissertation, Texas A&M University, 2006). *Dissertation Abstracts International, 67*, 127.

Ryan, A., Gheen, M., & Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teacher's social-emotional role, and the classroom goal structure. *Journal of Educational Psychology, 90*(3), 528-535.

Ryan, A., & Pintrich, P. (1998). Achievement and social motivational influences on help seeking in the classroom. In S. Karabenick (Ed.), *Strategic help seeking: Implications for learning and teaching* (pp. 117-139). Malwah, NJ: Lawrence Erlbaum.

Ryngala, D., Shields, A., & Caruso, J. (2005). Reliability generalization of the Revised Children's Manifest Anxiety Scale. *Educational and Psychological Measurement, 65*(2), 259-271.

+*Saglam, M. (2010). Students' performance awareness, motivational orientations and learning strategies in a problem-based electromagnetism course. *Asia Pacific Forum on Science Learning and Teaching, 11*(1).

+*Salamonson, Y., Everett, B., Koch, J., Wilson, I., & Davidson, P. (2009). Learning strategies of first year nursing and medical students: A comparative study. *International Journal of Nursing Studies, 46*, 1541-1547.

Schatt, M. (2011). High school instrumental music students' attitudes and beliefs regarding practice: An application of attribution theory. *Update: Applications of Research in Music Education, 29*(2), 29-40.

Schunk, D. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3 & 4), 207-231.

Schunk, D. (2004). *Learning theories: An educational perspective* (4th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

+*Schutz, P., Drogosz, L., White, V., & Distefano, C. (1998). Prior knowledge, attititude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10*(4), 291-308.

Scriven, M., & Paul, R. *Defining critical thinking*. National Council for Excellence in Critical Thinking Instruction. Retrieved from http://www.criticalthinking.org/pages/our-concept-and-definition-of-critical-thinking/411 on January 17, 2012.

Shields, A., Campfield, D., Miller, C., Howell, R., Wallace, K., & Weiss, R. (2008). Score reliability of Adolescent Alcohol Screening Measures: A meta-analytic inquiry. *Journal of Child & Adolescent Substance Abuse, 17*(4), 75-97.

Shields, A., & Caruso, J. (2004). A reliability induction and reliability generalization study of the Cage Questionnaire. Educational and Psychological Measurement, 64(2), 254-270.

Shuell, T. (1986). Cognitive conceptions of learning. *Review of Educational Research, 56*(4), 411-436.

+*Silagyi-Rebovich, E. (1997). Motivation and learning strategies among dietetic students at four-year colleges and universities (Doctoral dissertation, University of South Carolina, 1996). *Dissertation Abstracts International, 57*, 171.

Sims, B., & Sperling, R. (2001, February). *Changes in self-regulated learning within the first semesters of university enrollment.* Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, SC.

Slavin, R. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*, 43-69.

+*Spencer, D. (1999). Differences in motivated strategies and learning strategies of traditional and non-traditional age undergraduates enrolled in elective general education courses (Doctoral dissertation, University of San Francisco, 1999). *Dissertation Abstracts International, 60*, 157.

+*Sperling, R., Howard, B., Staley, R., & DuBois, N. (2004). Metacognition and self-regulated learning constructs. *Educational Research & Evaluation, 10*(2), 117-139.

Spitzer, T. (2000). Predictors of college success: A comparison of traditional and nontraditional age students. *NASPA Journal, 38*(1), 82-98.

+Stansbury, S. (1998). The effect of parental education, prior achievement, self-efficacy, goal orientation, and effort on undergraduate science performance of latinos and caucasians (Doctoral dissertation, University of Southern California, 1997). *Dissertation Abstracts International, 59*, 130.

Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.

Stevenson, H., & Chen (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science*, *259*(5091), 53-58.

Stigler, J., Smith, S., & Mao, L. (1985). The self-perception of competence by Chinese children. *Child Development, 56*, 1259-1270.

Streiner, D. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*(3), 217-222.

[+*]Sullivan, M. (2004). Self-regulated learning of medical students: Assessment of a social cognitive model (Doctoral dissertation, University of Southern California, 2003). *Dissertation Abstracts International, 65*, 128.

[+*]Sungur, S. (2007). Modeling the relationships among students' motivational beliefs, metacognitive strategy use, and effort regulation. *Scandinavian Journal of Educational Research, 51*(3), 315-326.

[+*]Sungur, S., & Tekkaya, C. (2006). Effects of problem-based learning and traditional instruction on self-regulated learning. *Journal of Educational Research, 99*(5), 307-317.

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55*, 525-534.

Thompson, B. (1999, January). *Reliability generalization: An important meta-analytic method, because it is incorrect to say, "the test is unreliable".* Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.

Thompson, B., & Snyder, P. (1998). Statistical significance and reliability analyses in recent Journal of Counseling and Development Research Articles. *Journal of Counseling and Development, 76*, 436-441.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement, 60*(2), 174-195.

Thorndike, R. (1951). Reliability. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 560-620), Washington DC: American Council on Education.

Thorndike, R. (2005). *Measurement and evaluation in psychology and education*. Columbus, OH: Prentice-Hall.

[+]Turner, J., & Schallert, D. (2001). Excpectancy-value relationships of shame reactions and shame resiliency. *Journal of Educational Psychology, 93*(2), 320-329.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6-20

Vacha-Haase, T., Henson, R., & Caruso, J. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*(4), 562-569.

Vacha-Haase, T., Kogan, L., Tani, C., & Woodall, R. (2001). Reliability generalization: Exploring variation of reliability coefficients of MMPI Clinical Scales scores. *Educational and Psychological Measurement, 61*(1), 45-59.

Vacha-Haase, T., Kogan, L, & Thompson, B. (2000). Sample compositions and variabilities in published studies versus those in test manuals: Validity of score reliability inductions. *Educational and Psychological Measurement, 60*(4), 509-522.

Vacha-Haase, T., Ness, C., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *The Journal of Experimental Education, 67*(4), 335-341.

+*Van Zile-Tamsen, C. (1997, February). *Self-directed learning and adults: The role of personal control beliefs*. Paper presented at the annual meeting of the Eastern Educational Research Association, Hilton Head, SC.

*VanZile-Tamsen, C. (2001). The predictive power of expectancy of success and task value for college students' self-regulated strategy use. *Journal of College Student Development, 42*(3), 233-241.

+VanZile-Tamsen, C., & Livingston, J. (1999). The differential impact of motivation on the self-regulated strategy use of high- and low-achieving college students. *Journal of College Student Development, 40*(1), 54-60.

Vassar, M., & Bradley, G. (2010). A reliability generalization study of coefficient alpha for the Life Orientation Test. *Journal of Personality Assessment, 92*(4), 362-370.

Vassar, M., & Crosby, J. (2008). A reliability generalization study of coefficient alpha for the UCLA Loneliness Scale. *Journal of Personality Assessment, 90*(6), 601-607.

+Vazquez, M. (2009). Problem-based learning and its influence on college preparation knowledge, motivation, & self-efficacy in high school students (Doctoral dissertation, University of Southern California, 2009). *Dissertation Abstracts International, 69*, 155.

Vogt, C. (2003). An account of women's progress in engineering: A social cognitive perspective. *Journal of Women & Minorities in Sciences & Engineering, 9*(3/4), 217-238.

Wallace, K., & Wheeler, A. (2002). Reliability generalization of the Life Satisfaction Index. *Educational and Psychological Measurement, 62*(4), 674-684.

+*Watson, M., McSorley, M., Foxcroft, C., & Watson, A. (2004). Exploring the motivation orientation and learning strategies of first year university learners. *Tertiary Education and Management, 10*, 193-207.

Weiner, B. (1986). *An attributional theory of motivation and motion*. New York: Springer-Varlag.

Weinstein, C., & Mayer, R. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *The handbook of research on teaching* (pp. 315-327). New York: Macmillan.

Whang, P., & Hancock, G. (1994). Motivation and mathematics achievement: Comparisons between Asian-American and non-Asian students. *Contemporary Educational Psychology, 19*, 302-322.

Wheeler, D., Vassar, M., Worley, J., & Barnes, L. (2011). A reliability generalization meta-analysis of coefficient alpha for the Maslach Burnout Inventory. *Educational and Psychological Measurement, 71*(1), 231-244.

Wigfield, A., & Eccles, J. (2000). Expectancy-value theory of motivation. *Contemporary Educational Psychology, 25*, 68-81.

+Wilke, R. (2001). The effect of active learning on college students' achievement, motivation, and self-efficacy in a human physiology course for non-majors (Doctoral dissertation, University of Texas at Austin, 2001). *Dissertation Abstracts International, 61,* 225.

Wilke, R. (2003). The effect of active learning on student characteristics in a human physiology course for nonmajors. *Advances in Physiology Education, 27*(4), 207-223.

Wilkinson, L., & The APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.

+Willems, P. (2001). The effects of situated cognition, academic efficacy, and intrinsic motivation on education students' learning of reciprocal teaching (Doctoral dissertation, University of Florida, 2000). *Dissertation Abstracts International, 61*, 120.

+*Williams, M. (2009). Components of self-regulated learning in high school students with learning disabilities (Doctoral dissertation, Indiana University of Pennsylvania, 2008). *Dissertation Abstracts International, 69*, 222.

Willson, V. (1980). Research techniques in AERJ articles: 1969 to 1978. *Educational Researcher, 9*(6), 5-10.

Winne, P., & Perry, N. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531-566). San Diego, CA: Academic Press.

Wittrock, M. (Ed.). (1977). *Learning and instruction*. Berkeley, CA, McCutchan Publishing Corp.

*Wolters, C. (2003). Understanding procrastination from a self-regulated learning perspective. *Journal of Educational Psychology, 95*(1), 179-187.

Wolters, C. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*(2), 236-250.

Yeager, M. (2010). A cross-validation study of the College Learning Effectiveness Inventory (CLEI) (Doctoral dissertation, Kansas State University, 2010). *Dissertation Abstracts International, 70*, 4579.

Yen, W., & Fitzpatrick, A. (2006). Item response theory. In R. L. Brennan, (Ed.), *Educational measurement* (4th ed. Pp. 111-153). Westport, CN: Greenwood Publishing Group.

[+*]Yukselturk, E., & Bulut, S. (2005). Relationships among self-regulated learning components, motivational beliefs and computer programming achievement in an online learning environment. *Mediterranean Journal of Educational Studies, 10*(1), 91-112.

[+*]Yukselturk, E., & Bulut, S. (2007). Predictors for student success in an online course. *Journal of Educational Technology & Society, 10*(2), 71-83.

[+]Yukselturk, E., & Bulut, S. (2009). Gender differences in self-regulated online learning environment. *Journal of Educational Technology & Society, 12*(3), 12-22.

[+*]Yusri, G., & Rahimi, N. (2010). Self-regulated learning strategies among students of Arabic language course and intensive Arabic course in Mara University of Technology Malaysia (Uitm). *International Journal of Applied Educational Studies, 8*(1), 57-67.

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum Press.

Zimmerman, B. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*, 307-313.

Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.) *Handbook of self-regulation* (pp. 13-39). New York: Academic.

Zimmerman, B., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal, 23*, 614-628.

Zimmerman, B., & Risenberg, R. (1997). Self-regulatory dimensions of academic learning and motivation. In G. Phye (Ed.), *Handbook of academic learning* (pp. 105-125). San Diego, CA: Acadmemic Press, Inc.

[+*]Zusho, A., Pintrich, P., & Coppola, B. (2003). Skill and will: The role of motivation and cognition in the learning of college chemistry. *International Journal of Science Education, 25*(9), 1081-1094.

Appendix A: MSLQ Instrument

Motivation Scales

<u>Value Components</u>
Intrinsic Goal Orientation
1. In a class like this, I prefer course material that really challenges me so I can learn new things.
16. In a class like this, I prefer course material that arouses my curiosity, even if it is difficult to learn.
22. The most satisfying thing for me in this course is trying to understand the content as thoroughly as possible.
24. When I have the opportunity in this class, I choose course assignments that I can learn from even if they don't guarantee a good grade.

Extrinsic Goal Orientation
7. Getting a good grade in this class is the most satisfying thing for me right now.
11. The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade.
13. If I can, I want to get better grades in this class than most of the other students.
30. I want to do well in this class because it is important to show my ability to my family, friends, employer, or others.

Task Value
4. I think I will be able to use what I learn in this course in other courses.
10. It is important for me to learn the course material in this class.
17. I am very interested in the content area of this course.
23. I think the course material in this class is useful for me to learn.
26. I like the subject matter of this course.
27. Understanding the subject matter of this course is very important to me.

<u>Expectancy Components</u>
Control of Learning Beliefs
2. If I study in appropriate ways, then I will be able to learn the course material in this course.
9. It is my own fault if I don't learn the material in this course.
18. If I try hard enough, then I will understand the course material.
25. If I don't understand the course material, it is because I didn't try hard enough.

Self-Efficacy for Learning and Performance
5. I believe I will receive an excellent grade in this class.
6. I'm certain I can understand the most difficult course material presented in the readings for this course.
12. I'm confident I can understand the basic concepts taught in this course.
15. I'm confident I can understand the most complex material presented by the instructor in this course.
20. I'm confident I can do an excellent job on the assignments and tests in this course.
21. I expect to do well in this class.
29.I'm certain I can master the skills being taught in this class.
31. Considering the difficulty of this course, the teacher, and my skills, I think I will do well in this class.

<u>Affective Components</u>
Test Anxiety
3. When I take a test I think about how poorly I am doing compared with other students.
8. When I take a test I think about items on other parts of the test I can't answer.
14. When I take tests I think if the consequences of failing.
19. I have an uneasy, upset feeling when I take an exam.
28. I feel my heart beating fast when I take an exam.

Learning Strategies Scales

<u>Cognitive and Metacognitive Strategies</u>

Rehearsal

39. When I study for this class, I practice saying the material to myself over and over.
46. When studying for this class, I read my class notes and the course readings over and over again.
59. I memorize key words to remind me of important concepts in this class.
72. I make lists of important terms for this course and memorize the lists.

Elaboration

53. When I study for this class, I pull together information from different sources, such as lectures, readings, and discussions.
62. I try to relate ideas in this subject to those in other courses whenever possible.
64. When reading for this class, I try to relate the material to what I already know.
67. When I study for this course, I write brief summaries of the main ideas from the readings and the concepts from the lectures.
69. I try to understand the material in this class by making connections between the readings and the concepts from the lectures.
81. I try to apply ideas from course readings in other class activities such as lecture and discussion.

Organization

32. When I study the readings for this course, I outline the material to help me organize my thoughts.
42. When I study for this course, I go through the readings and my class notes and try to find the most important ideas.
49. I make simple charts, diagrams, or tables to help me organize course material.
63. When I study for this course, I go over my class notes and make an outline of important concepts.

Critical Thinking

38. I often find myself questioning things I hear or read in this course to decide if I find them convincing.
47. When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence.
51. I treat the course material as a starting point and try to develop my own ideas about it.
66. I try to play around with ideas of my own related to what I am learning in this course.
71. Whenever I read or hear an assertion or conclusion in this class, I think about possible alternatives.

Metacognitive Self-Regulation

33. During class time I often miss important points because I am thinking of other things. (REVERSED)
36. When reading for this course, I make up questions to help focus my reading.
41. When I become confused about something I'm reading for this class, I go back and try to figure it out.
44. If course materials are difficult to understand, I change the way I read the material.
54. Before I study new course material thoroughly, I often skim it to see how it is organized.
55. I ask myself questions to make sure I understand the material I have been studying in this class.
56. I try to change the way I study in order to fit the course requirements and instructor's teaching style.
57. I often find that I have been reading for class but don't know what it was all about. (REVERSED)
61. I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when studying.
76. When studying for this course I try to determine which concepts I don't understand well.
78. When I study for this class, I set goals for myself in order to direct my activities in each study period.
79. If I get confused taking notes in class, I make sure I sort it out afterwards.

Resource Management Strategies

Time and Study Environment

35. I usually study in a place where I can concentrate on my course work.
43. I make good use of my study time for this course.
52. I find it hard to stick to a study schedule. (REVERSED)
65. I have a regular place set aside for studying.
70. I make sure I keep up with the weekly readings and assignments for this course.
73. I attend class regularly.
77. I often find that I don't spend very much time on this course because of other activities. (REVERSED)
80. I rarely find time to review my notes or readings before an exam. (REVERSED)

Effort Regulation

37. I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do. (REVERSED)
48. I work hard to do well in this class even if I don't like what we are doing.
60. When course work is difficult, I give up or only study the easy parts. (REVERSED)
74. Even when the course materials are dull and uninteresting, I manage to keep working until I finish.

Peer Learning

34. When studying for this course, I often try to explain the material to a classmate or a friend.
45. I try to work with other students from this class to complete the course assignments.
50. When studying for this course, I often set aside time to discuss the course material with a group of students from the class.

Help Seeking

40. Even if I have trouble learning the material for this class, I try to do the work on my own, without help from anyone. (REVERSED)
58. I ask the instructor to clarify concepts I don't understand well.
68. When I can't understand the material in this course, I ask another student in this class for help.
75. I try to identify students in this class whom I can ask for help if necessary.

Appendix B: Coding Sheet

Identification: # ______
Author(s):
Year:
Title:

Source of data: ___ Journal ___ Book ___ Thesis ___ Conf Paper ___Unpublished Report ___ Other ________

MSLQ Use: __Study Using Instrument[1] __MSLQ Validation Study[2] __References MSLQ[3] __Not English[4] __False Hit[5]

Data: __ Reliability[1] __ Means[2] __ Both μ/α[3] __Other Data[4] __None Reported[5]

Research Design: ___ Experimental[1] ___ Quasi-Experimental[2] ___ Comparative[3]___ Correlational[4] ___ Model Spec[5]
___Descriptive[6] ___ Qualitative[7] ___ Other[8](______________________)

Are the MSLQ scales used as a predictive variable? ___ Yes ___ No

Country of Study: **Translation:**

Notes:

Outcome Variable:
Mean: ***SD:***

Sample Characteristics:

Sample Size: ________ Mean Age: ______ *SD* Age: _____

Course(s): ____________________________________

Academic Population:

___ Elementary[1] ___ Middle[2] ___ Junior/High[3] ___ Undergraduate[4] ___ Graduate[5] __Other[6] (_____________)

Gender:

Male		Female	
N	%	*N*	%

Race:

Asian		Black		Hispanic		Nat Am		White		Other	
N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%

Other Characterisitcs:

N	%	*N*	%	*N*	%	*N*	%	*N*	%	*N*	%

Scale: __ MSLQ Collegiate scales (81) __Adapted MSLQ Coll. ___MSLQ Jr. High (44) ___MSLQ Adapted JH

Adapted Explained:________________________________

Reliability Measurements:
__Not report __ Manual / other study __ α coefficients provided __ α provided, but range or reference

Reliability Coefficient Used: ___ α ___ Other (________________) **Likert-Scale Used:**_____

Reported Statistics

		Total -			Group -			Group -			Group -		
		Reliability	Mean	*SD*	Reliability	Mean	*SD*	Reliability	Mean	*SD*	Reliability	Mean	*SD*
Motivation Scales [31]	**Value Components**												
	Intrinsic Goal (IG) [4]												
	Extrinsic Goal (EG) [4]												
	Task Value (TV) [6]												
	Expectancy Components												
	Control Beliefs (CB) [4]												
	Self-Efficacy (SE) [8]												
	Affective Components												
	Test Anxiety (TA) [5]												
Learning Strategy Scales [50]	**Cognitive & Metacognitive Strategies**												
	Rehearsal (R) [4]												
	Elaboration (E) [6]												
	Organization (O) [4]												
	Critical Thinking (CT) [5]												
	Metacognitive Self-Regulation (MSR) [12]												
	Resource Management Strategies												
	Time & Study Management (TSM) [8]												
	Effort Regulation (ER) [4]												
	Peer Learning (PL) [3]												
	Help Seeking (HS) [4]												
Jr High Version [44]	Intrinsic Value [9]												
	Self-Efficacy [9]												
	Test Anxiety [4]												
	Cognitive Strategy Use [13]												
	Self-Regulation [9]												