TOWARD A TREATMENT EFFECT OF AN INTERVENTION TO FOSTER SELF-REGULATED LEARNING (SRL): AN APPLICATION OF THE RASCH MODEL

by

Todd Douglas Reeves
August 4, 2009

A thesis submitted to the
Faculty of the Graduate School of
the University at Buffalo, State University of New York
in partial fulfillment of the requirements for the
degree of

Master of Arts

Department of Counseling, School and Educational Psychology

ii

ACKNOWLEDGEMENTS

The author would like to recognize the important contributions of three individuals, without whom this project would not have been possible. I would first like to thank Dr. Jaekyung Lee, my major advisor and thesis committee chair, for modeling a high standard against which graduate student research can be evaluated and for providing me with such great opportunities to develop as a researcher. Secondly, I would like to extend my sincerest gratitude to Dr. Kelly H. Ahuna, Director of the Methods of Inquiry Program (MOI), for allowing me access to the programmatic data used to conduct this research project and for being a most "excellent" mentor. I am very grateful for these individuals' willingness to serve as members of my thesis committee; their comments during the various phases of this project were tremendously helpful. Lastly, the author would also like to recognize Kelly M. Rosney for her assistance constructing one of the indicators used in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Abstract

This study investigated whether an intervention measurably contributed to the self-regulatory processes underlying undergraduate students' learning. The Rasch model was first applied to Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) data to examine the validity of inferences made from this instrument and to estimate interval person measures for the answering of substantive research questions. Psychometric analyses provided mixed evidence for the validity of DALI-R inferences. Practically meaningful growth in both active and dynamic self-regulation was observed throughout the course of the intervention. However, attempts to explain growth by motivational factors and indicators of intervention exposure were largely unsuccessful. Furthermore, differential effects by race/ethnicity and gender were not observed. Limitations and recommendations for future research are discussed.

CHAPTER ONE: INTRODUCTION

The lack of persistence from enrollment through graduation is a problem facing the American higher educational system. Of those students entering a 4-year post-secondary institution for the first time during the 1995-96 academic year, only 58.4% held a Bachelor's degree five years later. Gaps in college graduation rates across racial/ethnic and gender lines are furthermore evident in the literature. The five-year Bachelor's degree completion rates for Black (43.4%), Hispanic (44.0%), and male (54.6%) students are notably lower than those of their White (61.9%), Asian or Pacific Islander (69.1%) and female (61.6%) counterparts (Snyder, Dillow, & Hoffman, 2008).

These differential graduation rates signal a perpetuation of inequities in American society that exist along racial/ethnic and gender lines. Importantly, the failure to persist in higher education, and ultimately the failure to graduate, has consequences. The median annual earnings of a person holding a Bachelor's degree in 2006 was $43,500, while this economic indicator was $31,400 for those who only had completed some post-secondary education (Planty et al., 2008). The economic implications of this problem are similarly evidenced by a 1.6% higher unemployment rate in 2006 for those who dropped out relative to their Bachelor's degree-holding peers (Liming & Wolf, 2008).

These graduation statistics, and their (presumably) associated consequences, remind educational researchers and policy-makers that attrition before graduation is a problem at the tail end of the American educational system that needs to be met with more scholarly research. Consequently, this study was aimed at investigating one means by which this problem may be tackled systematically. Specifically, it explored whether a comprehensive, semester-length

1

intervention for undergraduate students made a measurable contribution to the self-regulatory processes underlying their academic learning.

Any discussion of a problem in the American educational system would be incomplete without a discussion of its causes, and there are myriad factors that may contribute to a student's failure to graduate. Tinto (1975) formulated a theoretical model of dropout from institutions of higher learning that conceptualizes the problem as a longitudinal process of interactions between individual- and institutional-level factors. Tinto (1975) contends that dropout ultimately occurs in response to insufficient integration into either the social or academic systems of an institution. For example, experiencing performance problems in the academic realm might prompt such a decision. Correspondingly, failing to experience favorable social interactions with peers or faculty also are predicted to have such an effect.

Individual background characteristics and attributes as well as educational aspirations and motivations are ascribed important roles as input variables in Tinto's (1975) model. Crucially, previous schooling experiences, familial backgrounds, individual learning, and other characteristics are purported to impact academic and social integration to the extent that they influence one's commitment to educational and institutional goals. In other words, even though events during a student's tenure at an institution are clearly important, characteristics of the learner him- or herself are also related (albeit indirectly) to the outcome. Given the role of individual characteristics in Tinto's (1975) model, I will now discuss two factors that may be related to this outcome.

If previous educational experiences are of import in Tinto's (1975) model, one might hypothesize that unpreparedness is a factor related to our nation's overall and subgroup degree completion rates. Such unpreparedness would be expected to contribute to suboptimal

performance in the academic realm and consequent decision making regarding persistence in degree attainment. Evidence for such unpreparedness for college-level work comes from Parsad and Lewis (2003), who reported that over 76% of degree-granting institutions offered at least one remedial course in reading, writing, or mathematics during the fall of 2000. Not unexpectedly, these remedial courses were most prevalent in public 2-year colleges (98%), and were more common in 4-year public degree-granting institutions (80%) than in their 4-year private counterparts (59%).

Unfortunately, Wirt et al. (2004) paint a bleak picture of the success of these remedial courses in preparing students. They report that those who did *not* take remedial courses while enrolled at a post-secondary institution are actually more likely to receive a Bachelor's degree than their remedial-course-taking peers. Specifically, for those enrolled in twelfth grade in 1992, and subsequently enrolled in a post-secondary institution, 69% of those not taking remedial courses had received a degree or certificate by 2000, while only 30-57% of those who had taken such courses managed to do so (depending on the number and type of courses).

Thus, it appears that remedial coursework is not necessarily providing students with what they need to be successful. If such classes were successful in preparing students, one might expect degree completion rates for remedial course-takers and non-course-takers to be more similar. Of course, the lower rate of degree completion for remedial course-takers is not likely a *result* of their having taken such courses; it more likely reflects student backgrounds, prior experiences and prerequisite knowledge and skills. Alternatively, what is provided by these domain-specific (e.g., reading, mathematics, writing) remedial courses might need to be considered in explicating their unsuccessfulness. Such courses might provide students with what they need to succeed in, say, another mathematics course, but not much else that could be called

upon in their other undergraduate courses. This research will investigate the success of a different form of intervention for undergraduate students.

In addition to previous educational experiences, Tinto's (1975) model also accounts for other characteristics upon which individuals may differ. Self-regulated learning (SRL; e.g., Zimmerman, 2000; 2002; 2008; Winne & Hadwin, 1998) might be such a construct that has implications for the capacity of students to experience success in the academic sphere of higher education. Put simply, SRL permits learners to autonomously take control of their learning processes. This notion of SRL will be taken up in much greater depth in the literature review below.

This research is grounded in the belief that a failure to engage in self-regulated learning (SRL) contributes to academic problems that precede the decision to discontinue higher education. Along these lines, instructors cannot simply assume that students engage in the kind of strategic self-regulation called for by successful higher education. It follows that if students do not possess the skills required by the level of learning at a four-year degree-granting institution, then they must acquire them. Therefore, instruction in the use of strategies to promote SRL might be one avenue toward improving the academic performance of undergraduate students so that they may graduate.

This study was aimed at evaluating the effectiveness of a theoretically-based intervention targeted for all undergraduate students. The intervention is a semester-long study-skills-type course at a large, public research university in the Northeast. The theoretical framework for much of the intervention is the biofunctional theory of self-regulation, which will be described in more detail below (Iran-Nejad, 1990), although other theoretical notions from the cognitive and educational psychology literatures are also invoked. More comprehensive descriptions of course

content and its theoretical grounding can be found elsewhere (e.g., Ahuna & Tinnesz, 2006; Tinnesz, Ahuna, & Keiner, 2006; Schapiro & Livingston, 2000).

One goal of the intervention is the mastery of cognitive strategies students can use to approach their learning of new material. To this end, students are taught concrete study techniques and required to employ them in their other classes. The various elements of these techniques are explicitly mapped onto components of the biofunctional self-regulation theory (Iran-Nejad, 1990). For example, note-taking and reading for comprehension[1] techniques taught and practiced require that students employ cognitive strategies such as summarizing material, generating questions and comprehension monitoring. Zimmerman (2008), an expert in self-regulated learning, suggests that note-taking and reading are appropriate areas of academic functioning targeted by these types of interventions. In addition to twice-weekly lectures, the course also involves a weekly 30-minute one-on-one meeting with an undergraduate peer monitor, who facilitates on-going student self-assessment of strategy implementation.

---

[1] An example version of the reading for comprehension technique can be found in Appendix C of Livingston (2000).

CHAPTER TWO: LITERATURE REVIEW AND RESEARCH QUESTIONS

This research was principally informed by the literature on self-regulated learning (SRL), although the literature on higher-order cognitive processes (i.e., metacognition) was also consulted. Additionally, this research called upon literature regarding the extent to which self-regulated learning (SRL) can be developed in students. Relevant literature from each of these areas of inquiry will now be discussed in turn.

SELF-REGULATED LEARNING (SRL)

Self-regulated learning (SRL) has received much attention from educational and other psychologists. Although Pajares (2003) notes that SRL has theoretical grounding in the work that William James conducted over a century ago, it has been studied increasingly by educational researchers and scholars during the past few decades (Zimmerman, 2008). Zimmerman (2008, pp. 1-2) describes SRL as:

> The self-directive processes and self-beliefs that enable learners to transform their mental abilities, such as verbal aptitude, into an academic performance skill, such as writing. SRL is viewed as *proactive* processes that students use to acquire academic skill, such as setting goals, selecting and deploying strategies, and self-monitoring one's effectiveness, rather than as a reactive event that happens to students due to impersonal forces.

Notwithstanding an ostensible emphasis on *process* in Zimmerman's (2008) definition of SRL, Murphy and Alexander (2000) contend that SRL is also very much contingent upon motivational factors. That is, SRL is more than just a simple knowledge and application of cognitive strategies. Instead, students who engage in SRL need not only the *skill* but also—and more importantly—the *will* to do so (McComb & Marzano, 1990).

Scholarly research has established self-regulated learning (SRL) as a construct upon which individual students differ with significant implications for educational settings. Specifically, SRL has been linked to various indicators of academic achievement, including track

placement and standardized achievement test scores (Zimmerman & Martinez-Pons, 1986; Zimmerman, 1990). SRL has also been shown to be distinct from measures of general cognitive ability, although they are correlated (Zimmerman & Bandura, 1994).

Any discussion of SRL is inextricably linked to a discussion of metacognition. Another important construct receiving much attention, metacognition is described by Flavell (1979) as "cognition about cognitive phenomena" (e.g., thinking about thinking) (p. 906). More recently, Kuhn (2000) defined metacognition as "[higher-order] cognition that reflects on, monitors, or regulates first-order cognition" (p. 178). Notably, she draws a distinction between metacognitive awareness, "an awareness of what one believes and how one knows," and metastrategic control, "[the] application of strategies that process new information" (p. 178).

Developments in the understanding of metacognition have largely framed developments in the study of self-regulated learning (SRL; Zimmerman, 2002). Put simply, metacognitive abilities permit learners to engage in the strategic aspects of SRL. That is, meta-level processes would provide a student with the awareness that he has a particular strategy to deploy and would also serve to direct the application of that strategy to a learning event. Similarly, monitoring a student's memory processes (i.e., metamemory) would allow her to evaluate the effectiveness of a particular cognitive strategy in a particular learning context. Kuhn (2000) affirms that enhancing these abilities should be an important educational goal, as they allow for both an awareness of one's cognitive processes and a means of managing or regulating them.

TRADITIONAL MODELS OF SELF-REGULATED LEARING (SRL)

There are various theoretical conceptions of self-regulated learning (SRL). Puustinen and Pulkinen (2001), who define SRL more broadly as "an intermediate construct describing the ways in which individuals regulate their own cognitive processes within an educational setting"

(p. 269), provide a review of five theoretical models that have considerable empirical support. Zimmerman (2002) notes that research on SRL—as it is traditionally understood—emerged from a social cognitive theoretical framework. Accordingly, the popular contemporary models of SRL by Zimmerman (2000; 2002) and by Winne and Hadwin (1998) both align with the social cognitive theoretical perspective. I will now discuss the former (and most prominent) model in detail.

Zimmerman (2000; 2002) proposed a cyclical model of self-regulated learning (SRL) comprised of three stages. Each stage includes processes that occur either before, during or after a learning event. In the *forethought* phase, a task is analyzed such that strategic planning can occur and performance goals can be set. Self-motivational factors (e.g., self-efficacy beliefs, outcome expectations and intrinsic interest) are germane to this phase since it concerns what will (or will not) happen prior to, during and after a particular learning event. During the next phase, *performance*, self-control strategies such as imagery, self-instruction, or attention focusing are applied to the learning of new material while the learner monitors the success of his or her efforts and progress. During the final phase, *self-reflection*, the learner reflects on his or her efforts during the learning event and uses self-evaluation to make causal attributions about one's successes or failures. During this phase, successfully applied strategies might be noted for future use; unsuccessful attempts, on the other hand, might prompt a modification of the strategic control approach used.

Zimmerman's (2000; 2002) and other (e.g., Winne and Hadwin, 1998) models of self-regulated learning (SRL) rely on the operations of the central executive, a cognitive mechanism theorized to control and manage the operation of human cognitive processes (Baddeley, 2000). Sternberg (2009) defines the central executive as "the gating mechanism that decides what

8

information to process further and how to process it [as well as] what resources to allocate to memory and related tasks, and how to allocate them" (p. 193). Importantly, the central executive would govern the attention focusing and other active mental processes called upon by self-regulated learning (SRL).

Although there are various theoretical models of self-regulated learning (SRL), this study focused on a particular theoretical account of the cognitive mechanisms by which learning processes are internally self-regulated. In particular, this study critically examined the biofunctional theory of self-regulation (Iran-Nejad, 1990) by applying the Rasch measurement model to its experimental operation. Secondly, this study attempted to evaluate the success of an intervention grounded in this particular account of how one's learning processes can be self-regulated.

THE BIOFUNCTIONAL THEORY OF SELF-REGULATION

Iran-Nejad (1990) challenges the notion that active executive control is the only mechanism by which learning is internally self-regulated, and he proposes a two-source theoretical model of learners' internal self-regulation: the biofunctional theory of self-regulation. His theory, similar to other biofunctional theories, attempts to explain the intrinsic nervous system (i.e., brain) processes that underlie cognition (e.g., learning). More specifically, this theory is concerned with the brain processes by which learning is internally self-regulated, allowing persons to ultimately commit information to long-term memory. In his article, Iran-Nejad (1990) posits both active and dynamic sources of internal self-regulation.

It is generally understood in cognitive psychology that explicit learning results from active, central executive control over one's learning processes. That is, committing information to long-term memory involves *actively* intending to learn something and strategically engaging

in control processes (e.g., rehearsal) to do so (Atkinson & Shiffrin, 1971). However, according to Iran-Nejad (1990), such learner-controlled processes are only one means by which learning processes are internally self-regulated. Iran-Nejad (1990) calls this first source of self-regulation *active* self-regulation and describes it as a voluntary, controlled process involving selective attention, self-questioning, prediction and procedural metacognition.

When one is actively self-regulating his or her learning processes, elaboration, organization, comprehension monitoring and rehearsal strategies are applied to the learning of new knowledge and skills. Elaboration involves integrating new information with one's existing knowledge structures (e.g., summarizing, using keywords or mnemonics). Organization involves relating ideas and concepts—in other words, categorizing (e.g., creating a hierarchy or graphic representation). Comprehension monitoring involves using strategies to ensure that goals are being met (e.g., self-testing). Rehearsal involves repeating the information one is learning. Thus, active self-regulation involves the performance of strategic, goal-directed acts. The similarity between the goal-directed nature of both biofunctional (Iran-Nejad, 1990) and other conceptions of self-regulation (e.g., Zimmerman, 2000; 2002) is notable. Moreover, the *active* application of cognitive strategies to the learning of new information is also a component of other self-regulated learning (SRL) models (e.g., Zimmerman, 2000; 2002).

Despite its similarities to alternative theoretical models of self-regulated learning (SRL), Iran-Nejad's (1990) biofunctional theory of self-regulation does have its differences. His conception of self-regulation as a *multi*-source phenomenon largely arose from evidence that not all learning is governed by the constrained processes of the central executive (Baddeley, 2000). In response to incidental or unintentional forms of learning that are not the result of such active

executive control, Iran-Nejad (1990) proposed a second source by which learning processes are internally self-regulated: *dynamic* self-regulation.

Iran-Nejad (1990) defines this other source of internal self-regulation as "self-regulation by the nonexecutive components of the system, implying that brain subsystems and microsystems must be capable of regulating local internal [knowledge] construction processes on their own." Dynamic self-regulation is "control exercised by the mindful brain as a whole," rather than just by the operations of the central executive (p. 587). According to Iran-Nejad and Chissom (1992), this nonexecutive (uncontrolled) influence on learning is marked by interest, curiosity, postdiction and reflective metacognition. Iran-Nejad (1990) notes that most learning occurs as a result of dynamic self-regulatory control. Implicit (or procedural) learning—that is, learning that occurs without an intention to do so (e.g., Schacter & Graf, 1986)—was not taken up by Iran-Nejad (1990).

Iran-Nejad and Chissom (1992) tested the hypothesis that two independent sources of control permit undergraduate students to internally self-regulate their learning processes (Iran-Nejad, 1990). In their study, undergraduate students ($N = 99$) completed the Dynamic and Active Learning Inventory Revised (DALI-R), which is purported to measure both active and dynamic self-regulation. To examine whether these two internal control processes do in fact explain differences in learning, participants were divided into two groups based on their overall grade point average (GPA). Mean GPAs were 2.56 and 3.22 for the low and high GPA groups, respectively. Independent samples *t*-tests revealed that both active ($p < .05$) and dynamic ($p < .001$) control scores distinguished the groups, favoring those with higher GPAs. Although effect sizes were not reported, *d*s were .40 and .67 for active and dynamic control, respectively. These effect sizes implied a practically meaningful differentiation of learners by these two constructs.

11

This differentiation of GPA groups by both active and dynamic measures was taken as evidence that dynamic control processes are an "important source of variation in individual differences in learning processes" (p. 131). However, to examine whether dynamic control represents an *independent* source of variation in learning processes, above and beyond the influence of active control, additional analyses were conducted. First, simple regression analyses predicting GPA on the basis of either active or dynamic DALI-R scale scores revealed respective statistically significant Pearson product-moment correlations of .22 and .42. However, when the influence of the other control source was accounted for, the partial correlation for only dynamic control remained statistically significant. That is, support was lent for the authors' hypothesis that dynamic internal control represents an independent contributor to learning; accounting for active self-regulation did not explain away the relationship between dynamic control and learning. In contrast, active self-regulation was found to *not* independently contribute to learning after parceling out the effect of dynamic self-regulation.

Iran-Nejad and Chissom (1992) concluded that dynamic self-regulation thus represents a real construct that contributes to learning. An alternative account of this finding, however, might take into consideration the instrumentation employed by Iran-Nejad and Chissom (1992). In their study, dynamic self-regulation was measured by approximately twice as many items as was active self-regulation. In trying to explain differences in learning, twenty additional items developed on the basis of prior research on individual differences in learning might very well be expected to explain additional variance in the dependent measure. In their article, Iran-Nejad and Chissom (1992) report preliminary analyses in which "dynamic control contributed about four times as much variance to cumulative grade point average as active control" (p. 129). To

evaluate the merits of this alternative hypothesis, this study takes a closer look at the instrumentation developed and used by Iran-Nejad and Chissom (1992).

Iran-Nejad and Chissom's (1992) other finding (i.e., that dynamic self-regulation explained the relationship between active self-regulation and learning) was unexpected. This could also be interpreted in terms of how these constructs were operationalized. For instance, the dynamic self-regulation items could be measuring the same thing as the active self-regulation items. Alternately, the dynamic self-regulation items could be capturing something more fundamentally related to learning, something without which students might not actively engage in strategy use (i.e., motivational factors). In either case, the instrumentation used in the drawing of these conclusions requires close attention.

Additionally, discovering that the use of active self-regulatory strategies is statistically unrelated to learning except through another construct—dynamic self-regulation—is inconsistent with prior research to the contrary. That is, the use of strategies (e.g., summarizing, self-questioning) is a well-established means of promoting comprehension and committing information to long-term memory (e.g., King, 1992). Furthermore, it is also a component of Iran-Nejad's (1990) theory. To explain this finding, Iran-Nejad and Chissom (1992) suggested that perhaps "active learning strategies contribute to learning to the extent that they influence the activity of dynamic sources" (p. 132). Further research is clearly needed to understand the nature of and the relationship between these two self-regulatory processes. This latter question (i.e., their relationship) is of particular importance since they are hypothesized by Iran-Nejad (1990) to govern different forms of learning.

In response to Iran-Nejad and Chissom (1992), Schapiro and Livingston (2000) investigated the role of dynamic self-regulation in academic learning. Specifically, they

attempted to examine whether dynamic control underlies active self-regulatory behavior. To examine the influence of dynamic self-regulation on learning, Schapiro and Livingston (2000) looked at its relationship with grade point average (GPA) in a large sample ($N > 300$). They drew upon programmatic data collected to evidence change in active and dynamic self-regulation as a function of the intervention described earlier. The authors hypothesized that higher levels of dynamic control would be associated with higher GPAs, regardless of participants' active self-regulatory control levels. This is exactly what they found. Participants were categorized into four quadrants based on post-intervention active and dynamic DALI-R subscale scores. After ruling out other potential confounds using simple regression analyses, a 2 (active control: high, low) x 2 (dynamic control: high, low) analysis of covariance (ANCOVA), with post-course GPA as the dependent variable and year in school as a covariate, was conducted. They report a statistically significant main effect of dynamic control on GPA in favor of those in the high dynamic control group. Neither a main effect of active control nor the interaction between active and dynamic control were statistically significant.

Schapiro and Livingston (2000) conclude that active control "does not uniquely contribute to GPA," (p. 30) and that instruction focused solely on the use of active strategies under executive control would be insufficient. Like in Iran-Nejad and Chissom's (1992) study, Schapiro and Livingston's (2000) non-significant main effect of active self-regulation is surprising. Their interpretation of this finding was that "while the dynamic may drive the student's desire to learn, active strategies or techniques provide the student with the tools to initiate learning and master the materials" (p. 34). However, if dynamic self-regulation "drive[s]" active self-regulation, an interaction might have been observed such that dynamic self-regulation

14

would make a particularly large contribution to GPA when coupled with the use of active self-regulatory strategies.

One aspect of the Schapiro and Livingston (2000) study warrants consideration. Specifically, institutional warehouse GPA and self-reported Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) data collected prior to the intervention may have been preferable for the answering of their research question. Although post-intervention data were selected so as to "put all students on an even playing field, all having been exposed to the same course set of interventions" (p. 30), the use of pre-intervention data would have ensured that the analysis was untangled from any effects of the intervention itself. That is, pre-intervention data may have better represented the relationship between these self-regulatory processes and learning as they exist for most undergraduate students; intervention participants may be different than their non-participating counterparts. Inspection of Table 1 in Schapiro and Livingston (2000) reveals a relatively small difference between the high and low active self-regulation groups' GPA marginal means. This group GPA difference might have been more pronounced at pre-course, which might have drastically changed their results and conclusions. For example, a larger pre-intervention GPA difference may have revealed a statistically significant main effect of active self-regulation.

Schapiro and Livingston (2002) conclude with a call for further study of "the nature of dynamic self-regulation" and "the factors that comprise, influence and stimulate dynamic process[es]" (p. 34). In light of these findings, it is clear that additional research is needed to sort out the nature of dynamic self-regulation (Iran-Nejad, 1990), as well as its relationship with active self-regulation. To this end, this study considers that which is claimed to be measured by the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992):

15

the experimental operation of Iran-Nejad's (1990) hypothetical self-regulatory control constructs. This is performed via the application of the Rasch measurement model (Bond & Fox, 2007).

FOSTERING SELF-REGULATED LEARNING (SRL)

This study was also informed by research on the teachability of those cognitive strategies employed by self-regulated learners (SRLs) as well as the development of academic self-regulation more generally in undergraduate and other students. For example, Rosenshine, Meister and Chapman (1996) reviewed 26 intervention studies with equivalent control groups that examined training in a particular cognitive strategy (i.e., generating questions to aid comprehension). The median effect size ranged between .35 and .88 (depending on the outcome measure) for training in this strategy, suggesting a practically meaningful overall treatment effect. Importantly, all five studies in which college students were trained to use this particular strategy demonstrated statistically significant effects in their favor. However, this review may be somewhat limited in its scope, as it only focused on one cognitive strategy.

Hattie, Biggs and Purdie (1996), in contrast, conducted a meta-analysis of 51 different learning skill interventions to determine their overall effectiveness. Some interventions included in the analysis specifically targeted self-regulatory processes, while others focused on study-related skills and strategies (e.g., mnemonics) or motivational factors. Importantly, all of the included studies exercised some form of statistical control. Their statistical aggregation technique revealed a practically meaningful overall weighted mean effect size of .45. However, this figure was somewhat lower (.28) for interventions with college students.

Furthermore, Hattie, Biggs and Purdie (1996) also attempted to determine which, if any, instructional characteristics were systematically related to cognitive strategy intervention effectiveness. Reviewing the literature, they first summarize previously indicated conditions for

successful strategy training, including "high and appropriate motivation, including self-efficacy and appropriate attributions (such as attributing failures to a lack of effort, and setting realistic and attainable goals)," "the strategic and contextual knowledge for doing the task," and "a teaching-learning context that supports and reinforces the strategies being taught" (p. 103). Based on their meta-analysis, they also report that strategy training was most effective when it was contextualized. For example, they recommend that strategy training should occur in the target context (e.g., writing strategies in a writing course). Furthermore, their meta-analysis suggested that "a high degree of learner activity and metacognitive awareness" (p. 131) and motivational support were associated with intervention effectiveness.

The review by Rosenshine, Meister and Chapman (1996) and the meta-analysis by Hattie, Biggs and Purdie (1996) suggest that students can benefit from training in the types of cognitive strategies employed by self-regulated learners. Hattie, Biggs and Purdie (1996) further note that metacognitive interventions, those which "focus on the self-management of learning, that is, on planning, implementing, and monitoring one's learning efforts, and on the conditional knowledge of when, where, why, and how to use particular tactics and strategies in their appropriate contexts" (p. 100), are successful. Not surprisingly, these interventions closely mirror the goals of interventions aimed at promoting self-regulated learning.

Lastly, clues regarding how to successfully promote self-regulated learning (SRL) via an intervention can also be garnered from educational practitioners who work with students rather than effect sizes. In response to an investigation of SRL in a research and writing class, Perry and Drummond (2002) suggest establishing a community of learners and a "positive, encouraging learning atmosphere" (p. 302). Further, they also point out the importance of engaging students in complex, meaningful tasks, self-monitoring and -evaluation to foster SRL.

These suggestions are consistent with other factors recommended by the literature, as well as by the social cognitive theoretical models (e.g., Zimmerman, 2000; 2002; Winne & Hadwin, 1998) in which they are grounded.

Like many educational and other interventions (see Ceci & Papierno, 2005), investigations of the effectiveness of interventions targeted at promoting cognitive strategy use have demonstrated differential effects as a function of participant characteristics. For example, Hattie, Biggs and Purdie (1996) found in their meta-analysis that these types of interventions were the most helpful for underachieving students. In contrast, the review by Rosenshine, Meister, and Chapman (1996) yielded a more mixed picture of the differential effectiveness of training in the question generation strategy they studied. They found some studies favoring lower achieving students and some favoring those who were higher achieving. Three of the five studies that focused on college students had results favoring the higher achieving students. However, the authors note that the number of studies reviewed limits their ability to defensibly argue their findings regarding the interaction between cognitive strategy training and participant characteristics such as age or achievement level. They accordingly recommend that such interactions be considered in future research.

Although training in the use of cognitive strategies can be successful, this is not always the case. For example, there is some evidence demonstrating that direct instruction in strategies coupled with simple self-monitoring fails to provide changes to study behavior. Cao and Nietfeld (2007) found that one's monitoring of his or her understanding of course material and an awareness of the difficulties one is having does not automatically lead to the application of appropriate remedial strategies. In their study, students in an Educational Psychology course covering various metacognitive strategies were asked to identify on a worksheet each day any

difficulties they were having (e.g., understanding a specific concept, application of information) and what they planned to do to improve their understanding. Using chi-square (χ^2) tests, they found that students did not appropriate their application of strategies to the type of problems they were aware they were having. That is, strategies supported by the literature were not applied systematically to meet specific learning difficulties, and participants more often relied upon common, passive rehearsal strategies (e.g., re-reading). The authors conclude that a "more intensive intervention is needed for students to not only understand but also master the recommended strategies" (p. 40).

It is important to consider that the Cao and Nietfeld (2007) study was conducted ancillary to an Educational Psychology course. This does not necessarily constitute a stand-alone intervention. As suggested by the authors, it is possible that a more directed or comprehensive intervention might be more successful in promoting the effective use of cognitive strategies. Furthermore, an intervention better grounded in the research literature regarding when interventions of this nature are effective might yield more favorable results.

The previous literature review demonstrates that the development of self-regulatory processes is possible given the right conditions. A fundamental assumption of this research is that active and dynamic self-regulation can be influenced by the intervention under study. Iran-Nejad (1990) posits that both active and dynamic self-regulation can be improved by instruction in learning strategies, clearly stating that "both sources of internal self-regulation must be the target of metacognitive instruction." He suggests that learning strategy instruction aimed at developing active self-regulation should "teach [students] to take advantage of executive internal self-regulation" (p. 592). It is assumed that the instructional conditions discussed earlier by other

scholars would similarly promote active self-regulatory abilities vis-à-vis Iran-Nejad's (1990) theory.

To develop dynamic self-regulation, Iran-Nejad (1990) purports that "academic contexts must be arranged in such a way that the dynamic or spontaneous learning approaches that worked for children before school continue to work for them during school" (p. 592). Specifically, he recommends "eradicating all signs suggesting that today's learning is memorization of isolated materials in preparation for tomorrow's learning of more authentic stuff" and "orchestrating learning opportunities to closely match authentic learning opportunities of real-world contexts" (p. 592). Furthermore, he advocates "changing one's learning intentions from those aimed at optimizing the conditions for encoding and retrieval under other-regulation to those aimed at optimizing the conditions for understanding and personal growth under self-regulation" (p. 592-593).

Tinnesz, Ahuna, and Keiner (2006) and Schapiro and Livingston (2000) contend that students enrolled in the present intervention receive overt instruction in both active and dynamic self-regulatory control. To promote active self-regulatory control, students engage in the contextualized application to mastery of concrete, prescribed techniques intended to foster the development and internalization of cognitive strategy use. It is noteworthy that this contextualization is consistent with the recommendations of Hattie, Biggs and Purdie (1996).

The extent to and the means by which Iran-Nejad's (1990) notion of dynamic self-regulation can be influenced by an intervention is less clear, but prior investigations have reported growth (e.g., Schapiro & Livinston, 2000). Schapiro and Livingston (2000) postulated that dynamic self-regulation may have been influenced by "giving students sufficient time to demonstrate their persistence, valuing…. [and] enthusiasm for unconventional ideas and

interpretations" as well as "reinforc[ing]…students' natural curiosity" (p. 33). They also suggest overtly valuing behaviors indicative of dynamic self-regulation and fostering a community of learning (i.e., peer interaction). In the present intervention, this is manifest in the explicit endorsement of behaviors[2] consistent with dynamic self-regulation both during formal instruction and 12 30-minute one-on-one meetings with an undergraduate peer monitor.

Prior evaluations of this intervention (VanZile-Tamsen, 2007; Ahuna, Tinnesz, & VanZile-Tamsen, 2009) have reported positive effects on student retention, graduation rates and quality point averages (QPAs) as a result of course participation. This research, however, centers on the effectiveness of the intervention as it relates to the biofunctional theory of self-regulation (Iran-Nejad, 1990) rather than variables of institutional interest. Crucially, this study seeks to rigorously address the question of whether growth in both active and dynamic self-regulatory control results from participation in the intervention.

Some quantitative data lend support for the success of this intervention in promoting undergraduate students' self-regulatory processes. Tinnesz, Ahuna and Keiner (2006) report on a large sample of students ($N = 680$) who took the course over four semesters. The authors measured students' active and dynamic self-regulatory processes by administering the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) at pre- and post-intervention. One-sample *t*-tests were conducted to examine if mean differences from pre- to post-intervention (i.e., growth) were statistically distinguishable from zero. Growth in both students' active and dynamic self-regulatory processes was observed (ps < .001) across all four semesters under study.

---

[2] These were assumed to be behavioral manifestations of the non-executive processes (e.g., interest) described by Iran-Nejad (1990), although the extent to which such self-regulatory processes are manifest behaviorally is unclear.

The authors also explored whether the growth observed in both active and dynamic self-regulatory processes differed by gender, year in college, class section, or race/ethnicity. Independent samples *t*-tests revealed no statistically significant gender differences (ps > .05). Similarly, separate analyses of variance (ANOVAs) conducted with growth in active and dynamic self-regulation as dependent variables revealed no statistically significant group differences (ps > .05) by year in college, class section, or race/ethnicity. That is, the growth observed was similar for all course participants. In light of other findings to the contrary (e.g., Hattie, Biggs, & Purdie, 1996) in similar interventions, this study attempts to re-visit this question using refined instrumentation.

Schapiro and Livingston (2000) also examined whether or not growth in active and dynamic self-regulation is evidenced throughout the course of this intervention. In order to also study growth by participant achievement level, they used post-intervention GPA[3] data to separate participants into two groups. Dependent samples *t*-tests were then conducted to examine self-reported change in active and dynamic self-regulation for these two achievement groups from pre- to post-intervention. They report increases (ps < .001) in both forms of self-regulation—regardless of achievement level—as a result of course participation. The authors also examined growth in dynamic self-regulation by examining changes in the quadrant status classification discussed earlier. They report that the percentage of participants classified as high dynamic increased from pre- to post-intervention. However, whether this represented a statistically significant difference was not reported.

The timing of posttest data collection in both the Tinnesz, Ahuna and Keiner (1006) and the Schapiro and Livingston (2000) studies was immediately after the intervention. To

---

[3] Pre-intervention data might have again been preferable for reasons similar to those discussed earlier.

investigate any sustained effects of the intervention, Livingston (2000) conducted a qualitative dissertation study ($N = 30$) to examine participants' perceptions of cognitive strategy instruction and its influence on their subsequent strategy use. Phenomenological interviews were conducted two semesters after course participation. Also, learning materials (e.g., notebooks, textbooks) were inspected, coded and used as a basis for discussion. Eighty percent of participants reported during the interviews that the intervention contributed to their current strategy use. In addition, inspection of participants' learning materials revealed that rehearsal (93%) and organization (87%) strategies were subsequently used by participants.

Livingston (2000) also considered participants' course grade as a proxy for participants' mastery of the prescribed techniques that were geared at fostering cognitive strategy use. Consistent with expectations, "A-grade students used a greater number of strategies more consistently than did C-grade students" (p. iv). An analysis of variance (ANOVA) revealed that students receiving an A were statistically more likely to employ organization strategies than students receiving a C. However, statistically significant differences in use of the three other self-regulatory strategies considered were not observed.

Students' motivations for taking the course were also considered in Livingston's (2000) study. In particular, she differentiated between students enrolled in the course for self-improvement versus other reasons (e.g., academic risk). Livingston (2000) hypothesized that students who were intrinsically motivated (i.e., in the self-improvement group) would be expected to "exhibit long-term strategy use and maintenance" (p. 9). However, conclusions regarding the relationship between motivation and subsequent strategy use were more nuanced. Specifically, students enrolled in the course for self-improvement subsequently used rehearsal and metacognitive self-regulation strategies more, while those with other motivations tended to

23

more often use elaboration and organization strategies. Nevertheless, analysis of variance (ANOVA) did not reveal any statistically significant differences. These failures to find significant differences may be related to the small sample size. The author calls for a larger, more representative sample in future follow-up research.

The Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1991) was also administered to participants in Livingston's (2000) study. Qualitative evidence of the use of organization, elaboration, comprehension monitoring and rehearsal strategies was compared to respective MSLQ subscales in order to triangulate using both ecologically valid and objective analytical approaches. The rehearsal and organization MLSQ subscales were found to be positively correlated with actual rehearsal ($r = .37$) and organization strategy ($r = .42$) use. In contrast, correlations between the elaboration and comprehension monitoring subscales were not statistically significant. Thus, although there were some inconsistencies in her data, Livingston's (2000) qualitative dissertation study did provide some evidence that students may acquire long-term benefits from the present intervention.

THE PRESENT STUDY

The purpose of this research was to critically examine previous findings and to address empirical questions that remain unanswered. Specifically, this study investigated the validity of inferences made from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad and Chissom, 1992) and addressed other threats to previously made causal inferences (e.g., maturation) regarding the effectiveness of the intervention discussed throughout this paper. This research thus represented an attempt to move toward estimating a defensible treatment effect of the intervention. Six research questions prompting this study are discussed next; specific research hypotheses are also presented when appropriate.

How confident can one be in inferences made from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992)? Prior investigations of this intervention have measured active and dynamic self-regulation using the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992), an experimental instrument undergoing continued development. Data from the DALI-R have been transformed and submitted to analysis in a number of ways. Tinnesz, Ahuna and Keiner (2006) computed scale score differences between pre- and post-intervention DALI-R administrations. Schapiro and Livingston (2000) used both post-intervention scale scores and also median splits to dichotomize DALI-R pre- and post-intervention scale score variables. Significantly, all of these uses of DALI-R data have begun with scale scores.

Like many measures used in research and practice, the DALI-R was developed in a manner consistent with the classical test theory (CTT) measurement approach. This approach has a number of limitations. These will be discussed thoroughly in the Method section. Most notably, scale scores yielded from measures developed using the CTT approach may violate the assumptions of the parametric inferential statistics to which they are submitted (Wright & Linacre, 1989). In other words, DALI-R and other CTT measure scale score data are not necessarily on the interval scale required by these analyses. The practical problems associated with the use of ordinal measures in the estimation of treatment effects have been discussed previously by Merbitz, Morris, and Grip (1989). The authors note that "misuse of [ordinal] scales…may mask ineffective treatment procedures and hide efficient procedures" (p. 308). Such errors of inference are clearly problematic.

Furthermore, the reader will recall that both Iran-Nejad and Chissom (1992) and Schapiro and Livingston (2000) studies found that active self-regulation was not statistically related to

achievement outcomes after accounting for the influence of dynamic self-regulation. Of note to this study, these findings contrasted the preponderance of evidence that strategic self-regulatory behavior does contribute to learning. There were various hypotheses presented as to why these unexpected findings were observed; they mostly centered on the role of dynamic self-regulation as requisite to active self-regulatory behavior. Alternately, the author hypothesized earlier that the instrumentation used to measure these constructs may have been problematic.

This study responds to these unexpected findings and alternative hypotheses by submitting DALI-R data to the Rasch measurement model (Bond & Fox, 2007). The refinement of the DALI-R via the application of this model in this study at the very least represents an improvement to the outcome measures employed by placing scores on an approximately interval scale. Additionally, submission of data to Rasch analysis and the analytic procedures involved in doing so also provided some evidence as to the validity of inferences that can be made using the DALI-R. For example, this study sheds light on the nature of dynamic self-regulation and whether it represents a unidimensional construct on which individuals differ—an important theoretical matter.

Do participants report increased active, dynamic self-regulatory control following the intervention? The primary research question asked whether growth in active and dynamic self-regulation is observed throughout the course of the intervention. This question has been addressed by prior research (Tinnesz, Ahuna, & Keiner, 2006; Schapiro & Livingston, 2000), although the inferences drawn previously may suffer from limitations related to the influence of unmeasured variables, as well other threats to internal validity related to the instrumentation used and maturation. The question, therefore, was re-addressed in this study using refined instrumentation. Subsequent research questions attempted to combat the other internal validity

threats. Given the findings of Tinnesz, Ahuna and Keiner (2006), it was hypothesized that statistically significant growth in active self-regulation would be evidenced throughout the course of the intervention. It was unclear at the outset of the analyses whether growth in dynamic self-regulation would be observed using refined instrumentation.

Does observed growth differ by instructor? Since parts of this intervention are implemented in the form of an undergraduate course, one might expect there to be differences in observed growth as a function of the instructor who is delivering the content. The data employed for this study were collected from students enrolled in course sections taught by two different instructors. Consequently, this study investigates whether such an instructor effect is present. Importantly, evidence that growth differs by instructor might lend support for the inference that the intervention itself was the cause of any observed growth. While this question has been addressed by Tinnesz, Ahuna & Keiner (2006), it was again re-addressed in this study using refined instrumentation. It was hypothesized that there would not be instructor differences in active and dynamic self-regulation growth, as no anecdotal or other evidence suggested that this would be the case.

Does observed growth differ by participant motivation? The influence of motivational factors on academic learning permeates the educational psychology literature. For example, Uguroglu and Walberg (1979) summarized 232 Pearson product-moment correlation coefficients describing the relationship between measures of motivation and achievement and reported a mean correlation of .338; motivation accounted for approximately 11.4% of the variance in the achievement measures. Although they did not include correlations from investigations with college students, they reported that the magnitude of the relationship grows with increases in age. It was assumed that this relationship is valid for the population targeted by the intervention

27

presently under investigation, as many of the students enrolled in it are only slightly older than the oldest students in Uguroglu and Walberg's (1979) study.

Prior quantitative investigations of the effectiveness of this intervention have not considered motivational factors. Such unmeasured variables threaten the internal validity of causal inferences made via observational research (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007). For example, it could be argued that students enrolled in the intervention demonstrate growth in active and dynamic self-regulation as a result of motivational factors rather than as an effect of the intervention itself. This research addresses whether any observed growth varies as a function of the motivations of its participants. More specifically, this research accounted for data on why students took the course. Given the importance of motivational factors in learning (e.g., Uguroglu & Walberg, 1979), and self-regulated learning (SRL) in particular (e.g., Zimmerman 2000; 2002; Winne & Hadwin, 1998), it was hypothesized that growth in active self-regulation would be moderated by participant motivations. It was again unclear at the outset whether any observed growth in dynamic self-regulation would be moderated by participant motivation.

Does observed growth differ by the extent to which participants receive the intervention as intended? The threat of maturation to the internal validity of longitudinal treatment investigations without a comparison group is a paramount concern (Fraenkel & Wallen, 2006). In this case, the maturation threat casts doubt on whether reported growth (Tinnesz, Ahuna & Keiner, 2006; Schapiro & Livingston, 2000) would have actually been observed without participants having undergone the intervention. Specifically, it could be argued that students' active and dynamic self-regulatory processes could develop naturally over the course of a semester.

The research attempts to combat this maturation threat. Prior scholarly (i.e., non-institutional) quantitative evaluations of the effectiveness of this intervention in promoting self-regulatory processes (Tinnesz, Ahuna & Keiner, 2006; Schapiro & Livingston, 2000) have treated all participants as a homogeneous group with respect to their receipt of the intervention. However, one might expect to observe differential growth as a function of the extent to which an intervention was actually received. VanZile-Tamsen (2007) lends some support to the hypothesis that differential gains might be observed as a function of success in the course. She demonstrated that both retention and graduation rates as well as quality point average (QPA) gains varied as a function of course grade. Ahuna, Tinnesz, and VanZile-Tamsen (2009) demonstrated a similar differential effect on retention and graduation rates as a function of course grade with a larger sample. Furthermore, Livingston's (2000) qualitative dissertation study reviewed above revealed differences in subsequent cognitive strategy use as a function of course grade.

The present study attempts to shed light on whether it is the intervention itself that can be claimed to have caused observed growth. To provide evidence of the validity of this causal inference, the degree to which participants received the intervention as intended was accounted for in the analyses. It was hypothesized that growth in active self-regulation would increase in concert with increases in various indicators of intervention exposure. It was again unclear at the outset whether growth in dynamic self-regulation would vary as a function of intervention exposure.

Does observed growth differ by participant race/ethnicity or gender? Given variability in college graduation rates along racial/ethnic and gender lines and the reported differential effectiveness of cognitive strategy training programs discussed earlier, it might also be worthwhile to investigate this intervention's effectiveness in terms of the races/ethnicities and

genders, of its participants. The demonstration of differential intervention effectiveness favoring more disadvantaged students might provide evidence for its utility in helping to close gaps in American college graduation rates.

Support for the hypothesis that differential gains might be observed is mixed. For example, Tinnesz, Ahuna and Keiner (2006) reported equivalent active and dynamic self-regulatory control growth by various participant characteristics (e.g., race/ethnicity). In contrast, VanZile-Tamsen (2007) reported an interaction between certain subject variables and course completion when examining its effect on quality point averages (QPAs). In particular, she found that students participating in the Equal Opportunity Program (EOP) seem to experience a larger quality point average gain as a function of course participation than non-EOP participants. This research question was re-addressed in this study using refined instrumentation.

It was likewise possible that this intervention might promote differential growth favoring more advantaged groups. This issue is taken up by Ceci and Papierno (2005), who note examples of interventions that have actually exacerbated extant gaps. The intervention under study is what they refer to as a *universalized intervention*; such interventions are not targeted at particular high-risk or disadvantaged groups. In their article, they reference research on universalized cognitive strategy interventions by Scruggs and Mastropieri (1988), and others, where differential effects favoring more advantaged groups were observed. For example, Scruggs and Mastropieri (1988) report an experiment in which gifted and non-gifted (regular) students were randomly assigned to either a control or one of three mnemonic strategy conditions. Participants were instructed to independently learn material—the hardness ratings of various North American minerals—presented on cards in a manner consistent with the condition to which they were assigned. Statistically significant differences between the control and mnemonic strategy

conditions were observed only for gifted students. Additionally, only gifted students transferred the strategy to a novel learning task. In light of the mixed findings reported previously, it was unclear whether all participants would experience similar growth patterns in active self-regulation throughout the course of the intervention. There was again no hypothesis regarding differential growth in dynamic self-regulation.

CHAPTER THREE: METHOD

This research was conducted in two phases. First, the Rasch measurement model (Bond & Fox, 2007) was applied to data collected previously to address the first research question. This process is described in Analytical Approach. Second, after some refinements were made to the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992), person measures for the two primary variables of interest that approximate an interval scale were estimated. These yielded person measures were then analyzed with other data to answer the five other research questions.

DATA SOURCE

No new data were collected for this research. Data for this study were collected during the spring of 2009 for programmatic purposes. These analyses thus constituted a secondary analysis. Participants provided informed consent for their data to be used as part of an on-going, programmatic "research" project. This research was exempt from review by the Social and Behavioral Sciences Institutional Review Board (SBSIRB) at the University at Buffalo, The State University of New York. Data from disparate sources were completely de-identified after their merger; they were maintained, analyzed and are reported here anonymously.

DALI-R. The Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) was administered both at the beginning and end of the intervention for programmatic assessment purposes. It was administered as part of a larger inventory containing additional items more closely mapping onto course content and of administrative interest. There were 38 and 37 non-DALI-R items included in the complete inventory at pre- and post-intervention, respectively. Examples of non-DALI-R inventory items include *I consciously use a*

variety of strategies to study and *I consciously try to write a winning script for myself*. Only DALI-R items were considered in this study.

Iran-Nejad and Chissom (1992) assert the Dynamic and Active Learning Inventory Revised (DALI-R) to measure the active and dynamic self-regulation underlying the learning processes of undergraduate students. Regarding the validity of inferences that would be made from their instrument, its authors report that the DALI-R was "carefully derived from theory and research in cognitive educational psychology on the relationship between sources of self-regulation and learning processes" (p. 130). Specifically, the instrument was developed on the basis of Iran-Nejad's (1990) two-source biofunctional theory of self-regulation. Thus, item development is consistent with that called for by the Rasch model, which requires that items are developed with respect to theory (Wilson, 2005). It was assumed that the content of the DALI-R items are representative of Iran-Nejad's (1990) biofunctional theory of self-regulation.

A modified version of the DALI-R was used in this study. The item *When I have an exciting new idea, I carefully consider how it will play out, rather than acting upon it quickly* was not administered. Six items were reverse scored before analysis. The original DALI-R featured seven negatively worded dynamic self-regulated items; one was administered with a positive wording in this context (i.e., *It's hard for me to picture myself achieving the goals I have in mind* on the DALI-R was administered *When I have a goal in mind, I can picture myself achieving it*). The response format for all items was a 7-point rating scale anchored by "never" (1), "sometimes" (3), "often" (5), and "always" (7). Internal consistency reliability (α) during development (*N* = 99) was .65, and item-total correlations ranged from .28-.69 (Iran-Nejad & Chissom, 1992). Ten DALI-R items were intended to measure rehearsal, organization, elaboration and comprehension monitoring—active self-regulatory strategies (see Appendix A).

33

When employed in practice, internal consistency reliabilities for the active self-regulation items (α) were .81 and .83 at pre- and post-intervention, respectively (Schapiro & Livingston, 2000).

Twenty-one DALI-R items were intended to measure "alertful attention, curiosity, postdiction, reflective metacognition, and other related dynamic factors such as suspense, anxiety and interest" (p. 130; see Appendix B). The breadth in item content is noteworthy and prompted the author's hypothesis that there may be sub-dimensions captured by these DALI-R dynamic self-regulation items. Schapiro and Livingston (2000) reported internal consistency reliabilities (α) for the dynamic items at .77 and .78 at pre- and post-intervention, respectively. Prior work (Iran-Nejad & Chissom, 1992; Livingston & Shapiro, 2000) has administered these items to undergraduate samples. It was assumed that students' self-reports of their active and dynamic self-regulatory processes were valid.

Demographics. Demographic data were obtained from the University at Buffalo, The State University of New York. Information regarding the race/ethnicity and gender of each participant was considered in the analyses. These data permitted investigation of the effects of the intervention for many of those subgroups discussed earlier. Specifically, Black, Asian or Pacific Islander and White students were represented in the analyses[4].

Descriptive analyses indicate that the sample was roughly representative of the target population. For example, the sample has slightly more males (51.6%) than females, which is similar to the total full-time undergraduate degree-seeking population of males (53.9%) at the institution at which these data were collected. The percentages in the sample and target population, respectively, of Asian or Pacific Islander (11.3% and 11.8%) and Hispanic (5.4% and

---

[4] Native American and Puerto Rican students were not represented in the sample. Hispanic students were not represented in the analyses owing to insufficient sampling of this group. Participants either indicating "Other" or for which race/ethnicity information was not available were also excluded from the analyses.

4.5%) students are also quite similar. In contrast, Black students are over-represented in the sample (11.8% versus 8.9%), while White students are under-represented (60.8% versus 74.6%) (Office of Academic Planning & Budget, 2008). Sample characteristics are presented in Table 1.

Motivation. To roughly capture motivational factors, data from a selected-response pre-intervention assessment item asking participants to report their primary reason for taking the course was used. There were five response categories: "It is mandatory"; "My advisor recommended it"; "I needed to improve my GPA"; I needed an elective"; and "The course content seemed interesting." These categories were developed by the author and a colleague to collect information for programmatic purposes. Therefore, they were not necessarily a priori categories aligning with current theoretical understandings of motivation. While a post hoc attempt is made later to interpret some of these categories with respect to theory, the limitations of this indicator were recognized at the outset. The "It is mandatory" response category was endorsed by an insufficient amount of participants ($n = 2$) and was excluded from analysis. The "I needed to improve my GPA" and "The course content seemed interesting" response categories were endorsed by at least twice as many participants as each of the remaining two. Response category frequencies are presented in Table 1.

Intervention exposure. The extent to which participants received the intended intervention was estimated by accounting for various elements of intervention participation and progress that were quantified and recorded for programmatic purposes. This study, therefore, considers the extent to which: (1) direct instruction in course content was received (exposure); (2) weekly, facilitated self-assessment meetings were attended (attendance); (3) strategies were employed in students' other classes (commitment); (4) strategies were mastered in relation to a set of specified criteria (practice); (5) and strategies were implemented correctly by course end

(final review). The course elements that were taken into consideration were assumed to serve as crude estimates (i.e., proxies) of their participation and engagement in the intervention (i.e., intervention exposure) as it relates to Iran-Nejad's (1990) theory.

ANALYTICAL APPROACH

Rasch analysis. As discussed earlier, the Rasch measurement model was applied in this study to investigate the validity of inferences that can be drawn from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) and to yield interval active and dynamic self-regulation person measures to assist in the answering of the other research questions. The author's rationale for the application of the Rasch model in this study is thoroughly discussed next. Later, the results of both previous and new psychometric analyses are presented.

The application of the Rasch measurement model constitutes a contribution to the measurement of these constructs for a few reasons, which correspond with the advantages of Rasch measurement (Bond & Fox, 2007). Like the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992), instruments are often developed according to the principles of classical test theory (CTT). In CTT, a person's observed score on some latent trait is considered to be a summation of their hypothetical (and unobservable) true score, and error. Error can include the influence of irrelevant factors such as testing context, test-taker motivation, anxiety, or characteristics of the instrument itself (e.g., item wording). Since CTT supposes two estimable elements necessary for the measurement of a mental construct, and the observed score is quantified, the estimation of error, then, becomes the focus of measurement. This second value (i.e., error) is used in conjunction with the first (i.e., observed score) in order to determine an estimate of the third, one's true hypothetical score. In CTT, error, or the standard

36

error of measurement (SEM), is estimated using an indicator of the reliability of the instrument, among other things. After the SEM is estimated, a confidence interval can be calculated around a person's observed score to yield—with some degree of confidence—that person's hypothetical true score.

The classical test theory (CTT) approach to measurement has a number of limitations. First, these operations are based on total scores and do not consider the characteristics of individual items (which could, in theory, be differentially difficult to endorse or error-laden). Next, the SEM is held constant for all test-takers, which is problematic given the vast literature on many relevant individual differences and unavoidable variability in the conditions under which tests are administered. Third, true and observed scores are dependent, such that a different test could yield a different "true" score estimate. Lastly, reliability, which is used to estimate error, is a function of test length such that longer tests are typically more reliable, which may not always be administratively efficient—or economical.

An important consequence of the limitations of classical test theory (CTT) involves the type of scale upon which scores are placed. Measures are often assumed to yield scores on an interval scale when developed using CTT, which can be an invalid assumption. Since scores are based on the simple summation of raw scores, items that require different cognitive demands or are differentially endorsable are weighted equally in the computation of a total score. This is problematic since interval (or ratio) scales are assumed by some inferential (i.e., parametric) statistics used in the drawing of generalizations from samples to populations. Importantly, the use of ordinal scales increases measurement error, which decreases statistical power.

There are additional limitations of classical test theory (CTT) when dealing with polytomous (i.e., rating scale) data. Notably, response category differences are not necessarily

identical across all items. Also, the "agree" category has generally been observed to be larger than other categories when employing a Likert-type scale response format. This phenomenon has been termed the acquiescence response bias (Knowles & Nathan, 1997). However, the CTT approach to measurement treats all response category thresholds as equivalent, which can be problematic. Fortunately, these issues can be handled by the Rasch approach to measurement (Bond & Fox, 2007).

The Rasch approach to measurement considers two parameters—item difficulty and person ability—both of which are measured on the same hypothetical scale. Rasch measurement is based on conjoint measurement, which involves predicting the probability that one answers a given item correctly by modeling it as an interaction between its difficulty and the his or her ability level (Bond & Fox, 2007). The Rasch measurement model assumes that the object of measurement is a unidimensional, linear trait which can be measured on an interval scale. Consequently, the scale upon which both person and item measures are placed using the Rasch approach is in log-odds probability units (logits). Significantly, this transformation to logits changes the scale from ordinal to interval, eliminating this limitation of classical test theory (CTT).

The reader will recall that the classical test theory (CTT) approach relies on total scores. In contrast, the Rasch measurement approach considers items individually, and thus does not rely on total scores. Therefore, in CTT, there is only one error estimate for a set of items, while in Rasch, an error estimate is provided for each item (and person). This feature of Rasch overcomes another major limitation of CTT. Lastly, in Rasch measurement, persons and items are independent, a concept known as invariance (Bond & Fox, 2007). This feature is uncharacteristic of measures developed in a manner consistent with CTT.

Because the Rasch model assumes that the focus of measurement is a linear, latent construct, submitting data from an instrument to Rasch analysis can provide evidence for the validity of inferences that are made using that instrument.[5] For example, Rasch analysis provides model fit statistics for items that provide a measure of the extent to which they conform with the expectations of the model. Additionally, principle components factor analysis (PCA) of the standardized model residuals (i.e., discrepancies between model predictions and empirical data) can provide evidence of the existence of any sub-dimensions captured by an instrument (Smith, 1996). The Rasch model was applied in this manner in this study using Bond&FoxSteps software (Bond & Fox, 2007).

It was assumed that both active and dynamic self-regulation represent unidimensional constructs suitable for the application of the Rasch model (Bond & Fox, 2007; Wilson, 2005). Preliminary support for the validity of this assumption was lent by Iran-Nejad's (1990) example descriptions of learners who make use of their active and dynamic self-regulatory control sources to varying degrees. Additionally, a previous application of the Rasch model to the active self-regulation items of the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) items described later supported this contention.

Statistical analysis. Both descriptive and inferential statistical analyses were conducted using SPSS software. To answer the question of whether growth in either active or dynamic self-regulation was observed throughout the course of the intervention, two one-sample *t*-tests were conducted to determine whether observed growth was statistically distinguishable from zero. To answer the research questions related to an instructor effect, differential effectiveness and

---

[5] The instrument must similarly be assumed to measure a linear construct. The Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) is posited to separately measure two linear constructs: active and dynamic self-regulation.

motivational factors, separate multivariate analyses of variance (MANOVAs) were conducted with active and dynamic self-regulatory control growth as a multivariate set of dependent variables. Separate analyses were conducted for each independent variable (e.g., race/ethnicity) because some cell sizes precluded sufficient statistical power to include all variables in the same model. To explain any observed growth as a function of intervention exposure, univariate multiple regression analyses were conducted to simultaneously predict growth in both active and dynamic self-regulation from each of five process indicators. Post hoc contrasts were also conducted secondary to omnibus inferential analyses. Strength-of-effect measures were also calculated for all statistically significant relationships.

Groups with insufficient counts were excluded from the analyses where appropriate. Two participants were excluded from the analyses owing to the fact that their post-intervention active self-regulation scale scores were artificially imputed by Bond&FoxSteps due to insufficient data. Both participants' post-intervention scale and growth scores were clear outliers at the high end of the distribution. Bond&FoxSteps did not impute any dynamic self-regulation scores. Results from the analyses conducted after removing these two participants are reported in this manuscript to ensure that treatment effects were not inflated. Substantive results were not changed by removing these two participants.

CHAPTER FOUR: RESULTS

RASCH ANALYSIS

The Rasch model was applied to ten DALI-R active self-regulation items for a course project during the spring of 2009. This project was concerned with calibration (i.e., ordering items and identifying their relative difficulties) and providing preliminary evidence for the validity of inferences made using these items. Findings are described next in conjunction with rationales for the analytical tools that were used. These results are then synthesized with findings from new applications of the Rasch model performed for this study.

Initially, the active self-regulation data were submitted to the Rasch measurement model for polytomous data, otherwise known as the rating scale model (RSM; Andrich, 1978). Important in the context of the present study, the RSM assumes the same category structure across all items. However, since all the items were rated by participants on the same 7-point ordinal rating scale without an explicit theoretical rationale for doing so, the data were also submitted to the partial credit Rasch model[6] (PCM; Wright & Masters, 1982). This allowed for category thresholds to vary across items and for the examination of their category structures individually.

In general, the data better fit the partial credit Rasch model (Masters, 1982), indicating that all items may not have the same optimal item design. However, since the rating scale model is preferable in that it is more parsimonious and would involve participants uniformly responding to all items, it was retained. Moreover, overall item measure instrument quality indicators (described later) were acceptably high when data were submitted to both of these models.

---

[6] It was recognized that model selection is not ordinarily performed on the basis of post hoc item fit (X. Liu, personal communication, May 12, 2009).

After model selection, each item's point-biserial correlation with the measure was estimated to ensure that it was making a contribution to the instrument. According to Mok, Cheong, Moore and Kennedy (2006), items with point-biserial correlations with the measure larger than .4 are generally not indicative of sub-dimensions. Thus, only positive, moderate correlations with the measure are acceptable. From the outset of the analyses, all items were moderately and positively correlated with the measure (.44-.64).

Infit and outfit mean square and standardized item fit statistics were consulted to determine the extent to which each item conformed to the expectations of the Rasch model. Non-fitting items were removed (Smith, 2001) if their mean square and standardized infit and outfit fit statistics were not within acceptable limits (Bond & Fox, 2007). One item (i.e., *When I study the textbook or my lecture notes, I underline or highlight important sentences*) was removed for exceeding acceptable limits for both its infit and outfit mean square and standardized item fit statistics (Bond & Fox, 2007; Adams & Khoo, 1996). This item may not have been measuring this same construct as were the others (Smith, 2001). Reeves (2009) hypothesized that this item may not conform to the pattern expected by the Rasch model because it represents a behavior that is quite common among students, regardless of their level of the target active self-regulatory construct. In contrast, infit and outfit mean square and standardized fit statistics for the nine other items were acceptable throughout the analyses.

A major aspect of instrument development when using the rating scale model is category structure analysis. Specifically, this phase of instrument development concerns whether participants are afforded enough response options or if there are too many. Linacre (2002) outlined a set of criteria for the optimization of rating scale effectiveness. First, category response counts for each item were inspected to ensure that each response category was used for

each item and that there was a minimum of 10 observations in each category for each item. Insufficient response counts were observed for some items' response categories. Next, the uniformity of observations across categories was examined. The normal distribution of responses across categories that was observed was less problematic than others.

Because multiple difficulties describe each item when using the rating scale model, category difficulty increases should represent, on average, higher abilities. To verify this, item category difficulties were explored to ensure that they increased monotonically (Linacre, 2002). This was examined both overall and by item, although overall acceptability was considered more important. Non-monotonic increases were initially observed for some items. Also, threshold step calibrations, or difficulty estimates for choosing one response category over another, were examined for monotonic increases of at least 1.4 but less than 5 logits, as recommended by Linacre (2002). Threshold step calibrations did not initially meet this criterion. Category fit statistics, using the criterion of less than 2.0 for each category's respective outfit mean square fit statistic, were also examined. Outfit mean square fit statistic values greater than 2.0 indicate unacceptable levels of noise (Linacre, 2002). Overall, category fit statistics were acceptable. Point-biserial correlations between each rating scale category and the instrument were also inspected. Some non-monotonic increases with increases in response category, in terms of their point-biserial correlations with the measure, were observed. Category response curves were also examined for the appearance of distinct peaks (Linacre, 2002). This visually demonstrates whether some categories are subsumed by others, indicating an excess of necessary response categories. Inspection of these curves suggested that respondents were provided with an excess of response categories.

After investigating the data with respect to these criteria, a category re-structuring (i.e., reducing the number of response categories) was ultimately performed for rating scale optimization (Linacre, 2002). Some original categories were collapsed with adjacent ones to ensure sufficient category response counts. Specifically, the 7-point rating scale structure was reduced to a 3-point rating scale, with categories one and two, categories three, four and five, and categories six and seven being merged. Collapsing the response categories was expected to remedy any problems observed. Most non-monotonic increases in response category difficulties and category point-biserial correlations with the measure were remedied by this particular re-structuring. Inspection of category response curves after re-structuring also revealed distinct peaks. Category fit statistics were also improved by this particular re-structuring. Lastly, item standard errors were inspected. These provide information regarding the extent to which one can be confident in the various item difficulty estimates. Standard errors for all items increased from .04 to .12 with category re-structuring, but were still acceptably small after this modification.

To further examine dimensionality of measurement (Smith, 1996) beyond the inspection of item fit statistics, standardized model residuals were submitted to a principle components factor analysis (PCA). Item loadings greater than .3 and less than -.3 on resultant factors (Tabachnick & Fidell, 1983) with eigenvalues larger than 1.4 (Smith, 1996) were considered indicative of possible sub-dimensions. PCA revealed only one factor with an eigenvalue larger than 1.4 (1.7). Six items loaded on this possible sub-dimension. Interestingly, those items with positive and negative factor loadings were the most and least difficult items, respectively. Reeves (2009) speculated that this possible sub-dimension could be related to motivational factors.

Data was also represented graphically by examining the person-item map and a bubble chart plotting items and persons on the same hypothetical scale. For example, the person-item

map was inspected to explore the presence of item redundancies. There were no redundancies. The distribution of persons approximated a normal curve with a mean of -.22 logits. Thus, although the items were generally well-targeted for the sample, the items tended to be somewhat more difficult, on average, than persons. There were also gaps in item coverage at the higher and especially the lower levels of person ability.

There was also a restricted range in the difficulty of the progression being measured initially (1.96 logits). This was not surprising given the measurement approach through which these DALI-R items were developed. The original approach to instrument development was consistent with classical test theory (CTT), and included those that most discriminated among persons (i.e., mean level difficulty). However, the range in item difficulty increased to a somewhat more acceptable 2.95 logits after the aforementioned changes were made to the instrument.

To examine overall instrument quality, item reliability and separation indices were estimated and referenced during the various phases of analysis. Item reliability provides information about the consistency of item placement in terms of its difficulty amongst the other items. The item separation index provides information as to the spread of the items, and the differentiation of items by the measure. More specifically, the item separation index is a ratio of the variance accounted for by the Rasch model to error variance (Wilson, 2005). For the separation index, a value greater than 2.0 is typically considered acceptable. For reliability, a value of .80 is typically acceptable (Bond & Fox, 2007; Wilson, 2005).

Person reliability and separation indices were also estimated and referenced during the analyses to examine overall instrument quality. The person separation index provides an estimate of the differentiation of persons by the measure. Person reliability is analogous to the indicator of

internal consistency reliability used in the classical test theory (CTT) measurement approach. Similar to item reliability and separation, person separation greater than 2.0 and person reliability greater than .80 are typically considered acceptable (Bond & Fox, 2007; Wilson, 2005).

Both changes to the instrument (i.e., removal of the non-fitting item, category re-structuring) were examined with respect to these indicators of overall instrument quality. These indices were expected to increase with these changes. Satisfactory differentiation of the items by the measure (i.e., item separation) was observed at the outset and improved with the removal of the misfitting item. Similarly, item reliability was acceptable from the outset of analysis; no change was observed with the removal of the misfitting item. In contrast, decrements were observed in both item separation and reliability after the category re-structuring. However, these changes did not move them away from acceptable levels.

The story of the person summary statistics during these analyses was somewhat different. Only person reliability, but not separation, was acceptable initially. The removal of the misfitting item slightly decreased estimates of both person reliability and separation. Similarly, category re-structuring resulted in further decrements in both person summary statistics. These changes ultimately resulted in estimates of person separation and reliability that were outside acceptable limits. Although it was recognized that caution should be exercised if instrument quality indicators decrease with item removal or category re-structuring (X. Liu, personal communication, n.d.), the revised category structure was retained to meet the criteria outlined by Linacre (2002).

In summary, inadequate person summary statistics were still observed after conducting changes intended to improve overall instrument quality. It was recognized at the outset that the ten items included in this analysis may not be sufficient for a good instrument (X. Liu, personal

communication, March 26, 2009). Thus, the inadequate person summary statistics were not surprising. They were likely due to the small number of items. Barring these problems and in recognition of the iterative nature of instrument development, the results of these analyses (i.e., item fit, reliability and separation) supported the contention that active self-regulation was a measurable, unidimensional construct worthy of further study (Reeves, 2009).

This study builds on these earlier analyses by applying the Rasch model to Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) data collected from a new sample. The Rasch psychometric analyses reported next are also the first to consider dynamic self-regulation DALI-R items. These new analyses were aimed at addressing the question of the extent to which one can be confident in inferences regarding active and dynamic self-regulation made from the DALI-R (Iran-Nejad, 1990). These new analyses were conducted in a manner consistent with those just described earlier.

Both active and dynamic self-regulation DALI-R items were analyzed first simultaneously. The purpose of this combined analysis was to examine whether Iran-Nejad's (1990) notions of active and dynamic self-regulation are not actually a single self-regulatory construct. For this analysis, initial and final status DALI-R data from 186 intervention participants were submitted to Rasch analysis, representing each person twice in the same analysis. The sample size (N) for this psychometric analysis thus exceeded the traditional rule of thumb of at least five participants per item. Analyses presented later also included sufficient Ns.

To investigate whether active and dynamic self-regulation actually represent a single construct, two methods were used to examine dimensionality of measurement. First, infit and outfit mean square and standardized item fit statistics were calculated by Bond&FoxSteps (Bond & Fox, 2007). In total, 16 of the 31 items were outside acceptable limits according to at least two

47

standardized fit statistics. Four had fit previously in the earlier application of the Rasch model to the DALI-R active self-regulation items. The active self-regulation item that had been removed previously was again not behaving appropriately. The other 11 were DALI-R dynamic self-regulation items.

Second, a principal components factor analysis (PCA) of the standardized model residuals was conducted. The PCA revealed five factors with considerable eigenvalues by the criterion discussed earlier. The reader will recall that the PCA conducted previously with the active self-regulation items only yielded one possible sub-dimension. The presence of five possible sub-dimensions in this combined set of items suggested that assumed dimensionality of measurement may be invalid in this case.

Because of the large share of misfitting items and the five possible sub-dimensions indicated by principal components factor analysis (PCA), these analyses did not lend support to the hypothesis that active and dynamic self-regulation represent a single construct. Moreover, the range in difficulty for the 31 items submitted to analysis was only 1.91 logits, which was less than the range initially observed with just ten active self-regulation items. If the combined set of items were measuring the same thing as those ten previously submitted to Rasch analysis, one might have expected the range in item difficulty to increase with more items. This restricted range in item difficulty also cast doubt on the hypothesis that active and dynamic self-regulation represent a single construct.

Consequently, the Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992) active and dynamic self-regulation items were submitted to separate Rasch psychometric analyses. These analyses were aimed at substantiating Iran-Nejad's (1990) theoretical account of active and dynamic self-regulation as separate, linear constructs.

Importantly, these analyses could provide empirical evidence for the validity of inferences made using this instrument. The results of these analyses are presented next and then synthesized with those found previously.

Active self-regulation. Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992) active self-regulation data were submitted to the same analytic procedures discussed earlier. Not surprisingly, the results of this new Rasch analysis of the active self-regulation items were quite similar to those obtained earlier. Those changes possible during one round of data collection were conducted. Specially, misfitting items were removed and revisions were made to the rating scale structure. These changes ultimately resulted in person measures for active self-regulation that approximated an interval scale. Results of a principal components factor analysis (PCA) of the standardized residuals and its implications are also presented next.

One Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992) active self-regulation item was removed for exceeding acceptable limits in both its infit and outfit standardized fit statistics. This item was the same as that excluded during the previous analysis. Two other items' infit standardized fit statistics were outside acceptable limits, but these items were retained to ensure content coverage. Additionally, these items' infit and outfit mean square and standardized fit statistics were within acceptable limits previously with a larger sample. Item statistics are presented in Table 2.

The same category re-structuring performed previously was applied to these new Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992) active self-regulation data after consideration of Linacre's (2002) criteria. As expected, the revision of category structure improved insufficient response category counts, non-monotonic increases in response category difficulties, as well as other important aspects of rating scale structure outlined

by Linacre (2002). Distinct peaks were furthermore evident after but not before category re-structuring (see Figures 1 and 2). Altering category structure did not adversely affect item fit for the active self-regulation items. The revised rating scale structure is summarized in Table 3.

After removing the misfitting item and conducting a category re-structuring, principal components factor analysis (PCA) of the standardized model residuals revealed two factors with substantial eigenvalues. The first factor had an eigenvalue of 1.6 and explained 8.9% of the variance in the observations. Six items had considerable loadings on this factor (see Figure 3). It is notable that the same six items loaded on a factor identified in the previous analysis. The second factor had an eigenvalue of 1.5 and explained 8.6% of the variance in the observations. Six items also had considerable loadings on this factor (see Figure 4). The items loading on this second factor were not the same as those loading on the other, although there was some necessary overlap.

Consideration of person and item reliabilities and separation indices throughout the various stages of this psychometric analysis reveal a similar story to that reported earlier (see Table 4). Both person[7] and item summary statistics were acceptable by conventions at the outset of analysis. However, the removal of the misfitting item and the revision of category structure brought person (but not item) summary statistics outside of acceptable limits. This pattern was similarly seen during previous analyses. Performing these changes independently indicated that these decrements were owed primarily to the category re-structuring as opposed to item removal. However, the category re-structuring was again retained to meet the criteria for rating scale

---

[7] Only person reliability was acceptable initially during the previous application of the Rasch model the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) active self-regulation items (Reeves, 2009).

optimization outlined by Linacre (2002). Notably, the category re-structuring did have the favorable impact of doubling the range in difficulty seen across these items.

Inspection of the person-item map (see Figure 5) and a bubble chart plotting persons and items (see Figure 6) on the same hypothetical scale also provided some information about overall instrument quality for these nine remaining active self-regulation items. These figures generally indicate that some gaps in item difficulty coverage existed, especially near those individuals with the highest and lowest estimated active self-regulation scores. The implications of these gaps will be discussed in greater detail later.

Dynamic self-regulation. The dynamic self-regulation items from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) were submitted to the same set of Rasch analyses. Seven items were removed[8] for exceeding acceptable limits in at least both their infit and outfit standardized fit statistics. One remaining item's infit standardized fit statistic was also outside acceptable limits, but this item was retained. Final dynamic self-regulation item statistics are presented in Table 5.

The same category re-structuring performed for the DALI-R active self-regulation data was performed with the dynamic self-regulation items after consideration of Linacre's (2002) criteria. As expected, the revision of category structure improved insufficient response category counts, non-monotonic increases in response category difficulties, and other important aspects of rating scale structure outlined by Linacre (2002). Distinct peaks, furthermore, were evident after

---

[8] K. Ahuna (personal communication, August 4, 2009) noted that one misfitting (and removed) item ("I find it difficult to keep my mind on the topic I am studying") was very similar to another ("I have a hard time concentrating when I am studying") that behaved in accordance with the model and was retained. Despite this perhaps unexpected difference in model fit, removing one of these items had the favorable effect of eliminating an item redundancy (Bond & Fox, 2007).

but not before category re-structuring (see Figures 7 and 8). Altering category structure did not adversely affect item fit for the dynamic self-regulation items. The revised rating scale structure is summarized in Table 6.

After removing the misfitting items and conducting a category re-structuring, principal components factor analysis (PCA) of the standardized model residuals revealed two factors with substantial eigenvalues. The first factor had an eigenvalue of 2.9 and explained 6.5% of the variance in the observations. Ten items had considerable loadings on this factor (see Figure 9). The second factor had an eigenvalue of 1.5 and explained 3.4% of the variance in the observations. Six items had considerable loadings on this factor (see Figure 10). The items loading on this second factor were not entirely the same as those loading on the other, although some overlap existed.

Consideration of person and item reliabilities and separation indices throughout the various stages of this psychometric analysis reveal a similar story to that reported twice earlier (see Table 7). Both person and item summary statistics were acceptable by conventions at the outset of analysis. However, the removal of the misfitting items and the revision of category structure brought person (but not item) summary statistics outside of acceptable limits. Performing these changes independently indicated that these decrements were owed primarily to the removal of misfitting items as opposed to the revision of category structure. This was the opposite pattern of that seen with analysis of the active self-regulation items for this study. However, this is likely due to the removal of multiple items as opposed to just one. The category re-structuring was again maintained to meet the criteria for rating scale optimization outlined by Linacre (2002). Notably, the category re-structuring again had the favorable impact of almost doubling the range in difficulty seen across these items.

Inspection of the person-item map (see Figure 11) and a bubble chart plotting persons and items (see Figure 12) on the same hypothetical scales also provided some information regarding overall instrument quality for the dynamic self-regulation items. These figures generally indicate that some gaps in item difficulty coverage existed, especially outside of mean-level person ability and at the lower end of the person ability distribution. The implications of these gaps in item difficulty coverage will be discussed in more detail later.

Conclusions. The previous analyses were intended to provide evidence for the validity of claims made from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) regarding active and dynamic self-regulation. This was done primarily by investigating dimensionality of measurement. Dimensionality was explored by way of item model fit and the results of the principal components factor analyses. Information regarding broader instrument quality was also provided by inspection of the person-item maps and various indicators of overall instrument quality.

While only one active self-regulation item was removed, there were seven clearly misfitting dynamic self-regulation items. This suggested that the original dynamic self-regulation item set was less coherent in terms of its measurement of a single construct. Principal components analyses (PCAs) of both the active and dynamic self-regulation items each revealed two possible sub-dimensions[9] captured by the items. Speculation as to the nature of these sub-dimensions likely would be tenuous. However, they do suggest non-dimensionality in both sets of items. The eigenvalue of the first factor revealed in the dynamic set, moreover, was larger than that observed in active self-regulation items. The loadings of individual items on these

---

[9] The previous PCA analysis of the active self-regulation items (Reeves, 2009) only yielded one possible sub-dimension.

factors were also larger in the dynamic set. This suggests that there may be more of a non-dimensionality problem in the dynamic self-regulation items.[10]

Although the ranges in both active and dynamic self-regulation item difficulty were small initially, improvements were seen with the revision of category structure. However, the final item difficulty ranges were still not as large as those which might ensure confidence that one is measuring a real, unidimensional construct (Bond & Fox, 2007). Gaps were also observed, especially at the upper and lower ends of the distribution of participants' abilities, during all Rasch calibrations[11]. This finding again might be best understood by considering the manner in which the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) was developed. That is, those items retained for the most recent version of the DALI-R may represent only those items that maximally discriminated among participants. They therefore may not represent the entire linear progressions.

Despite evidence of non-dimensionality from the PCA, fit statistics mostly supported the notion that the active self-regulation items were measuring something of a linear character. Because of the advantages of the Rasch measurement approach, a variable was constructed from the nine remaining items. This decision was also made in light of other theoretical models of self-regulated learning (SRL; e.g., the performance stage in Zimmerman, 2000; 2002) that are remarkably similar to Iran-Nejad's (1990) notion of active self-regulation.

Again despite the results of PCA, fit statistics for the remaining 14 dynamic self-regulation items also provided some support for the claim that they too might be measuring something of a linear character. Person measures were thus computed from these 14 remaining

---

[10] However, the amounts of explained variance in the active self-regulation observations reveal a pattern suggesting the opposite.
[11] Gaps in item difficulty coverage in the dynamic self-regulation items were more problematic at the lower levels than at the higher levels.

items. Although the variable constructed from this amalgamation of items is hereafter referred to as dynamic self-regulation, the reader should be skeptical at this point as to whether or not it necessarily reflects the non-executive self-regulation of learning processes. The potential limitations of conclusions regarding dynamic self-regulation will be discussed later in much greater detail.

Subsequent analyses were conducted with scale scores created at the conclusion of the psychometric analyses of the DALI-R active and dynamic self-regulation items[12]. Logit scores were re-scaled to a mean of 500 as recommended by Wright and Stone (1979). The value of one logit was also set at 100. This avoided dealing with negative logit values. Active and dynamic self-regulation growth measures were then calculated by subtracting initial from final status Rasch person measure scale scores.

DESCRIPTIVE STATISTICS

Means, standard deviations and other descriptive statistics for all continuous variables are presented in Table 8. There was a restriction in the possible ranges of three out of the five intervention exposure variables (i.e., attendance, commitment and practice). Means and standard deviations for Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) scale score and growth variables for all levels of the factors of classification (e.g., race/ethnicity) are presented in Table 9. Relevant descriptive mean differences will be discussed where appropriate throughout the remainder of this manuscript. Additionally, frequencies and valid percentages for all categorical variables are presented in Table 1.

---

[12] It is notable that Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) items were removed solely on the basis of Rasch model item fit. Thus, although the researcher's opinion regarding the remaining items' representativeness of Iran-Nejad's (1990) theory are presented later, person measures were yielded from only those items identified as behaving in accordance with the Rasch model.

Data were represented graphically and inspected before proceeding with the inferential statistical analyses. The distributions of all continuous variables were examined for normality and skewness. The five intervention exposure variables were negatively skewed. All six variables resulting from Rasch analysis were normally distributed. Scatter plots of the relationships between continuous variables were also inspected for linearity. The relationships between all intervention exposure variables were approximately linear, as were those between the intervention exposure and active and dynamic self-regulation growth variables. Furthermore, the relationships between all Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) scale score and computed growth variables were also approximately linear.

Pearson product-moment correlation coefficients (rs) describing the linear relationships between the five intervention exposure variables ranged from .362 to .766 and were all statistically significant (ps < .01). Pearson product-moment correlation coefficients (rs) describing the linear relationships between the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) scale score and growth variables are presented in Table 10. There were statistically significant correlations between all of these variables except for the correlations between initial status in dynamic self-regulation and final status in both active (r = -.042, p > .05) and dynamic (r = .120, p > .01) self-regulation.

Most notably, there was a moderate negative correlation between active self-regulation initial status and growth (r = -.637, p < .01) and a moderate positive correlation between active self-regulation final status and growth (r = .646, p < .01). This suggests that participants with an initial lower status tended to evidence more growth throughout the course. A similar pattern was observed for dynamic self-regulation. That is, there was a moderate negative correlation (r = -

56

.631, $p < .01$) between dynamic self-regulation initial status and growth, and a moderate positive correlation ($r = .694$, $p < .01$) between dynamic self-regulation final status and growth. Together, these findings suggest that participants with more room to grow in active and dynamic self-regulation demonstrated more growth over the course of the intervention. Also notable, both initial status ($r = .528$, $p = .10$), final status ($r = .540$, $p = .10$) and growth ($r = .599$, $p < .01$) in active and dynamic self-regulation were also moderately positively correlated; these interesting correlations will be discussed later.

Only five bivariate Pearson product-moment correlations between intervention exposure and Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) scale score and growth variables were statistically significant. Attendance was slightly positively related to initial status in both dynamic[13] ($r = .156$, $p < .05$) and active self-regulation ($r = .179$, $p < .05$) and slightly negatively related ($r = -.146$, $p < .05$) to active self-regulation growth. This suggests a tendency for participants with higher initial levels of dynamic and active self-regulation to attend weekly peer monitored self-assessment meetings more often than their counterparts. The (unexpected) small negative correlation between attendance and growth in active self-regulation suggests that students more often attending these meetings tended to experience less growth in active self-regulation. The conclusion that attendance of these meetings is detrimental to growth in active self-regulation, however, is not necessarily warranted based on this small negative correlation. This indicator of meeting attendance is confounded with tardiness, as attendance points were deducted if a student arrived later than ten minutes into the meeting (irrespective of whether or not everything planned for the meeting was accomplished). Given the restricted range and negatively skewed nature of this variable, students who perhaps consistently arrived late—consequently reducing their attendance points—but otherwise self-

---

[13] This bivariate correlation will be expounded in detail later.

reporting growth in active self-regulation might be fueling this small negative correlation. Alternatively, students more often attending these meetings could have possessed higher initial levels of active self-regulation such that they did not experience as much growth throughout the course of the intervention. Empirical support for this latter hypothesis is presented later. Lastly, commitment ($r = .159$, $p < .05$) and practice ($r = .173$, $p < .05$) were slightly positively related to final dynamic self-regulation status. These correlations suggest that there was a slight tendency for those who attempted employing the prescribed techniques (commitment), and for those who ultimately mastered these techniques against a set of stated criteria, to self-report higher final levels of dynamic self-regulation.

INFERENTIAL STATISTICS

Do participants report increased active, dynamic self-regulatory control following the intervention? To examine whether growth in active or dynamic self-regulation occurred over the course of the intervention, on average, one sample *t*-tests were conducted to determine whether mean growth in both active and dynamic self-regulation was statistically distinguishable from zero. Growth in self-reported active self-regulation was statistically significant, $t(183) = 6.971$, $p < .001$. The magnitude of this growth in standard deviation units[14] was estimated at .66. This suggests practically significant growth in active self-regulation over the course of the intervention (Cohen, 1988). Growth in self-reported dynamic self-regulation was also statistically significant, $t(185) = 5.871$, $p < .001$. The magnitude of this growth in standard deviation units was estimated at .55, also suggesting practically meaningful growth in dynamic self-regulation over the course of the intervention (Cohen, 1988).

---

[14] Active and dynamic self-regulation mean growth was divided by the standard deviation of the respective posttest scale scores, which were slightly larger than those at pretest.

Given the moderate negative correlations between initial status in both active and dynamic self-regulation and the respective amounts of growth observed during the course of the intervention, additional analyses were conducted to estimate the amount of growth observed as a function of initial status. Participants were divided into quartiles[15] on the basis of their active and dynamic self-regulation initial statuses for these analyses. Consistent with expectations,[16] the amount of growth observed in both active and dynamic self-regulation throughout the course of the intervention varied as a function of initial status.

To investigate whether these various initial status groups were in fact different in terms of their active or dynamic self-regulation as self-reported on the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad, 1992) at the outset of the intervention, significance tests for initial status scale score mean differences were conducted. Also, estimates of the magnitudes of these differences in standard deviation[17] units were calculated. There were statistically significant differences on the DALI-R between all quartiles for both active and dynamic self-regulation initial status. The magnitude of the difference in standard deviation units between the highest and lowest initial status quartiles was 2.58 for active self-regulation and 2.43 for dynamic self-regulation. Standardized differences on this measure between each of the other quartiles for both active and dynamic self-regulation were also practically meaningful, ranging from .59 to 1.0. These analyses suggested that the individuals comprising these various

---

[15] The groupings are only approximate quartiles (*n*s) because some participants were assigned identical Rasch person measure estimates. This impeded the ability to make the exact quartile cuts specified by the analyst.

[16] The threat of regression toward the mean was kindly brought to the author's attention by J. Lee (personal communication, July 7, 2009).

[17] Mean quartile differences were divided by the respective pooled standard deviations of the active and dynamic initial status scale score variables.

groupings were meaningfully different—according to the DALI-R—at the beginning of the intervention.

The next set of analyses was geared at investigating whether growth in both active and dynamic self-regulation varied as function of participants' respective initial status. For active self-regulation (see Table 11), only three of the initial status quartiles experienced statistically significant change throughout the course of the intervention. Specifically, the first and second quartiles demonstrated statistically significant growth from pre- to posttest. The respective magnitude of the growth for these quartiles in standard deviation units was 1.71 and .67. Although descriptive statistics revealed that the third quartile also experienced growth, it was not statistically distinguishable from zero. Interestingly, the fourth quartile experienced a statistically significant decrease in self-reported active self-regulation throughout the course of the intervention (.36 of a standard deviation).

This pattern was similar for dynamic self-regulation (see Table 12), although the third quartile also experienced statistically significant growth throughout the course of the intervention. The magnitude of self-reported growth in dynamic self-regulation for the first, second and third quartiles, respectively, was 1.42, .77, and .37. As in active self-regulation, the fourth quartile also self-reported a statistically significant decrease in dynamic self-regulation throughout the course of the intervention. The magnitude of this decrease in standard deviation units was .36 of a standard deviation.

To explore the problematic finding that some participants self-reported decreases in active and dynamic self-regulation throughout the course of the intervention, the upper initial status quartiles for each construct were next examined more closely. First, descriptive statistics and a graphic representation of the data were inspected. The ranges and standard deviations in

60

initial status were larger in the upper quartiles for both active and dynamic self-regulation. Histograms for both the active and dynamic self-regulation lowest initial status quartiles revealed that the distributions were positively skewed[18]. Second, initial status person measures for this quartile were considered with respect to Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) Rasch item difficulty estimates. For active self-regulation initial status, 24 of the 37 individuals in the upper quartile had person ability estimates which exceeded the estimate of the most difficult item. That is, they were initially very high on active self-regulation as measured by this instrument. If one was using a ruler, this would be analogous to making a measurement estimate that is above those very highest marks on the ruler itself. The lack of items with estimated difficulties higher than the initial statuses of these individuals implies that the instrument may not necessarily have featured items to further differentiate individuals within an even higher ability range. In contrast, for dynamic self-regulation initial status, only 1 of the 42 individuals in the upper quartile had a person ability estimate that exceeded the highest item difficulty estimate. Unlike with active self-regulation, this does not necessarily suggest that an initial lack of room for growth is behind their failure to demonstrate such growth at posttest using this instrument.

These analyses might shed light on the reasons why some individuals did not evidence growth in active self-regulation throughout the course of the semester. They do not, however, provide evidence as to why this was the case with dynamic self-regulation. Further, these analyses do not explain why there was a statistically significant *decrease*, on average, for these highest initial status groups. An alternative explanation for these problematic findings is

---

[18] This was particularly the case for active self-regulation initial status within this upper range. However, the distributions for the entire range of initial status in both active and dynamic self-regulation were normally distributed.

presented later in Conclusions. Additionally, the implications of these descriptive analyses for the suitability of using these items with this calibration sample are presented later.

Does observed growth differ by instructor? Descriptive analyses (see Table 9) revealed that there were sample mean differences in both active and dynamic self-regulation growth favoring different instructors. However, a multivariate analysis of variance (MANOVA) did not yield a statistically significant multivariate difference in active and dynamic self-regulation growth by instructor, $F(2, 181) = 1.105$, $p = .333$. Univariate results for both active, $F(1, 182) = .498$, $p = .481$, and dynamic, $F(1, 182) = .387$, $p = .535$, self-regulation mean growth differences were also not statistically significant. These results suggest that there were not differences in self-reported active and dynamic self-regulation growth observed during the course as a function of the instructor.

Does observed growth differ by participant motivation? Descriptive analyses revealed mean differences in active self-regulation growth between those participants enrolling in the intervention for different primary reasons (see Table 9), favoring those students indicating "I needed to improve my GPA," "My advisor recommended it" and "It was mandatory[19]" over those enrolled because "The course content seemed interesting" and "I needed an elective" (in that order). The pattern of mean growth differences for dynamic self-regulation was somewhat different, although the same three groups experienced the most growth (see Table 9); dynamic self-regulation growth favored those responding "My advisor recommended it," "It is mandatory" and "I needed to improve my GPA" over those responding "I needed an elective" and "The course content seemed interesting" (in that order).

---

[19] The "It was mandatory" response category growth means for both active and dynamic self-regulation were based on only two participants.

After excluding the "It was mandatory" group ($n = 2$), a multivariate analysis of variance (MANOVA) revealed a statistically significant multivariate difference in active and dynamic self-regulation as a function of participant motivation, $F(6, 352) = 2.421$, $p < .05$. However, inspection of univariate analysis of variance (ANOVA) results demonstrated that only growth in active self-regulation was distinguishable by this motivation grouping, $F(3, 176) = 3.162$, $p < .05$; growth in dynamic self-regulation was only marginally so, $F(3, 176) = 2.507$, $p = .06$. At first blush, this suggested that participant motivations were relevant to the amount of growth observed in active (and possibly dynamic) self-regulation.

However, the reader will recall that there was a negative bivariate Pearson product-moment correlation between initial status in both active and dynamic self-regulation and the respective amount of growth observed. Consequently, analyses were conducted to rule out the possibility that the relationship between participant motivation and growth was a regression artifact. After including both active and dynamic self-regulation initial statuses in the model, the multivariate difference was no longer statistically significant, $F(6,348) = 1.437$, $p = .199$. Univariate differences were also not observed for either active, $F(3,174) = 1.948$, $p = .124$, or dynamic, $F(3,174) = .148$, $p = .931$, self-regulation after controlling for initial status in these variables. This suggests that motivational factors were relevant in explaining observed growth only to the extent that they reflected the amount of room for growth participants had initially.

Six post hoc contrasts were conducted secondary to the omnibus multivariate analysis of variance (MANOVA) to determine whether there were specific statistically significant differences in active and dynamic self-regulation growth between groups indicating particular primary reasons for enrolling in the intervention. There was not a statistically significant multivariate difference in active and dynamic self-regulation growth between the "My advisor

recommended it" and the "I needed to improve my GPA" groups, $F(2, 91) = .232$, $p = .793$, nor were the univariate differences in active, $F(1, 92) = .054$, $p = .816$, or dynamic, $F(1, 92) = .120$, $p = .730$, self-regulation statistically significant. There also was not a statistically significant multivariate difference in active and dynamic self-regulation growth between the "My advisor recommended it" and the "I needed an elective" groups, $F(2, 48) = 2.371$, $p = .104$. However, there was a statistically significant univariate difference in active self-regulation, $F(1, 49) = 4.766$, $p < .05$, favoring those that endorsed "My advisor recommended it." There was no statistically significant difference observed between these two groups in terms of their growth in dynamic self-regulation, $F(1, 92) = 1.751$, $p = .192$. After controlling for initial status in active self-regulation, however, there was no longer a statistically significant difference between these two groups in terms of the amount of active self-regulation growth they self-reported, $F(1, 48) = 3.401$, $p = .071$. This again suggested that it was not motivational factors that fueled this difference but rather participants' levels of active self-regulation at the beginning of the intervention. There was not a statistically significant multivariate difference in active and dynamic self-regulation between the "My advisor recommended it" and "The course content seemed interesting" groups, $F(2, 82) = 2.442$, $p = .093$. The univariate difference for active self-regulation also was not statistically significant, $F(1, 83) = 1.883$, $p = .174$. However, there was a statistically significant difference in terms of dynamic self-regulation growth, $F(1,83) = 4.886$, $p < .05$, favoring those that endorsed "My advisor recommended it." After controlling for initial status in dynamic self-regulation, this difference was again no longer statistically significant, $F(1, 82) = .619$, $p = .434$. There was a statistically significant multivariate difference in active and dynamic self-regulation between those responding "I needed to improve my GPA" and "I needed an elective," $F(2, 92) = 3.670$, $p < .05$. However, only the univariate difference between

these two groups in active, $F(1, 93) = 6.600$, $p < .05$, but not dynamic, $F(1, 93) = .896$, $p = .346$, self-regulation was statistically significant. This difference favored those indicating "I need to improve my GPA." Unlike the previous analyses, after controlling for active self-regulation initial status for this comparison, this difference remained statistically significant, $F(1, 92) = 5.552$, $p < .01$. To investigate the magnitude of this difference in standard deviation units after controlling for initial status in active self-regulation, a univariate multiple regression analysis was conducted to simultaneously predict active self-regulation growth on the basis of participants' initial status and a dummy variable in which "I needed to improve my GPA" was coded 1 and "I needed an elective" was coded 0. The resulting standardized beta coefficient was .181; this suggests a small effect of being motivated to improve one's GPA relative to taking the course as an elective. The contrast between "I needed to improve my GPA" and "The course content seemed interesting" did not yield a statistically significant multivariate difference in active and dynamic self-regulation, $F(2, 126) = 2.842$, $p = .062$. However, both univariate tests for differences in active, $F(1, 127) = 4.025$, $p < .05$, and dynamic, $F(1, 127) = 5.140$, $p < .05$, self-regulation were statistically significant. When respectively controlling for active and dynamic self-regulation initial status, however, growth in both active, $F(1, 125) = .553$, $p = .458$, and dynamic, $F(1, 125) = .002$, $p = .965$, self-regulation was no longer distinguishable by this grouping. The last contrast between "I needed an elective" and "The course content seemed interesting" did not yield either a multivariate difference, $F(2, 83) = .054$, $p = .098$, or univariate differences in either active, $F(1, 84) = 1.556$, $p = .216$, or dynamic, $F(1, 84) = .795$, $p = .375$, self-regulation growth.

Does observed growth differ by the extent to which participants receive the intervention as intended? Multiple regression analyses were conducted to determine whether the five indicators of intervention exposure were systematically related to any observed growth in either active or dynamic self-regulation. Only attendance of weekly self-assessment meetings with a peer monitor was related statistically to growth in active self-regulation while controlling for the other intervention exposure variables (Beta = -.321, t = -2.801, p < .01). Specifically, a one standard deviation increase in the attendance of these meetings was associated with a .321 of a standard deviation decrease in active self-regulation growth. Like the unexpected negative bivariate correlation between these variables reported earlier, this suggests that attendance of weekly meetings was associated with less growth in active self-regulation.

However, in light of the finding that initial status in active self-regulation was related to the amount of growth observed, a second model that included active self-regulation initial status was estimated. This was done to combat the threat of regression toward the mean to the internal validity of this finding. This analysis examined the extent to which these various intervention exposure variables were systematically related to the growth outcome while controlling for the others as well as accounting for where individuals were at the beginning of the intervention. After accounting for initial status in active self-regulation, attendance was no longer systematically related to the active self-regulation growth outcome, t = 1.527, p = .129. This lends support to the author's hypothesis regarding this unexpected negative correlation that students more often attending weekly self-assessment meetings simply had less room to grow in active self-regulation over the course of the intervention.

Only the extent to which students mastered the prescribed techniques (practice) was related systematically to growth in dynamic self-regulation (Beta = .199, t = 2.019, p < .05)

while controlling for the other intervention exposure variables. Specifically, a one standard deviation increase in practice was associated with a .199 of a standard deviation increase in dynamic self-regulation growth. However, after controlling for initial status in dynamic self-regulation, practice was no longer related to dynamic self-regulation growth ($t = 1.404$, $p = .162$).

Does observed growth differ by participant race/ethnicity or gender? After excluding Hispanic ($n = 10$) participants, as well as those indicating "Other" ($n = 1$) and whose race/ethnicity was unknown ($n = 19$), a multivariate analysis of variance (MANOVA) did not yield a statistically significant multivariate difference in active and dynamic self-regulation growth by race/ethnicity, $F(4, 302) = 1.476$, $p = .209$. Univariate results for both active, $F(2, 151) = 2.558$, $p = .081$, and dynamic, $F(2, 151) = .209$, $p = .812$, self-regulation mean growth differences were also not statistically significant. These results suggested that there were not omnibus race/ethnicity differences in self-reported growth in active and dynamic self-regulation observed throughout the course.

Three post hoc contrasts, however, were conducted secondary to the omnibus multivariate analysis of variance (MANOVA) to determine whether there were specific statistically significant differences in active and dynamic self-regulation growth between particular racial/ethnic subgroups. There was not a statistically significant multivariate difference between White and Black students in terms of their growth in active and dynamic self-regulation, $F(2, 131) = .457$, $p = .634$. Univariate results for both active $F(1, 132) = .920$, $p = .339$, and dynamic, $F(1, 132) = .288$, $p = .593$, self-regulation also were not significant for this contrast. There also was not a statistically significant multivariate difference between Asian or Pacific Islander and Black participants in terms of their growth in both active and dynamic self-

regulation $F(2, 38) = .637$, $p = .535$. Univariate results for both active, $F(1, 39) = .919$, $p = .344$, and dynamic, $F(1, 39) = .010$, $p = .923$, self-regulation again were not significant. The multivariate difference between Asian or Pacific Islander and White students also was not statistically significant, $F(2, 130) = 2.868$, $p = .060$. However, there was a statistically significant univariate difference in growth in active self-regulation, $F(1, 131) = 4.797$, $p = .05$, favoring Asian or Pacific Islander students. There was not a statistically significant univariate difference between these groups for dynamic self-regulation, $F(1, 131) = .182$, $p = .670$.

These findings suggested that although there were not omnibus race/ethnicity differences in terms of growth in active and dynamic self-regulation, there was a specific difference in active self-regulation growth favoring Asian or Pacific Islander over White participants. However, to ensure that this difference in growth was not a regression artifact, initial status in active self-regulation was simultaneously included in this same statistical model just reported. Results indicated that after including initial status in the model, there was no longer a statistically significant difference, $F(1, 130) = 2.456$, $p = .120$, between Asian or Pacific Islander and White students in terms of the active self-regulation growth they observed during the intervention. This suggests that the growth differences observed were more an effect of initial differences in the dependent measure than they were the result of differential intervention effectiveness. Inspection of Table 9 lends support to this contention.

Descriptive analyses (see Table 9) revealed slight differences in active and, in particular, dynamic self-regulation growth means as a function of participant gender. However, a multivariate analysis of variance (MANOVA) did not yield a statistically significant multivariate difference in active and dynamic self-regulation growth by gender, $F(2, 181) = .160$, $p = .853$. Univariate results for both active, $F(1, 182) = .036$, $p = .851$, and dynamic, $F(1, 182) = .292$, $p =$

.589, self-regulation mean growth differences were also not statistically significant. These results suggest that there were not gender differences in self-reported active and dynamic self-regulation growth observed throughout the course.

CHAPTER FIVE: CONCLUSIONS

This study first investigated the validity of Dynamic and Active and Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) inferences regarding active and dynamic self-regulation via the application of the Rasch measurement model. The data were collected both before and after an intervention targeted at fostering active and dynamic self-regulation (Iran-Nejad, 1990) in undergraduate students. Subsequent analyses attempted to estimate the amount of growth observed in both active and dynamic self-regulation over the course of the intervention and to explain the variance in any observed growth in terms of participant motivations and intervention exposure. Furthermore, differential effects of the intervention by initial status, race/ethnicity and gender were also investigated. Findings from analyses aimed at answering these various research questions are now discussed.

Results of those analyses aimed at providing evidence for the validity of inferences made from the DALI-R (Iran-Nejad & Chissom, 1992) were somewhat different than the author's expectations. The author had hypothesized that while the analyses would provide support for the DALI-R measuring active self-regulation as a linear, unidimensional construct, they would fail to provide similar evidence regarding dynamic self-regulation. In contrast, mixed evidence for assertions that the DALI-R measures these two distinct, linear constructs was lent by the analyses. Removal of one and seven active and dynamic self-regulation items, respectively, left nine active and 14 dynamic items behaving mostly with the expectations of separate Rasch models. Remaining item fit suggested these two sets of items were each capturing something of a linear character.

However, principal components analysis (PCA) of the standardized model residuals for both active and dynamic self-regulation item sets indicated that there were two sub-dimensions

possibly captured by each. Notably, PCA eigenvalues and factor loadings appeared more problematic (i.e., larger) in the dynamic self-regulation items, but the amount of variance explained by the PCA resultant factors in the active self-regulation item observations suggested the opposite. Nevertheless, in light of the remaining item sets' model fit, variables representing active and dynamic self-regulation were computed for answering additional research questions. These results will be discussed following the implications of the psychometric analyses.

Iran-Nejad's (1990) account of active and dynamic self-regulation as independent mechanisms by which learning processes are internally self-regulated does not exist in isolation from other accounts of this phenomenon (i.e., self-regulated learning). The literature review presented earlier discussed other contemporary theoretical accounts of these findings. For example, Zimmerman's (2000; 2002) account of self-regulated learning (SRL) was presented in great detail. The reader will recall that Zimmerman's (2000; 2002) and other (e.g., Winne & Hadwin, 1998) models of SRL include the performance of strategic, goal-directed acts. This aspect of these models of SRL is remarkably similar to Iran-Nejad's (1990) notion of active self-regulation.

In light of these other theoretical accounts of self-regulated learning, the analyses reported earlier are interpreted by the author to provide some support for the claim that the DALI-R active self-regulation items are measuring the active, strategic self-regulation of learning. Care should be taken not to exaggerate the strength of this evidence, however, since possible sub-dimensions were observed in data collected from two separate samples. Future research should unquestionably investigate any systematicities among the items loading on these potential sub-dimensions.

Conclusions as to the validity of DALI-R inferences regarding dynamic self-regulation are more elusive. Despite the acceptable model fit of 14 remaining items, a question arises as to what is actually being measured. Inspecting the items in Appendix B for face validity reveals that some of the remaining items do not appear to reflect non-executive self-regulatory processes. For example, it is unclear how the item *When I do NOT understand a concept, I ask the teacher or other fellow students for clarification* reflects something that occurs outside the control of the central executive. There are additional examples in this set of 14 items. Interestingly, many of these face invalid items did behave in accordance with the Rasch model when submitted together with the others.

The nature of dynamic self-regulation thus still remains clouded. Iran-Nejad and Chissom (1992) state that dynamic self-regulation is marked by interest, curiosity, postdiction and reflective metacognition, among other things. One, however, could argue that these are each actually linear constructs themselves. To counter such an argument, it would be possible to define dynamic self-regulation as a broader unidimensional construct that is evidenced by things like interest, curiosity, postdiction and reflective metacognition. Yet, even if this argument was made, the question would still remain as to why some items do not, at face value, appear to reflect non-executive self-regulatory intrinsic brain processes. This *non*-executive characteristic is noted by Iran-Nejad (1990) to be fundamental to the nature of dynamic self-regulation.

Aside from qualitative observations on the part of the author, some findings from this research project might shed light on this issue. In particular, the initial correlation between active and dynamic self-regulation was statistically significant and moderate in magnitude, $r = .528, p < .01$. This observed correlation might have been unexpected for two reasons. First, Iran-Nejad (1990) theorized that these two sources of self-regulation govern different forms of learning and

are distinct mechanisms by which individuals' learning processes are internally self-regulated. If these two sources of self-regulation represent separate constructs, such a large statistical relationship might not have been expected. Second, Iran-Nejad (1990) asserts that as students grow older, they rely less and less on dynamic self-regulation in favor of engaging in active, strategic self-regulatory behavior. According to Iran-Nejad (1990), individuals enrolled in post-secondary institutions would thus be relying primarily on active self-regulation. If this was the case, the initial correlation between active and dynamic self-regulation might have been less robust, as the data were collected from a sampling of this population of students. As it was, individuals who often engaged in active self-regulation were also somewhat inclined to experience the influence of dynamic self-regulation on their learning processes. Furthermore, the moderate bivariate correlations between active and dynamic self-regulation at the end of the intervention, as well as the amount of growth observed throughout its course are also noteworthy. These correlations too suggest that there might be some overlap between these two constructs.

It is also interesting that the extent to which participants attended weekly self-assessment meetings was positively related to initial status in *both* active and dynamic self-regulation.[20] The extent to which a participant attends such meetings clearly has a motivational aspect to it, and this correlation suggests that motivational factors were related to both dynamic and active self-regulation prior to the intervention. The correlation between attendance and the former (i.e., active self-regulation) might have been expected, as they both involve actively behaving in a manner that contributes to success in one's academic endeavors (e.g., engaging in strategic self-regulation or going to a required course component). Furthermore, Murphy and Alexander

---

[20] The magnitudes of these two bivariate correlations were also similar.

(2000) and McComb and Marzano (1990) underscore the importance of motivational factors in similar conceptions of self-regulated learning (SRL).

In contrast, the correlation between attendance and one's level of dynamic self-regulation at the beginning of the course might not have been expected. Working off of Iran-Nejad's (1990) theory, the author would hypothesize[21] that the extent to which one makes an effort to attend weekly self-assessment meetings would not necessarily be related to the extent to which their learning processes are internally self-regulated by non-executive (i.e., *uncontrolled*) intrinsic brain processes. This correlation suggests—albeit tenuously—that there may be somewhat of a motivational character to "dynamic self-regulation" as measured by the DALI-R. However, it is recognized that these perhaps unexpected bivariate correlations cannot be used to solely frame such an argument. Nonetheless, inferences that follow regarding "dynamic self-regulation" (Iran-Nejad, 1990) should be interpreted with caution, as additional research into its nature is still clearly needed.

On the whole, self-reported growth in both active and dynamic self-regulation was observed throughout the course of the intervention. The magnitude of this overall growth was on the order of .66 of a standard deviation for active self-regulation and .55 of a standard deviation for dynamic self-regulation. After taking into consideration participants' initial status, however, conclusions regarding growth throughout the course of this intervention were more nuanced. The approximately 50 percent of participants with the lowest initial self-reported levels of active self-regulation experienced practically meaningful growth in this construct by the end of the intervention. Moreover, the approximately 75 percent participants with the lowest self-reported

---

[21] There were no a priori hypotheses regarding the relationships between the intervention exposure variables and Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) initial and final status scale score variables; hypotheses were only made regarding some the growth variables. The consideration of this issue occurred only after descriptive statistical analyses.

initial levels of dynamic self-regulation demonstrated practically meaningful growth, on average, in that which is captured by this set of items[22]. These findings could be taken to suggest that this intervention is effective in promoting active and dynamic self-regulation in most of its participants.

Then again, not all participants experienced self-reported growth in active and dynamic self-regulation. For example, those initially between approximately the 50th and 75th percentiles in active self-regulation did not demonstrate statistically significant growth throughout the course of the intervention. This finding suggests that some participants do not benefit, in terms of active self-regulation, from the intervention. However, this finding might be better understood by considering the statistical power of the inferential statistical analyses employed. Although the amount of growth self-reported was statistically indistinguishable from zero, inspection of descriptive statistics (see Table 9) reveal that there was a positive growth mean for this group. Yet, the sample size may not have afforded enough statistical power to confidently detect an effect of that magnitude.

Also, those initially in approximately the top 25 percent in terms of their active and dynamic self-regulation actually reported respective statistically significant decreases in active and dynamic self-regulation from pre- to posttest. This unexpected finding that some students actually self-report less active and dynamic self-regulation at posttest is a much more problematic finding than no growth at all. That is, this intervention is intended to foster self-regulatory abilities, not diminish them among those whose learning is already self-regulated. Descriptive analyses reported earlier suggested that the range in active self-regulation item

---

[22] For example, Hattie, Biggs, and Purdie's (1996) meta-analysis revealed that motivational factors can successfully be targeted by these kinds of interventions.

difficulty might have prohibited demonstrating growth for those participants initially very high on active self-regulation. No such evidence was found to provide a similar instrumentation-related explanation for a lack of observed growth for the upper quartile in dynamic self-regulation. Furthermore, those analyses could not elucidate why *decreases* were observed for these initially highest groups. To explain this, one might consider the accuracy with which participants initially responded to the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) at the beginning of the intervention. For example, it is possible that those participants indicating initially high levels of either active or dynamic self-regulation may have been overconfident at pretest only to perhaps more realistically report their self-regulatory processes by the intervention's end. This conclusion, however, is purely speculative on the part of the author.

These findings that some participants experienced decreases in active and dynamic self-regulation could be taken to indicate that this *universalized* intervention should instead be implemented as a *targeted* intervention (Ceci & Papeirno, 2005). Importantly, targeting the intervention might ensure that it is provided to students in whom growth is likely to be observed. Programmatic changes in response to these findings, however, would be premature in light of some problematic findings yielded from the psychometric analyses reported here. A qualitative follow-up study of the intervention experiences of those students self-reporting decreases in active or dynamic self-regulation might shed some light on this issue.

It was reported earlier that growth in both active and dynamic self-regulation did not vary as a function of the instructor who provided direct instruction in course content. This is consistent with findings reported by Tinnesz, Ahuna & Keiner (2006). This failure to explain variability in growth via an aspect of the intervention itself thus cannot be used to support the

76

inference that the intervention is in fact the cause of any observed growth. However, this null effect also cannot necessarily be taken as evidence that the intervention is *not* the cause of any observed growth. Assuming that other aspects of instructor quality (e.g., delivery) were similar, such differential growth as a function of the instructor might not necessarily have even been expected. The interpretation of this finding as it relates to the fidelity with which the intervention was implemented by the instructors will be elaborated later.

The next research question attempted to address roughly the influence of motivational factors on the amount of active and dynamic self-regulation growth that was observed throughout the course of the intervention. Although there were a fair number of initial growth differences between groups endorsing various primary reasons for enrolling in the course, all but one became non-significant after controlling for initial status in the relevant measure. The only remaining statistically significant difference was that between "I needed to improve my GPA" and "I needed an elective" in active self-regulation growth. The mean difference in active self-regulation was most pronounced for this pair, so it is not surprising that this was the only difference that remained statistically significant after controlling for initial status.

It is difficult to provide a generalizable interpretation of this finding in which the author can be confident. These five response categories were not explicitly grounded in the motivation literature when they were developed by the author and his colleague and subsequently administered for programmatic purposes. However, this particular contrast[23] might correspond roughly to the distinction between intrinsic ("I needed an elective") and extrinsic ("I needed to

---

[23] The contrast was coded 1 for "I needed to improve my GPA" and 0 for "I needed an elective" because the link between the former and extrinsic motivation was stronger than that between the latter and intrinsic motivation. This variable might thus be considered a rough dummy for extrinsic motivation rather than a contrast between intrinsic and extrinsic motivation.

improve my GPA") motivation (e.g., Ryan & Deci, 2000). Assuming that these response categories roughly captured these two motivations for engaging in behavior, it would appear that individuals who were extrinsically motivated (i.e., by the reward of an improved GPA) were slightly more successful in evidencing growth in active self-regulation throughout the intervention. Nevertheless, cautions are in order owing to the post hoc nature of this hypothesis and the crude programmatic indicator employed.

These analyses demonstrated that only one intervention exposure variable was related to the amount of growth observed in both active and dynamic self-regulation. Specifically, only attendance of weekly self-assessment meetings was associated with growth in active self-regulation, while only the extent to which students mastered the prescribed techniques was related to growth in dynamic self-regulation. However, after accounting for participants' respective initial statuses in either active or dynamic self-regulation, the relationships between these intervention exposure indicators and growth were no longer statistically significant. Thus, these findings do not provide evidence that can support an inference that the intervention itself is the *cause* of observed growth in active or dynamic self-regulation. The extent to which participants had room for growth was instead more important in providing an explanation for the growth that was observed.

While this last set of findings might make the reader skeptical of causal inferences regarding intervention effectiveness, it is necessary that care is taken to carefully consider the intervention exposure variables employed in this study. The reader will recall that there were restricted ranges in some of the intervention exposure variables and that all five were negatively skewed. Moreover, these variables all demonstrated somewhat of a ceiling effect where most participants were assigned scores at the high end of the distribution. Thus, while the failure to

78

explain variance in the growth measures could in part suggest that observed growth is due to maturation or other factors (e.g., a testing effect), the intervention exposure measures employed also might simply not have shown enough variability to evidence systematic relationships with the outcomes investigated.

The next research question involved whether or not this intervention was differentially effective for certain racial/ethnic and gender groups. The only difference observed in the amount of self-reported growth was between Asian or Pacific Islander and White students (favoring the former) for active self-regulation. However, this difference was null after accounting for initial status in participants' active self-regulation. That is, the observed difference in growth as a function of race/ethnicity was owed to the amount of room there was for growth rather than differential effectiveness of the intervention itself. The analyses reported earlier also did not demonstrate statistically distinguishable growth in either active or dynamic self-regulation as a function of participant gender. That is, both males and female intervention participants appear to experience similar growth patterns in active and dynamic self-regulation throughout the course of the intervention. This is not inconsistent with expectations, as there was no a priori directional hypothesis predicting that there would be differential growth observed by gender.

Taken together, these findings indicate that beyond their initial status all participants[24], in terms of their race/ethnicity[25] and gender, experience similar growth patterns in active and dynamic self-regulation. Thus, while this does not provide evidence that this intervention can potentially help close extant race/ethnicity and gender gaps in undergraduate degree-completion rates, it also does not suggest that it could potentially exacerbate such gaps. The finding that

---

[24] As discussed earlier, growth in active self-regulation favored participants indicating that their primary reason for enrolling in the intervention was "I needed to improve my GPA" over those indicating "I needed an elective." Also, interactions between subject variables (addressed later) were not investigated.

[25] Conclusions regarding differential intervention effectiveness by race/ethnicity are limited to the present comparisons between White, Black and Asian or Pacific Islander intervention participants.

growth was not moderated by race/ethnicity or gender was consistent with earlier reviewed research by Tinnesz, Ahuna & Keiner (2006) who found that growth did not differ by these and other (i.e., year in school) participant characteristics.

LIMITATIONS AND FUTURE DIRECTIONS

Although this study attempted to combat threats to the validity of prior empirical investigations, it certainly was not without limitations of its own. First, there remain threats to the validity of drawn conclusions related to the instrumentation employed. For example, the psychometric analyses conducted with Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) data were somewhat limited. Second, the timing of data collection also poses a threat to the validity of drawn conclusions. Third, the fidelity with which the intervention was implemented was largely neglected in this study. Lastly, threats related to the data and design of the study were also present. Fortunately, many of these threats can be addressed or mitigated by additional research. These threats will now be discussed in turn; how additional research may combat them will be discussed where appropriate.

Instrumentation. The proposed study attempted to provide evidence as to the validity of inferences that are made from the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992). These analyses, however, only considered items, not persons. Future analyses might also want to consider persons, given that any sources of bias on the part of respondents would preclude optimal test development as they would be inherently confounded with the behavior of items. Furthermore, the conclusions regarding active, and in particular, dynamic self-regulation are somewhat tenuous in light of mixed findings from the psychometric analyses regarding dimensionality. Also, the findings reported here regarding growth in dynamic self-regulation should be interpreted with extreme caution, as inspection of these items for face

80

validity casts doubt on what, in fact, grew throughout the course of the intervention. Consistent with recommendations by Schapiro and Livingston (2000), future research is still clearly needed to examine the nature of dynamic self-regulation.

The present study did not investigate the measurement invariance of the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) items by way of differential item functioning (DIF) analyses. Such analyses could be conducted using those grouping variables (e.g., race/ethnicity) that were considered during the answering of the substantive research questions. Specifically, DIF analysis could be conducted with respect to gender, and race/ethnicity. DIF analysis similarly could be conducted to evidence the invariance of the instrument between its pre- and posttest administrations. These analyses could help to identify bias in the instrument's items that could preclude sound inferences.

Although the development of both theoretically and empirically defensible instruments is iterative and involves multiple phases of data collection, this work still represented an improvement to the measurement and estimation of growth performed in prior research (e.g., Tinnesz, Ahuna & Keiner, 2006). Barring the results of principal components factor analyses (PCAs) and other concerns introduced in response to the psychometric analyses, the placing of persons on an interval scale allowed person measures to not violate the assumptions of the inferential statistics to which they were submitted. This would have not been the case had persons' individual item scores simply been summed to yield total scores. However, the calibration that was performed in this study utilized a relatively small sampling of respondent data (J. Lee, personal communication, July 7, 2009). This casts doubts on the generalizability of the claims presented here regarding the validity of DALI-R inferences. Inconsistencies between the previous and new analyses of the active self-regulation items support this contention.

81

In addition to a large calibration sample (or samples), however, rigorous instrument development also requires that the range of the sample's abilities represent those for which the instrument is ultimately intended (Bond & Fox, 2007). These items were developed for undergraduate students and were piloted with a sample of these students enrolled in an Introductory Psychology course at a large, public institution in the South. It is generally assumed in Psychological research that such courses are representative of those populations to which generalizations are made. Students enrolled in the present intervention ranged from those on academic probation to those taking it as a course elective. It could be assumed that this sample represents the range of abilities for which the DALI-R is intended. However, earlier and new analyses were geared at investigating the suitability of this instrument for the particular sample employed in the present study.

Aside from issues related to those possible sub-dimensions captured by the items, the reader will recall that inadequate person reliability and separation summary statistics were observed in both sets of items. While these could be related to features of the instrument itself (e.g., the number of or difficulty range covered by the items), the sample employed during the calibration should also considered (J. Lee, personal communication, August 4, 2009). The range in initial status person estimates was 7.6 logits for active self-regulation and 5.6 logits for dynamic self-regulation. The ranges in these estimates suggest that there was in fact variability within this calibration sample in terms of both active and dynamic self-regulation.

However, the reader will recall that many of those initially highest in active self-regulation received Rasch person measure estimates that exceeded the highest item difficulties. This provided some—albeit mixed—information as to the appropriate targeting of persons by the measure. To further shed light on the appropriate targeting of persons by the measure, the lower

initial status quartiles for both active and dynamic self-regulation were also examined in a manner consistent with that performed earlier. Similar to the upper quartiles, the ranges and standard deviations in initial status were larger in the lower quartiles for both active and dynamic self-regulation. Histograms for both of these lower quartiles revealed that the distributions were negatively skewed.[26] For active self-regulation initial status, all 53 of the individuals in the lower quartile had person ability estimates that were beneath the lowest item difficulty estimate. For dynamic self-regulation initial status, only 13 of the 64 individuals in the lowest quartile had person ability estimate that were beneath the highest item difficulty estimate.

In summary, with the exception of the upper range of dynamic self-regulation, the calibration sample included many individuals whose abilities lie outside the range of difficulties covered by the instrument's items. Not surprisingly, these conclusions are not different from those drawn from inspection of the person-item maps (see Figures 5 and 11). The analogy of this instrument as a ruler measuring both active and dynamic self-regulation in this particular calibration sample presented earlier again adequately conveys the implications of this matter. Together, these findings cast doubt on whether the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) instrument was well-suited for use with the entirety of this particular sample.

There is one additional finding revealed at the culmination of the psychometric analyses that deserves attention. Namely, there were notably low correlations between participants' initial and final statuses in both active and dynamic self-regulation. Despite the fact that change was self-reported throughout the course of the intervention, one might have expected there to be more of a tendency for individuals to somewhat remain around their initial levels at the end of the

---

[26] This was particularly the case for the dynamic self-regulation initial status variable within this lower range.

intervention. These small correlations between both of these constructs at the earlier and later time points could cast double on the stability of the measure (J. Lee, personal communication, August 4, 2009). Alternatively, the nature of growth in these constructs over the course of the intervention could be considered to explain these low correlations. If growth was happening consistently across all the intervention participants, one would expect those with higher and lower initial levels to still remain at relatively higher and lower levels, respectively. The lack of consistency between participants' initial and final statuses in active and dynamic self-regulation could suggest that growth may have occurred less neatly across intervention participants.

Although the psychometric analyses conducted for the present study could be advanced via additional research, future investigations of this intervention could also attempt to measure self-regulated learning (SRL) using other operations of this construct. For example, Zimmerman (2008) reviews three other well-established questionnaire and interview measures of SRL: the Learning and Study Strategies Inventory (LASSI; Weinstein, Schulte & Palmer, 1987); the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993); and the Self-Regulated Learning Interview Scale (SRLIS; Zimmerman & Martinez-Pons, 1986, 1988). These measures, or still others described by Zimmerman (2008), could in the future be administered either in lieu of or in tandem with the Dynamic or Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992). The latter would permit additional information regarding the validity of inferences made using the DALI-R.

This study also suffers from limitations related to how motivational factors and intervention exposure were operationalized. The reader will recall that a programmatic assessment question asking students to report their primary reason for enrolling in the intervention was used as a crude indicator of participants' motivations. Given the purpose of its

84

addition to the assessment by the author, these response categories were loosely (at best) aligned with the motivation literature. Using this operation of motivational factors, only one difference between the response category groupings remained statistically significant after controlling for initial status. Given the relationship between motivational factors and educational outcomes discussed earlier (i.e., Uguroglu & Walberg, 1979), this crude indicator of participant motivation could be to blame for the failure to explain growth as a function of participant motivation. A different indicator might have been better able to capture real differences—if they exist. Had such an indicator been employed in this study, conclusions regarding the role of motivational factors in the amount of growth observed might have been different.

Furthermore, most the indicators of intervention exposure employed in this study may have also been problematic. The reader will recall that all five indicators demonstrated little variability and somewhat of a ceiling effect. It is possible that variance in observed growth might have been explicable by these indicators had they shown more variability. There were also other potential problems with some of these indicators. First, the precision of the attendance and exposure indicators might have been impeded by the particular grading scheme employed in the course. For example, participants received the same amount of exposure points if they had missed either zero, one or two class sessions. Next, the commitment indicator was likely inflated owing to attempts to reward effort. For example, students would have received full credit for attempting to implement the prescribed techniques if they attempted at least half of them. Lastly, the practice points indicator employed may also have been problematic owing to a modified mastery learning (e.g., Block & Burns, 1976) philosophical approach used in the course. That is, two participants could have the same score for practice regardless of whether or not they successfully implemented a technique for either one or multiple weeks. Ultimately, this resulted

85

in an indicator of technique mastery that only captured information regarding where participants were at the intervention's end. Specific recommendations for how to advance these indicators are presented later.

This research also suffers from a limitation related to the unavailability of valid and reliable information on participants' baseline levels of academic achievement (K. Ahuna, personal communication, July 7, 2009). Although this study investigated the possibility of differential intervention effectiveness by participant race/ethnicity and gender to provide evidence of its ability to tackle extant gaps in Bachelor's degree-completion rates, these subject variables were assumed to represent prior educational experiences and academic achievement histories. That is, the target constructs in these analyses were not actually race/ethnicity or gender but rather participants' levels of academic achievement. The use of these subject variables in this manner is furthermore a limitation of this study.

The author must note that it was not expected that this intervention would be differentially effective owing simply to participants' skin color, area of origin or anatomical makeup. However, the presentation of findings in terms of participant characteristics such as race/ethnicity and gender did permit conclusions that could easily be interpreted with respect to obtained statistics regarding graduate rate inequities. Notwithstanding a difference in operational definitions, these proxies did provide convergent evidence of findings by Schapiro & Livingston (2000), who reported that growth in both active and dynamic self-regulation did not differ by achievement (i.e., grade point average) level.

Timing of data collection. The next limitation involves the timing of posttest data collection. The data for this study were collected on two occasions from participants in class, three months apart, at the beginning and end of the academic semester during which they

participated in the intervention. It is important to note that the pretest was administered before any course content was delivered and the posttest was administered after all course content was received and all self-assessment meetings had been completed. Justifiably, posttest data were collected at the immediate end of the course in order to ensure a sufficient number of respondents for evaluation purposes. Consequently, any sustained effect of the intervention was not investigated.

The ultimate goal of a metacognitive intervention such as this is the high road, forward-reaching transfer of domain-general knowledge and skills that can be called upon in students' subsequent academic endeavors. In the present intervention, this is attempted by purposefully having students diversify the contexts in which they apply the prescribed techniques. While Livingston's (1990) qualitative dissertation study examined subsequent cognitive strategy use with previous intervention participants, its focus was limited to active self-regulatory control. Follow-up quantitative data collection is clearly warranted to examine the longevity[27] of any observed growth in both active *and* "dynamic" self-regulation. Furthermore, a quantitative follow-up might serve to corroborate findings by Livingston (2000) that what is gained, in terms of active self-regulation, is sustained beyond the immediate end of the intervention. The reader will recall that Livingston (2000) also made a call for such a study.

Indicating increased self-regulation on an instrument administered at the end of an intervention that is identical[28] to that given at the beginning can also be problematic. That is, the increases observed during a repeated administration of the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992) could be owed to a testing effect.

---

[27] This is not a concern specific to investigations of this particular intervention. Perkins and Grotzer (1997) previously noted the lack of longitudinal investigations regarding the effects of metacognitive and similar interventions.

[28] There were very minor differences between the assessments administered at the beginning and end of the intervention.

This phenomenon (i.e., a testing effect) occurs when simply being administered some instrument more than once results in a demonstrated increase in the target of measurement. Along these lines, Fraenkel and Wallen (2006) note that a testing effect can also arise from the interaction between being administered a pretest and participating in an intervention. In this study, it is possible that administering the DALI-R prior to the intervention alerted participants of what would be targeted by it, making them perhaps more sensitive or receptive to the subsequent treatment. To control for the threat of a testing effect, Fraenkel & Wallen (2006) recommend modifying a study's design. They specifically recommend certain experimental, quasi-experimental or factorial designs. One quasi-experimental design (i.e., matching-only pretest-posttest control group) as well as a more complex, factorial design (without randomization) that could be feasibly employed in future research will be discussed later.

Fidelity of implementation. Various scholars discuss the need for any evaluation of an intervention to consider the extent to which it was implemented as planned—referred to as implementation integrity or (henceforth) fidelity (Gresham, 1989; Dane & Schneider, 1998; Moncher & Prinz, 1991; Gresham & Gansle, 1993). Importantly, doing so can provide evidence as to whether any non-significant results were observed due to either a conceptually poor intervention or instead poor implementation of an intervention that was otherwise valid. Alternatively, as in the present case, collecting information related to implementation fidelity can provide evidence that some statistically significant effect (e.g., growth in active or "dynamic" self-regulation) was in fact owed to the intervention itself and not exogenous factors. Lastly, collecting information related to implementation fidelity can also provide information as to what changes occur in the outcomes when various prescribed intervention elements are varied.

This study assumed that all of those individuals involved in implementing the intervention did so with equivalent fidelity. However, the possibility exists that there was variability in implementation fidelity by a number of players. For example, the reader will recall that one portion of the intervention is wholly implemented by an undergraduate peer monitor. This was neglected in this study. There is likely variability between peer monitors that could influence the amount of growth observed throughout the course of the intervention. Furthermore, the undergraduate peer monitors for each course section are also supervised by separate graduate teaching assistants, who could also introduce variability into how the intervention is implemented. This was also not considered. Future studies should include more elaborate data collection related to implementation fidelity in order to strengthen claims regarding a causal relationship between intervention participation and growth in participants' self-regulatory processes.

Clues regarding the target of such efforts and additional data collection can be found in the literature. In a review of the extent to which researchers assessed implementation fidelity in intervention studies, Dane and Schneider (1998) considered a number of factors, including but not limited to: the use of training manuals, the training of facilitators, and the supervision of implementers. The authors also examined whether information related to various other aspects of implementation fidelity was collected during the evaluation. Specifically, they considered adherence, quality of delivery, responsiveness, program differentiation and exposure. This later element of implementation fidelity was clearly targeted by analyses reported earlier, although this literature was not reviewed a priori.

Other findings provide information that can be used to promote fidelity of implementation. For example, consistent supervision of program implementers was shown to be

related to implementation fidelity in a program for latchkey children (Peterson, et al, 1988). Next, Dane and Schneider's (1998) review found that outcomes were improved when implementers adhered to manuals and protocols. Also, trained observers have been shown to be more successful in explaining intervention outcomes than are implementers when providing ratings; Dane and Schneider (1998) hypothesize that the former are less prone to demonstrate social desirability response biases. Lastly, after conducting reviews similar to Dane and Schneider's (1998), Moncher and Prinz (1991) and Gresham and Gansle (1993) both recommend that clearly operationally defining a treatment should be the first step in any investigation considering implementation fidelity.

This study made two preliminary attempts at investigations along these lines. First, this study considered teacher variability by examining differences in the growth outcomes. No differences were observed and it was essentially concluded that there was not differential implementation (e.g., in quality of delivery) by instructor. It is noteworthy that the course content delivered by the two instructors was highly scripted, which is consistent with recommendations in the literature (e.g., Moncher & Prinz, 1991; Gresham & Gansle, 1993). Without having collected information related to the various aspects of instructor implementation fidelity, however, it remains unclear from where this null effect arose. For example, instructors could have differentially implemented various elements of the intervention but still fostered the same amount of growth. Alternatively, one instructor could have implemented the intervention as prescribed, while the other could have made modifications yet still fostered the same amount growth. There are countless other possibilities about which the author could speculate in the absence of implementation fidelity data.

This study also considered the extent to which participants were exposed to the intervention via a set of quantitative indicators used for grading purposes. Some of the limitations of these indicators were discussed earlier. Moreover, these indicators provided information regarding intervention exposure that was expressed more in terms of *quantity* than *quality*. While they were employed as indicators of intervention exposure, they were not created on the basis of what actually constitutes exposure to the intervention (i.e., its constituent components). This constitutes another important limitation of these indicators. In contrast, indicators more aligned with what actually constitutes exposure to the various aspects of this intervention would certainly be more preferable.

The limitations of these two approaches can be addressed by future research if implementation fidelity is more rigorously considered. First, consistent with recommendations by both Moncher and Prinz (1991) and Gresham and Gansle (1993), the treatment (i.e., intervention) should be clearly operationally defined. This will allow for systematic documentation of the extent to which various intervention elements are implemented by the various players (e.g., instructors, peer monitors, and graduate teaching assistants) involved in doing so. Despite using training manuals, training, and supervision, this would allow the researcher to examine whether the intervention was actually implemented with fidelity by the various individuals involved. Subsequently, these data can be used to explain growth in the investigated outcomes; this could provide information as to which components of the intervention are more or less effective in fostering growth in participants' self-regulatory processes.

For participants' intervention exposure, future investigations might consider ratings made by the peer monitors with whom they work or ratings by others involved (or not) with the

intervention; the latter might be preferable for reasons discussed earlier by Dane and Schneider (1998). These exposure ratings for intervention participants should also align with the operational definition of the treatment. Alternatively, video or audio transcripts of weekly self-assessment meetings could be coded with respect to pre-defined criteria by trained observers to yield yet other process exposure indicators that could be used to explain observed growth (Gresham & Gansle, 1993). Consistent with assertions by Dane and Schneider (1998), and others, it may be possible to refine the estimate of this treatment effect by accounting for such factors.

Data and design. Still other limitations of this study are related to its data and design. The first limitation is owed to the size of the sample from which data were collected for the analyses. The reader will recall that the statistical analyses reported here largely produced null effects. Because of the size of the sample, failure to achieve statistical significance may be related to insufficient statistical power. That is, if real effects exist but are small in magnitude, a larger sampling might have meant that statistically significant differences could have been demonstrated. Inspection of the descriptive statistics in Table 9 does reveal that there were many mean differences between those groups who were studied, but the sizes of individual sample groupings may have precluded statistical confidence that these differences were not sample flukes. Nevertheless, it is also possible that a larger sample might not have yielded statistically significant effects, as the variability within groups could have been too large.

Because of the particular sample that was employed, this study was not able to investigate whether Hispanic, Native American or Puerto Rican[29] students experience more or less growth in active and dynamic self-regulation than their counterparts. Also, interactive effects

---

[29] These are additional groups recognized at the institution from which these data were collected.

on the outcomes investigated between particular subject variables could not be investigated. Although this study independently investigated whether differential growth in both active and dynamic self-regulation was observed as a function of race/ethnicity and gender, it was unable to include both variables in the same model due to insufficient cell sizes. Importantly, these interactive effects can still be present in the absence of any observed main effects of a particular independent variable (e.g., race/ethnicity or gender). It is possible that such interactive effects exist in these data although there was not sufficient statistical power to allow the analyst to explore them. Future analyses might want to consider interactions between such subject variables in an investigation with a larger sample.

Lastly, the design of this study was non-experimental (i.e., observational) in nature, which severely limited the ability to make any causal inferences whatsoever. While experimental research would be impractical given the nature of the intervention (i.e., an optional undergraduate course), conducting a quasi-experimental investigation would greatly move toward a causal inference regarding its effectiveness. For example, future research could also administer the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992), or another measure, to a comparable group of students who are not enrolled in the intervention. Additionally, collecting data from all participants on covariates that potentially could confound the comparison between these two groups (i.e., statistical matching) would allow the researcher to approximate a true experimental research study. Alternatively, a factorial design with a matched comparison group and data collected regarding additional independent variables of interest such as race/ethnicity and gender could be conducted to address additional research questions, including those regarding possible interactive differential effects of the intervention. Such designs would provide much greater confidence in claims regarding this intervention's

effectiveness in fostering self-regulated learning (SRL) in undergraduate students. They could possibly put to rest specific concerns that observed growth was the result of maturation or a testing effect.

APPENDIX A

Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992)

Active Self-Regulation Items Submitted to Rasch Analysis.

For each section of a unit I study, I make a list of questions and try to answer them

I make a list of possible exam questions and memorize the answers to them

As I read the textbook, I make predictions about the upcoming information

I make an outline of the main ideas I am studying

I start a study session by surveying the chapter summaries and chapter headings

I summarize the information in the textbook or my class notes and review the summaries when I study

When I study, I try to stay on task by continuously rehearsing the main ideas, names, and principles from my class notes or textbook

When I study the textbook or my lecture notes, I underline or highlight important sentences

I organize my class notes to consist mainly of the important concepts, definitions, and relevant examples from class and readings

I locate the definitions of the key terms in the textbook, lecture notes, or a glossary and learn those definitions

Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992)

Dynamic Self-Regulation Items Submitted to Rasch Analysis.

When I wake up in the morning, the first thing that springs to mind is the topic I have been studying in school

I wake up in the morning or the middle of the night with an insight about what I have been studying in school

When I do NOT understand a concept, I ask the teacher or other fellow students for clarification

When I study, what keeps me going is curiosity and interest

When I have a goal in mind, I can picture myself achieving it

Discovering new ideas causes excitement in me

I get so involved in learning that studying feels almost like watching a suspenseful movie

When studying, I cannot help making frequent stops to think about the ideas I am learning

The topics that I study activate many relevant thoughts in my head

Even for the courses that I consider important, studying feels like a chore to me

I can easily see the relationship of what I am studying to my own personal experiences

I find it difficult to keep my mind on the topic that I am studying

I am good at detecting inconsistencies or flaws in the ideas to which I am exposed

When I study, I find myself reading the same information over and over again without remembering much of what I read

I find it difficult to get as much meaning as I would like from what I study

I have a hard time concentrating when I am studying

I get a lot of satisfaction when I solve conflicts or inconsistencies in the information I encounter

I have trouble relating the different sections of the chapter I study

I have insights about what I learn during my study sessions

When I am doing things outside school, I spontaneously remember the insights that I have gained in school

At the moment that I first experience them, some of my school-related insights feel very much like pleasant surprises to me

REFERENCES

Ahuna, K. H., & Tinnesz, C. G. (2006). *Methods of Inquiry: Applied Critical Thinking* (2nd ed.). Dubuque, IA: Kendall Hunt Publishing Company.

Ahuna, K. H., Tinnesz, C. G., & VanZile-Tamsen, C. M. (2009). *Methods of Inquiry: Using critical thinking to retain students.* Manuscript submitted for publication.

Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4)*,* 561-573.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225,* 82-90.

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.). *The Oxford handbook of memory* (pp. 77-92). New York: Oxford University Press.

Block, J. H., & Burns, R. B. (1976). Mastery learning. *Review of Research in Education*, *4*¸ 3-49.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cao, L., & Nietfeld, J. L. (2007). College Students' Metacognitive Awareness of Difficulties in Learning the Class Content Does Not Automatically Lead to Adjustment of Study Strategies, *Australian Journal of Educational & Developmental Psychology, 7*, 31-46.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more. *American Psychologist, 60*(2), 149-160.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23-45.

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

Fraenkel, J. R., & Wallen, N. E. (2006). *How to Design and Evaluate Research in Education* (6th ed.). New York, NY: McGraw-Hill.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*(1), 37-50.

Gresham, F. M., & Gansle, K. A. (1993). Treatment integrity of school-based behavioral intervention studies: 1980-1990. *School Psychology Review, 22*(2), 254-273.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of Learning Skills Interventions on Student Learning: A Meta-Analysis. *Review of Educational Research*, 66(2), 99-136.

Iran-Nejad, A. (1990). Active and dynamic self-regulation of learning processes. *Review of Educational Research, 60*(4), 537-602.

Iran-Nejad, A., & Chissom, B. S. (1992). Contributions of active and dynamic self-regulation to learning. *Innovative Higher Education, 17,* 125-136.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. American Educational Research Journal, 29, 303-325.

Knowles, E. S., & Nathan, K. T. (1997). Acquiescent responding in self-reports: Cognitive style or social concern. *Journal of Research in Personality*, 31, 293-301.

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.

Liming, D., & Wolf, M. (2008). Job outlook by education, 2006-16. *Occupational Outlook Quarterly, 52*(3), 1-29.

Linacre, J. M. (2002). Optimizing rating scale effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Livingston, J. A. (2000). *A qualitative study of university students' post-course strategy use.* Unpublished doctoral dissertation, University at Buffalo, The State University of New York.

Martinez-Pons, M. (2002). Parental influences on children's academic self-regulatory development. *Theory Into Practice, 41,* 126-131.

Merbitz, C., Morris, J., & Grip, J. C. (1989). Ordinal scales and foundations of misinference. *Archives of Physical Medicine and Rehabilitation, 70,* 308-312.

McCombs, B. L., & Marzano, R. J. (1990). Putting the self in self-regulated: The self as agent in integrating skill and will. *Educational Psychologist, 25,* 51-70.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review, 13,* 247-26.

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25,* 3-53.

Office of Academic Planning & Budget (2008). *Common Data Set 2008-2009.* Buffalo, New York: University at Buffalo.

Pajares, F. (2003). William James: Our father who begat us. In B. Zimmerman and D. Schunk (Eds.). *Educational psychology: A century of contributions.* Mahwah, NK: Lawrence Erlbaum Associates.

Parsad, P., & Lewis, L. (2003). Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000 (NCES 2004-010). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Perkins, D. N. & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52*(10), 1125-1133.

Peterson, L., Mori, L., Selby, V., & Rosen, B. N. (1988). Community interventions in children's injury prevention: Differing costs and benefits. *Journal of Community Psychology, 16,* 188-204.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MLSQ). *Educational and Psychological Measurement*, *53*, 801–813.

Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., et al. (2008). The Condition of Education 2008 (NCES 2008-031). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45,* 269-286.

Perry, N. E., & Drummond, L. (2002). Helping young students become self-regulated researchers and writers. *The Reading Teacher, 56,* 298-310.

Reeves, T. D. (2009). *An Application of the Rasch Model to the Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992).* Unpublished manuscript, University at Buffalo, The State University of New York.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching Students to Generate Questions: A Review of the Intervention Studies. *Review of Educational Research*, *66*(2), 181-221.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory & Cognition, 12*(3), 432-444.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs* (report from the Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association.

Scruggs, T. E., & Mastropieri, M. A. (1988). Acquisition and transfer of learning strategies by gifted and nongifted students. *Journal of Special Education, 22*(2), 153-166.

Schapiro, S. R., & Livingston, J. A. (2000). Dynamic self-regulation: The driving force behind academic achievement. *Innovative Higher Education, 25*(1), 23-35.

Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling, 3,* 25-40.

Smith, E. V. (2001). Evidence for the Reliability of Measures and Validity of Measure Interpretation: A Rasch Measurement Perspective. *Journal of Applied Measurement, 2*(3), 281-311.

Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2008). *Digest of Education Statistics 2007* (NCES 2008-022). National Center for Education Statistics, Institute of Education Sciences, U. S. Department of Education. Washington, DC.

Sternberg, R. J. (2009). *Cognitive Psychology*. Belmont, CA: Wadsworth.

Tabachnick, B.G., & Fidell, L. S. (1983). Using multivariate statistics. New York: Harper.

Tinnesz, C. G., Ahuna, K. H., & Keiner, M. (2006). Toward College Success: Internalizing Active and Dynamic Strategies. *College Teaching¸ 54*(4), 302-206.

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Research. *Review of Educational Research, 45*(1), 89-125.

Uguroglu, M. E., & Walberg, H. J. (1979). Motivation and achievement: A quantitative synthesis. *American Educational Research Journal, 16*(4), 375-389.

Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., and Tobin, R. (2004). The Condition of Education 2004 (NCES 2004-077). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

VanZile-Tamsen, C. M. (2007). *The impact of Methods of Inquiry on retention, graduation rates, and grades*. Buffalo, New York: University at Buffalo, Office of Institutional Analysis.

Wilson, M. (2005). Constructing measures: an item response theory approach. Mahwah, NJ: Lawrence Erlbaum Associates.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J., Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum.

Winne, P. H. (2005). Key issues in modeling and applying research on self-regulated learning. *Applied Psychology: An International Review, 54,* 232-238.

Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., and Tobin, R. (2004). The Condition of Education 2004 (NCES 2004-077). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Wright, B. D., & Linacre, J. M. (1989). Observations are always ordinal; Measurements, however, must be interval. *Archives of Physical Medicine and Rehabilitation, 70*, 857-860.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3-17.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41,* 64-70.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*, 845–862.

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

Zimmerman, B. J., Martinez Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, *80*, 284–290.

Table 1

Frequencies and Percentages for All Categorical Variables

	Frequency	Percent
Teacher		
Dr. Analytical	129	69.4
Dr. Global	57	30.6
Gender		
Male	96	51.6
Female	90	48.4
Race/ethnicity		
Black	22	11.8
Hispanic	10	5.4
Asian or Pacific	21	11.3
Islander		
White	113	60.8
Other	1	.5
Unknown	19	10.2
Motivation[a]		
It was mandatory	2	1.1
My advisor	25	13.6
recommended it		
I needed to	70	38.0
improve my GPA		
I needed an	26	14.1
elective		
The course	61	33.2
content seemed		
interesting		

Note. Valid percentages are presented.
[a]Two participants did not indicate their primary reason for enrolling in the course.

Table 2

Final DALI-R Active Self-Regulation Item Statistics

Entry	Item	Raw score	Count	Item measure	S.E.	Infit	Outfit	Point-biserial corr.
10	generate questions	625	371	1.32	.10	-1.6	-1.6	.66
2	predict questions	654	372	1.02	.10	1.2	1.1	.60
5	predict upcoming information	700	371	.52	.10	-2.1	-1.4	.56
9	outline main ideas	727	371	.24	.10	1.4	1.2	.61
8	survey readings	764	371	-.15	.10	2.4	1.9	.56
7	Summarize	787	370	-.41	.10	-1.1	-1.3	.62
1	organize notes	804	370	-.59	.10	-.3	-.4	.59
3	continuously rehearse	811	372	-.63	.10	-.9	-.2	.58
4	locate definitions	876	372	-1.32	.10	.2	.2	.47

Note. DALI-R = Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992). Infit and outfit mean square fit statistics not included owing to editorial constraints; both were acceptable for all nine items.

Table 3

Summary of Rating Scale for DALI-R Active Self-Regulation Items after Category Re-
Structuring

Category	Observed count	Average measure	Infit mean square	Outfit mean square	Threshold calibration
1	644	-1.68	1.00	.99	-
2	1984	.05	.96	.97	-1.91
3	694	1.70	1.02	1.02	1.91

Note. DALI-R = Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992)

Table 4

Overall Instrument Quality for Active Self-Regulation Items as a Function of Rasch Analysis Stage

	ISI	IR	PSI	PR	Number of misfitting items[a]	Range	PCA factor eigenvalues[b]
Initial analysis	8.66	.99	2.04	.81	1	1.26	1.8, 1.6
After removal of misfitting item	8.90	.99	1.96	.79	0	1.31	1.7, 1.6
After category re-structuring[c]	7.79	.98	1.61	.72	0	2.55	1.6, 1.5

Note. PCA = Principal components factor analysis. ISI = Item separation index. IR = Item reliability. PSI = Person separation index. PR = Person reliability.
[a]Misfitting was defined as more than two unacceptable fit statistics. [b]Only eigenvalues for factors larger than 1.4 are shown. [c]Category re-structuring was conducted secondary to removal of the misfitting item.

Table 5

Final DALI-R Dynamic Self-Regulation Item Statistics

Entry	Item	Raw score	Count	Item measure	S.E.	Infit	Outfit	Point-biserial corr.
2	morning or middle of night insight	542	371	1.98	0.11	1.1	1.1	0.48
3	ask teacher or students	828	372	-0.94	0.1	-1	-1	0.54
4	curiosity and interest	758	369	-0.29	0.1	-0.5	-0.6	0.62
5	picture achieving goal	904	370	-1.78	0.1	0.6	0.6	0.5
7	suspenseful movie studying	525	372	2.2	0.11	-0.5	-1.8	0.6
8	stop frequently	721	371	0.11	0.1	-1.1	-1.3	0.42
10	studying a chore	645	372	0.89	0.1	0.7	0.6	0.51
13	detecting inconsistencies	746	369	-0.18	0.1	-1.8	-1.7	0.46
14	re-reading without remembering	758	371	-0.26	0.1	1.4	1.3	0.46
15	not enough meaning	739	371	-0.07	0.1	1.2	1.2	0.4
16	hard time concentrating	704	371	0.28	0.1	1.7	1.5	0.49
17	satisfied by resolving inconsistencies	866	371	-1.35	0.1	0.7	0.5	0.51
20	remember insights	769	370	-0.39	0.1	-0.4	-0.6	0.58
21	insights pleasantly surprising	751	371	-0.19	0.1	-2.1	-2.1	0.62

Note. Infit and outfit mean square fit statistics not included owing to editorial constraints; both were acceptable for all nine items.

Table 6

Summary of Rating Scale for DALI-R Dynamic Self-Regulation Items after Category Re-Structuring

Category	Observed count	Average measure	Infit mean square	Outfit mean square	Threshold calibration
1	1150	-1.73	1.00	1.00	-
2	3017	-.01	.93	.90	-1.80
3	1024	1.45	1.04	1.04	1.80

Note. DALI-R = Dynamic and Active Learning Inventory Revised (DALI-R; Iran-Nejad & Chissom, 1992)

Table 7

Overall Instrument Quality for Dynamic Self-Regulation Items as a Function of Rasch Analysis Stage

	ISI	IR	PSI	PR	Number of misfitting items[a]	Range	PCA factor eigenvalues[b]
Initial analysis	11.49	.99	2.55	.87	9	2.07	4.0, 2.2, 1.7
After category structuring[c]	10.06	.99	2.21	.83	6	4.14	3.3, 2.1, 1.5
After removal of six items	10.11	.99	1.83	.77	1	4.01	3.1, 1.6
After removal of a seventh item	10.41	.99	1.76	.76	0	3.98	2.9, 1.5

Note. PCA = Principal components factor analysis. ISI = Item separation index. IR = Item reliability. PSI = Person separation index. PR = Person reliability.
[a]Misfitting was defined as more than two unacceptable fit statistics. [b]Only eigenvalues for factors larger than 1.4 are shown. [c]Category re-structuring was conducted prior to removal of the misfitting item.

Table 8

Descriptive Statistics for All Continuous Variables

	Range	Minimum	Maximum	Mean	Standard error	Standard deviation	Variance
Exposure	10.00	0.00	10.00	8.94	0.16	2.12	4.49
Attendance	3.50	2.50	6.00	5.68	0.04	0.56	0.32
Commitment	8.50	3.50	12.00	10.89	0.12	1.60	2.55
Practice	37.40	7.60	45.00	40.94	0.37	5.04	25.42
Final review	10.00	0.00	10.00	8.47	0.12	1.65	2.72
DALI-R active initial status	762.72	68.67	831.39	454.83	10.13	137.47	18897.07
DALI-R active final status	851.98	68.67	920.652[a]	545.97	10.24	138.96	19309.66
DALI-R dynamic initial status	556.62	178.12	734.74	466.97	8.15	111.21	12367.70
DALI-R dynamic final status	834.96	167.52	1002.48	533.02	8.79	119.93	14383.77
DALI-R active growth	1145.03	-382.31	762.72	91.14	13.07	177.33	31445.62
DALI-R dynamic growth	1217.05	-424.74	792.31	66.05	11.25	153.42	23538.85

Note. DALI-R = Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992)
[a]The maximum statistic was 1054.22 prior to removing two person measures imputed by Bond&FoxSteps.

Table 9

Means and Standard Deviations for DALI-R Scale Score and Growth Variables by Subgroups

	DALI-R active initial status		DALI-R dynamic final status		DALI-R active initial status		DALI-R dynamic final status		DALI-R active growth		DALI-R dynamic growth	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Gender												
Male	436.8	153.1	466.8	114.5	525.5	142.4	526.8	111.6	88.7	199.7	60.0	143.7
Female	474.1	116.3	467.1	108.2	567.8	132.5	539.6	128.5	93.7	151.0	72.5	163.7
Race/ethnicity[a]												
Black	492.8	134.8	500.7	108.5	605.3	136.6	579.3	109.8	112.5	172.1	78.7	159.7
Hispanic[b]	407.6	220.7	478.3	153.5	508.8	205.2	522.5	120.7	101.2	311.6	44.2	196.1
Asian or Pacific Islander	414.9	150.8	472.8	80.7	574.8	117.6	546.6	99.5	159.9	142.1	73.8	109.1
White	462.8	126.1	466.4	109.0	538.0	128.7	521.1	120.6	75.2	162.2	54.7	142.3
Teacher												
Dr. Analytical	451.7	138.0	464.1	111.7	536.7	138.1	535.3	108.3	84.9	174.9	71.3	130.3
Dr. Global Motivation	461.7	137.2	473.5	110.7	566.6	139.8	527.8	143.7	104.9	183.5	54.3	196.9
It is mandatory[c]	471.9	31.8	468.9	19.9	583.3	62.4	565.0	76.7	111.4	30.5	96.1	56.8

My advisor recommended it	431.2	142.3	440.3	115.8	550.6	131.0	542.6	97.5	119.4	175.2	102.3	131.8
I needed to improve my GPA	435.8	143.1	438.4	108.9	565.7	147.2	528.9	134.8	129.9	197.7	90.5	167.9
I needed an elective	472.4	113.7	474.5	89.6	492.3	123.3	529.4	95.4	19.9	149.8	54.9	123.5
The course content seemed interesting	478.5	141.0	507.9	112.7	544.3	140.1	535.1	124.9	65.8	159.5	27.2	154.0

Note. DALI-R = Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992)
[a]Descriptive statistics for Other (*n* = 1) and Unknown (*n* = 19) response categories not shown. [b]Descriptive statistics for Hispanic response category based on only ten cases. [c]Descriptive statistics for "It is mandatory" response category based on only two cases.

Table 10

Bivariate Pearson Product-Moment Correlations between DALI-R Scale Score and Growth Variables

	DALI-R active initial status	DALI-R dynamic initial status	DALI-R active final status	DALI-R dynamic final status	DALI-R active growth	DALI-R dynamic growth
DALI-R active initial status	1.000	.528**	.177*	.082	-.637**	-.319**
DALI-R dynamic initial status	-	1.000	-.042	.120	-.442**	-.631**
DALI-R active final status	-	-	1.000	.540**	.646**	.449**
DALI-R dynamic final status	-	-	-	1.000	.359**	.694**
DALI-R active growth	-	-	-	-	1.000	.599**
DALI-R dynamic growth	-	-	-	-	-	1.000

Note. DALI-R = Dynamic and Active Learning Inventory Revised (Iran-Nejad & Chissom, 1992)
*p < .05. **p < .01

Table 11

Growth in Active Self-Regulation by Initial Status

	Active self-regulation initial status		Active self-regulation final status		Active self-regulation growth			
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	t	Growth in standard deviation units[a]
First quartile	290.15	72.83	541.37	146.52	251.22	147.86	12.369***	1.71
Second quartile	428.43	22.60	505.89	116.45	77.44	117.39	4.438***	0.67
Third quartile	513.58	22.42	551.15	140.83	37.57	143.16	1.837	-
Fourth quartile	645.01	66.59	594.45	140.12	-50.57	146.97	-2.093*	-0.36

[a]Mean differences in growth for each group were divided by respective final status standard deviations.

*$p < .05$. ***$p < .001$

Table 12

Growth in Dynamic Self-Regulation by Initial Status

	Dynamic self-regulation initial status		Dynamic self-regulation final status		Dynamic self-regulation growth			
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	t	Growth in standard deviation units[a]
First quartile	348.35	56.95	510.60	114.16	162.25	140.63	9.230***	1.42
Second quartile	442.06	14.32	527.79	111.67	85.73	110.90	4.304***	0.77
Third quartile	507.68	20.95	540.59	89.08	32.90	94.27	2.443*	0.37
Fourth quartile	618.64	46.66	562.23	157.20	-56.40	158.14	-2312*	-0.36

[a]Mean differences in growth for each group were divided by respective final status standard deviations.
*$p < .05$. ***$p < .001$

Figure 1
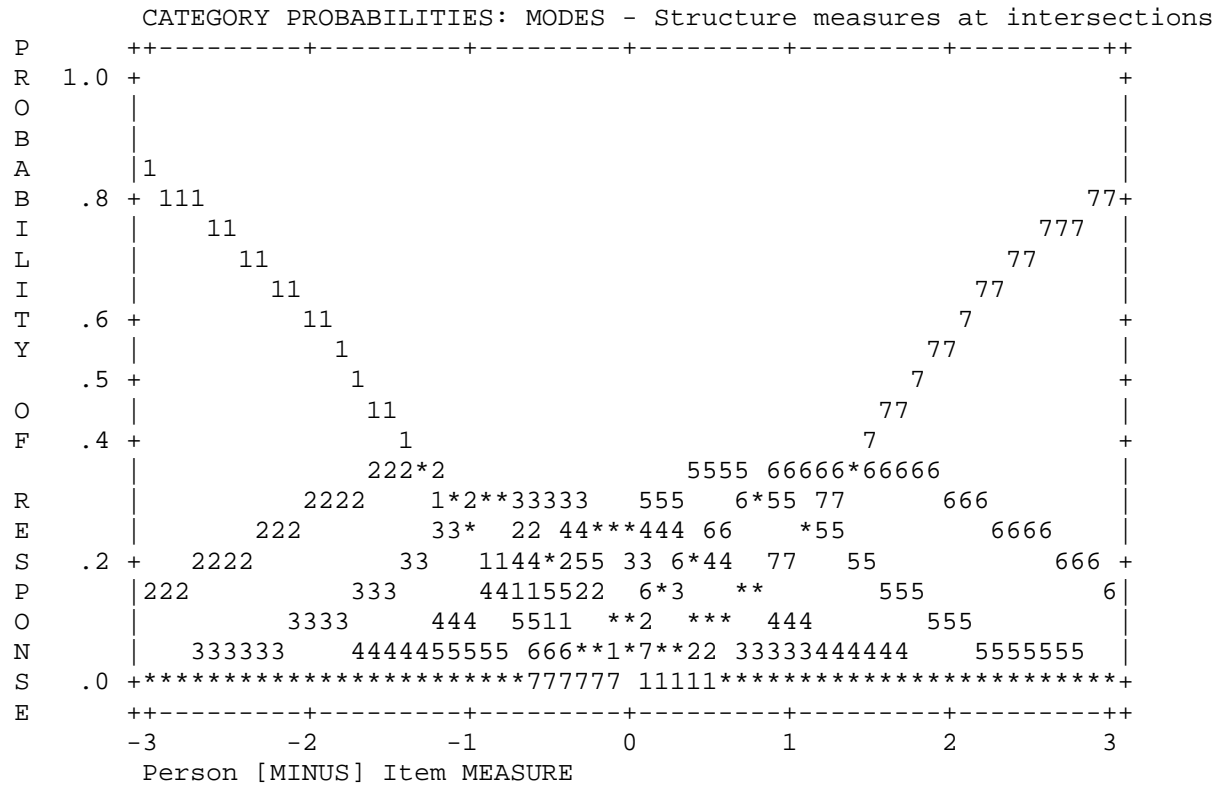
```
          CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P       ++---------+---------+---------+---------+---------+---------++
R   1.0 +                                                              +
O       |                                                              |
B       |                                                              |
A       |1                                                             |
B    .8 + 111                                                        77+
I       |    11                                                   777  |
L       |      11                                                77    |
I       |       11                                             77      |
T    .6 +         11                                          7        +
Y       |          1                                       77          |
     .5 +            1                                     7           +
O       |           11                                  77             |
F    .4 +             1                                7               +
        |             222*2                 5555 66666*66666           |
R       |           2222     1*2**33333   555   6*55 77       666      |
E       |        222           33*   22 44***444 66    *55        6666 |
S    .2 +     2222          33   1144*255 33 6*44  77   55          666 +
P       |222          333     44115522  6*3    **          555         6|
O       |          3333     444   5511  **2  ***   444         555     |
N       |    333333     4444455555 666**1*7**22 33333444444    5555555  |
S    .0 +********************777777 11111*************************+
E       ++---------+---------+---------+---------+---------+---------++
         -3        -2        -1         0         1         2         3
          Person [MINUS] Item MEASURE
```

Category response curves (CRCs) for the initial 7-point rating scale active self-regulation item design submitted to Rasch analysis.

Figure 2

```
          CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P         ++------+------+------+------+------+------+------+------++
R   1.0 +                                                            +
O         |                                                          |
B         |1                                                       3|
A         | 111                                                 333 |
B    .8 +    11                                               33     +
I         |       11              2222222222              33        |
L         |         11          22        22          33           |
I         |          1         22          22        3             |
T    .6 +         11      2                    2      33             +
Y         |          1  22                      22  3               |
     .5 +            1*                            *3                +
O         |          22 1                        3 22                |
F    .4 +          2    1                        3    2              +
          |         22      11                  33      22           |
R         |       22          11              33          22         |
E         |      22            1              3            22        |
S    .2 +      22              11      33              22    +
P         | 222                 111 333                       222 |
O         |2                     33*11                          2|
N         |                   3333333      1111111                  |
S    .0 +3333333333333333333                   1111111111111111111+
E         ++------+------+------+------+------+------+------+------++
          -4      -3     -2     -1      0      1      2      3      4
          Person [MINUS] Item MEASURE
```

Category response curves (CRCs) after a category re-structuring to yield a 3-point rating scale active self-regulation item design.

Figure 3
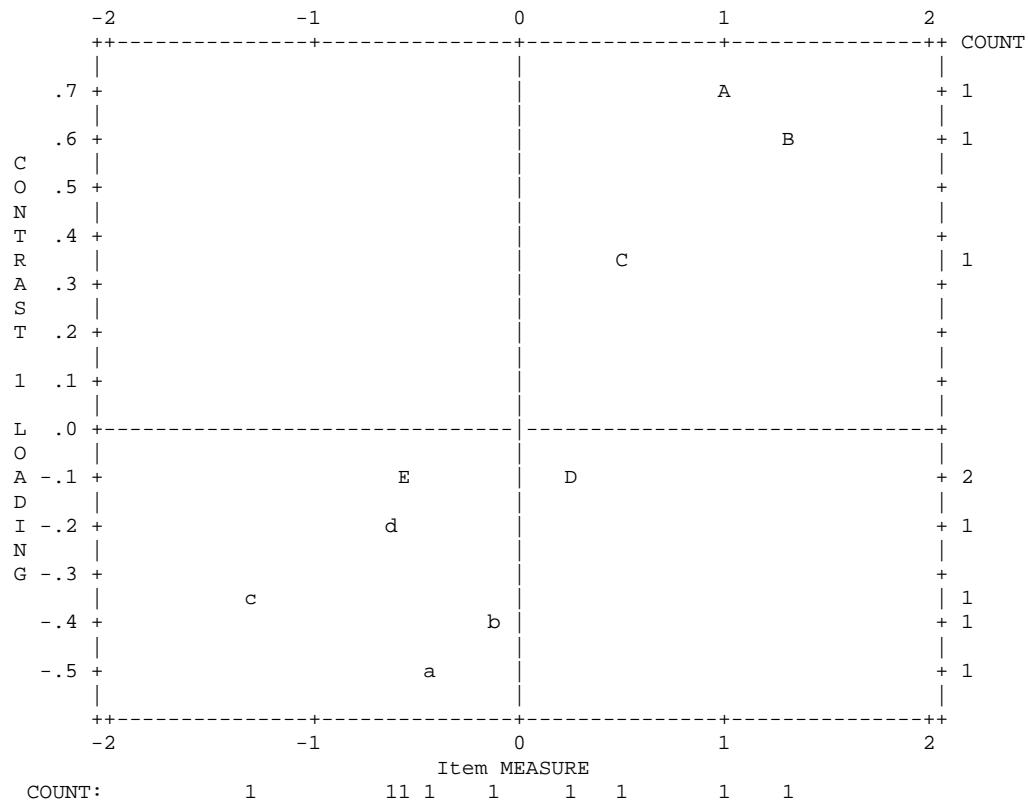
```
                -2              -1               0               1               2
                ++---------------+---------------+---------------+---------------++ COUNT
                |                                |                               |  |
          .7 +  |                                |               A               |  + 1
                |                                |                               |  |
          .6 +  |                                |                     B         |  + 1
     C          |                                |                               |  |
     O    .5 +  |                                |                               +
     N          |                                |                               |  |
     T    .4 +  |                                |                               +
     R          |                                |        C                      |  | 1
     A    .3 +  |                                |                               +
     S          |                                |                               |  |
     T    .2 +  |                                |                               +
                |                                |                               |  |
     1    .1 +  |                                |                               +
                |                                |                               |  |
     L    .0 +--+--------------------------------|-------------------------------+
     O          |                                |                               |  |
     A   -.1 +  |              E                 |    D                          +  2
     D          |                                |                               |  |
     I   -.2 +  |              d                 |                               +  1
     N          |                                |                               |  |
     G   -.3 +  |                                |                               +
                |          c                     |                               |  | 1
         -.4 +  |                           b    |                               +  1
                |                                |                               |  |
         -.5 +  |                      a         |                               +  1
                |                                |                               |  |
                ++---------------+---------------+---------------+---------------++
                -2              -1               0               1               2
                                         Item MEASURE
         COUNT:          1            11 1      1        1   1         1       1
```

Active self-regulation items loading on the first possible sub-dimension indicated by principle components factor analysis (PCA).

Figure 4

```
    -2              -1               0               1               2
     ++--------------+---------------+---------------+--------------++ COUNT
      |                               |                              |
 .6 + |                               |    D                         + 1
      |                               |                              |
 .5 + |                          b    |                              + 1
C     |                               |                              |
O  .4 + |                             |                              +
N     |                               |                              |
T  .3 + |                             |            B                 + 1
R     |                               |                              |
A  .2 + |                             |                              +
S     |              a                |                              | 1
T  .1 + |                             |                              +
      |                               |                              |
2  .0 +-------------------------------|------------------------------+
      |                               |                              |
L -.1 + |                             |  C                           + 1
O     |                               |       A                      | 1
A -.2 + |                             |                              +
D     |                               |                              |
I -.3 + |          c                  |                              + 1
N     |                               |                              |
G -.4 + |                             |                              +
      |                               |                              |
  -.5 + |                             |                              +
      |              E                |                              | 1
  -.6 + |              d              |                              + 1
     ++--------------+---------------+---------------+--------------++
    -2              -1               0               1               2
                             Item MEASURE
   COUNT:          1        11 1    1      1    1      1    1
```
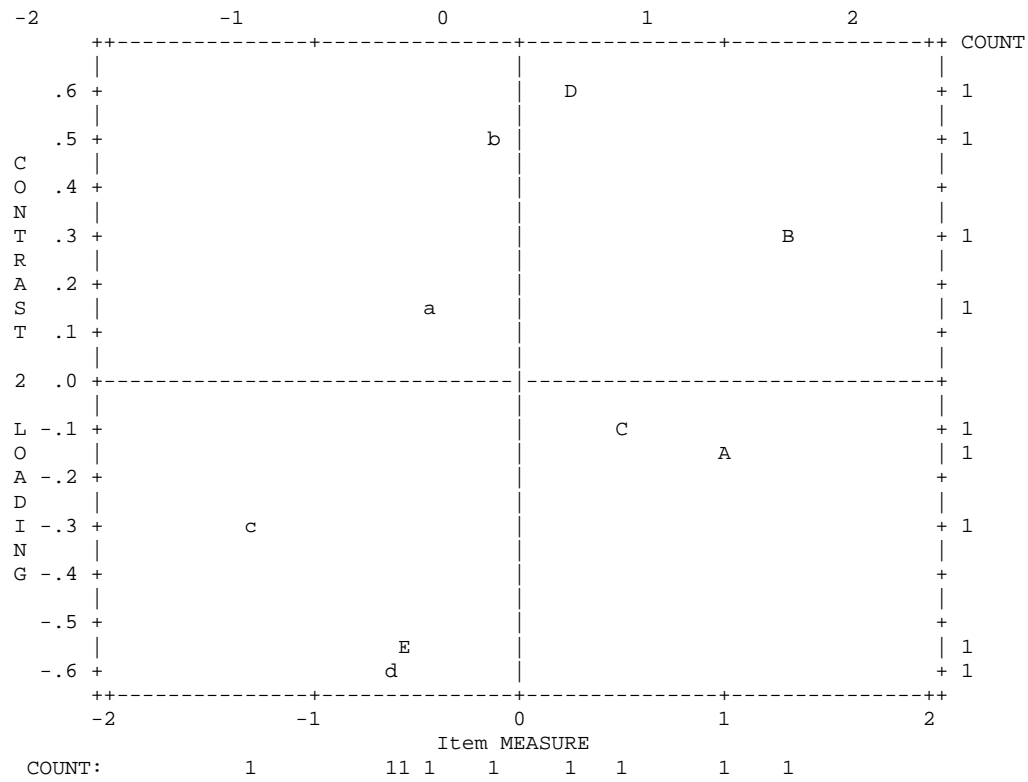
Active self-regulation items loading on the second possible sub-dimension indicated by principle components factor analysis (PCA).

Figure 5

```
Persons MAP OF Items
              <more>|<rare>
    5           .  +
                   |
                   |
                   |
                   |
                .  |
                   |
    4              +
                   |
                   |
                   |
              .#   |
    3           T+
              .#  |
                   |
                   |
                   |
              .##  |
    2              +
            .##### |
                   |T
                  S|
            .##### |   A10PRE generate questions
                   |
    1              +   A02PRE predict questions
           .###### |S
                   |
          ######## |   A05PRE predict upcoming informtion
                   |
                   |   A09PRE outline main ideas
    0 .############ M+M
                   |   A08PRE survey readings
              .    |   A07PRE summarize
         .######### |
                   |   A01PRE organize notes
                   |   A03PRE continuously rehearse
           .###### |S
   -1              +
                   |
          ##### S|   A04PRE locate definitions
              .    |
                   |T
            .###   |
   -2              +
                   |
             .##   |
              .    |
                   |
              .  T|
   -3              +
                   |
             #     |
                   |
                   |
                   |
   -4              +
                   |
              .    |
                   |
                   |
                   |
   -5              +
              <less>|<frequ>
EACH '#' IS 5.
```

Person-item map for nine active self-regulation items showing the relative locations of persons (left) and items (right) on the same hypothetical scale (*y*-axis).

Figure 6

Persons & Items

Measures

t Infit Zstd

Bubble chart representing the relative locations of 372 persons (grey) and nine active self-regulation items (black) on the same hypothetical scale (*y*-axis), in terms of their infit standardized fit statistics (*x*-axis) and standard errors (size).

Figure 7

```
        CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P       ++---------+---------+---------+---------+---------+---------++
R   1.0 +                                                             +
O       |                                                             |
B       |                                                             |
A       |11                                                           |
B    .8 +  111                                                      7+
I       |     11                                              777 |
L       |       11                                             77  |
I       |        1                                           77    |
T    .6 +         11                                        7       +
Y       |          1                                      77        |
     .5 +           1                                    7          +
O       |            11                                77           |
F    .4 +             1                           6666*             +
        |              1                  55   666  77 66666        |
R       |         2222222***3333334444*555   **55  7        666     |
E       |      2222        33122244433*5 44466    **5           666  |
S    .2 +     222         333    14422 55 33 66444  7    55        666+
P       |2222         333      4411 5*2    6*3      **       555       |
O       |        333       444    5*1  ***   3**7  444       5555      |
N       |   3333333    44444 55555 66**11 ****   3333   444444   555555 |
S    .0 +*********************777777*1111***********************+
E       ++---------+---------+---------+---------+---------+---------++
        -3        -2        -1         0         1         2         3
         Person [MINUS] Item MEASURE
```

Category response curves (CRCs) for the initial 7-point rating scale dynamic self-regulation item design submitted to Rasch analysis.

Figure 8

```
          CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P      ++---------+---------+---------+---------+---------+---------++
R  1.0 +                                                            +
O      |                                                            |
B      |                                                            |
A      |                                                            |
B   .8 +                                                            +
I      |111                     2222222222                       333|
L      |   11               2222          2222               33     |
I      |      11          222                222           33       |
T   .6 +        11       22                    22        33         +
Y      |         11    22                        22    33           |
    .5 +           1*2                              2*3             +
O      |         222 11                          33 222             |
F   .4 +        22       11                    33       22          +
       |      22           11                33           22        |
R      |    22               11            33               22      |
E      |222                    111      333                    222  |
S   .2 +                         111  333                          +
P      |                          111 333                           |
O      |                          3333*1111                         |
N      |                 333333333        111111111                 |
S   .0 +333333333333333333                      11111111111111111111+
E      ++---------+---------+---------+---------+---------+---------++
        -3        -2        -1         0         1         2         3
           Person [MINUS] Item MEASURE
```

Category response curves (CRCs) after category re-structuring to yield a 3-point rating scale dynamic self-regulation item design.

Figure 9

```
          -2            -1             0            1             2            3
          ++-----------+-----------+-----------+-----------+-----------++ COUNT
           |                                   |                        |  |
      .7 + |                               B|  A                    + 2
           |                               C  |                      |  | 1
      .6 + |                                   |                      +
  C        |                                   |                      |  |
  O   .5 + |                                   |       D              + 1
  N        |                                   |                      |  |
  T   .4 + |                                   |                      +
  R        |                                   |                      |  |
  A   .3 + |                                   |                      +
  S        |                                   |                      |  |
  T   .2 + |                                   |                      +
           |                                   |                      |  |
  1   .1 + |                                   |                      +
           |                                   |                      |  |
  L   .0 + +-----------------------------------|----------------------+  |
  O        |                                   |                      |  |
  A  -.1 + |                        E          |                      + 1
  D        |              F                    |              G       | 2
  I  -.2 + |    f                              |           g          + 2
  N        |                                   |                      | 1
  G  -.3 + |                        e          |                      +
           |                                  |d                      | 1
     -.4 + |         c                         |                      + 1
           |                        b          |                      | 1
     -.5 + |                                   |                      +
           |                        a          |                      | 1
          ++-----------+-----------+-----------+-----------+-----------++
          -2            -1             0            1             2            3
                                   Item MEASURE
     COUNT:    1      1       1   11121 1   1        1          1   1
```

Dynamic self-regulation items loading on the first possible sub-dimension indicated by principle components factor analysis (PCA).

Figure 10

```
        -2          -1           0           1           2           3
        ++-----------+-----------+-----------+-----------+-----------++ COUNT
    .7 +                          |                          g        +  1
       |                          |                                   |
    .6 +                          |                                   +
       |                          |                                   |
C   .5 +                          |                       G           +  1
O      |                          |                                   |
N   .4 +                          |                                   +
T      |                          |           D                       |  1
R   .3 +                          |                                   +
A      |                          |                                   |
S   .2 +              bE          |                                   +  2
T      |                          |                                   |
    .1 +                          |                                   +
2      |                  a       |                                   |  1
    .0 +-------------------------|---A-------------------------------+  1
L      |            F             |d                                  |  2
O  -.1 +                          |                                   +
A      |                      B   |                                   |  1
D  -.2 +                          |                                   +
I      |                  C       |                                   |  1
N  -.3 +                          |                                   +
G      |        f                 |                                   |  1
   -.4 +                          |                                   +
       |                          |                                   |
   -.5 +          c               e |                                 +  2
       |                          |                                   |
       ++-----------+-----------+-----------+-----------+-----------++
        -2          -1           0           1           2           3
                               Item MEASURE
        COUNT:   1    1     1   11121 1  1        1         1  1
```
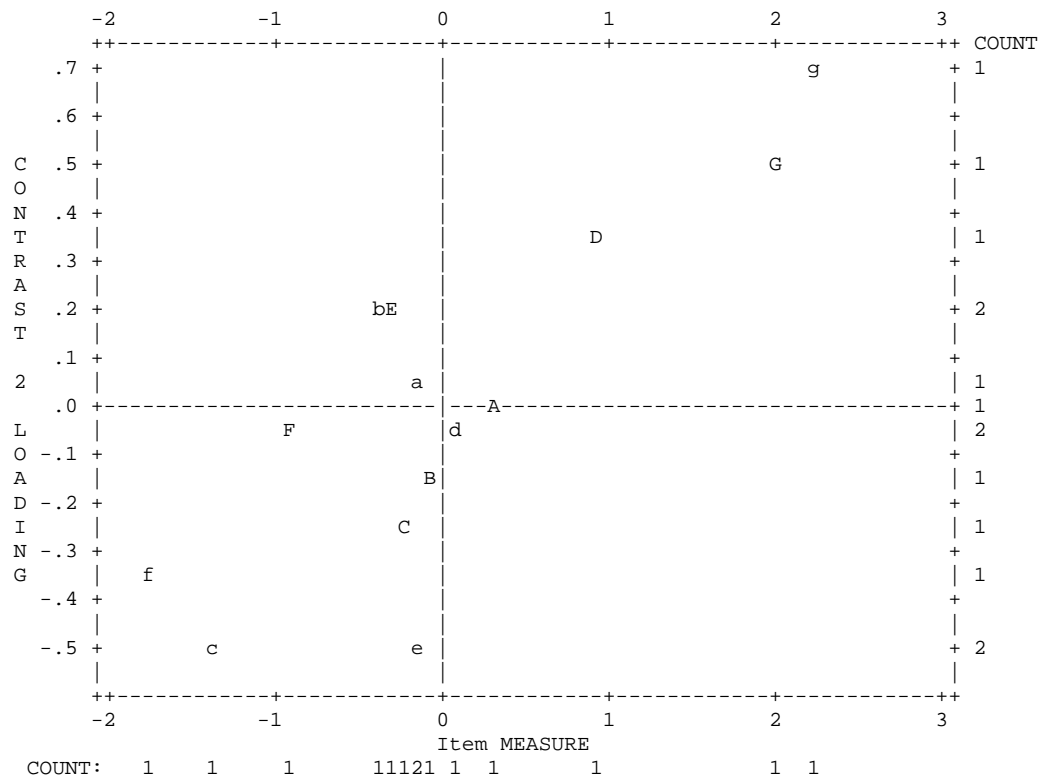
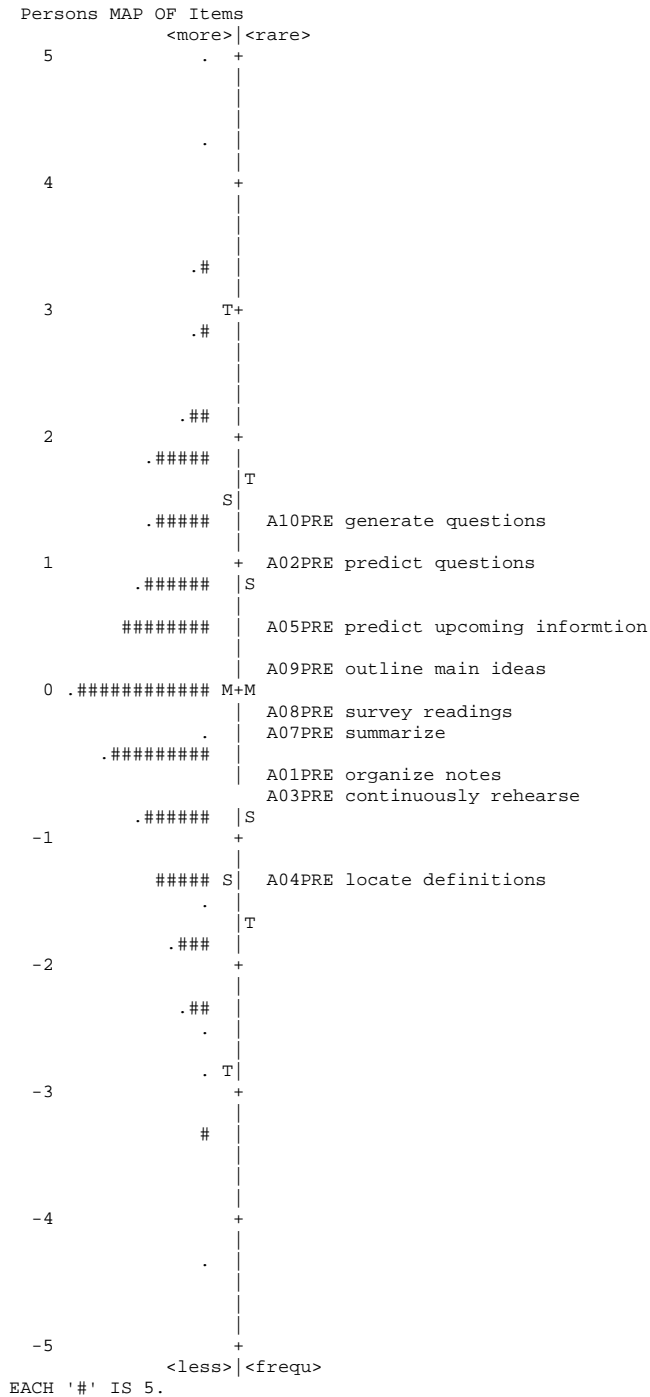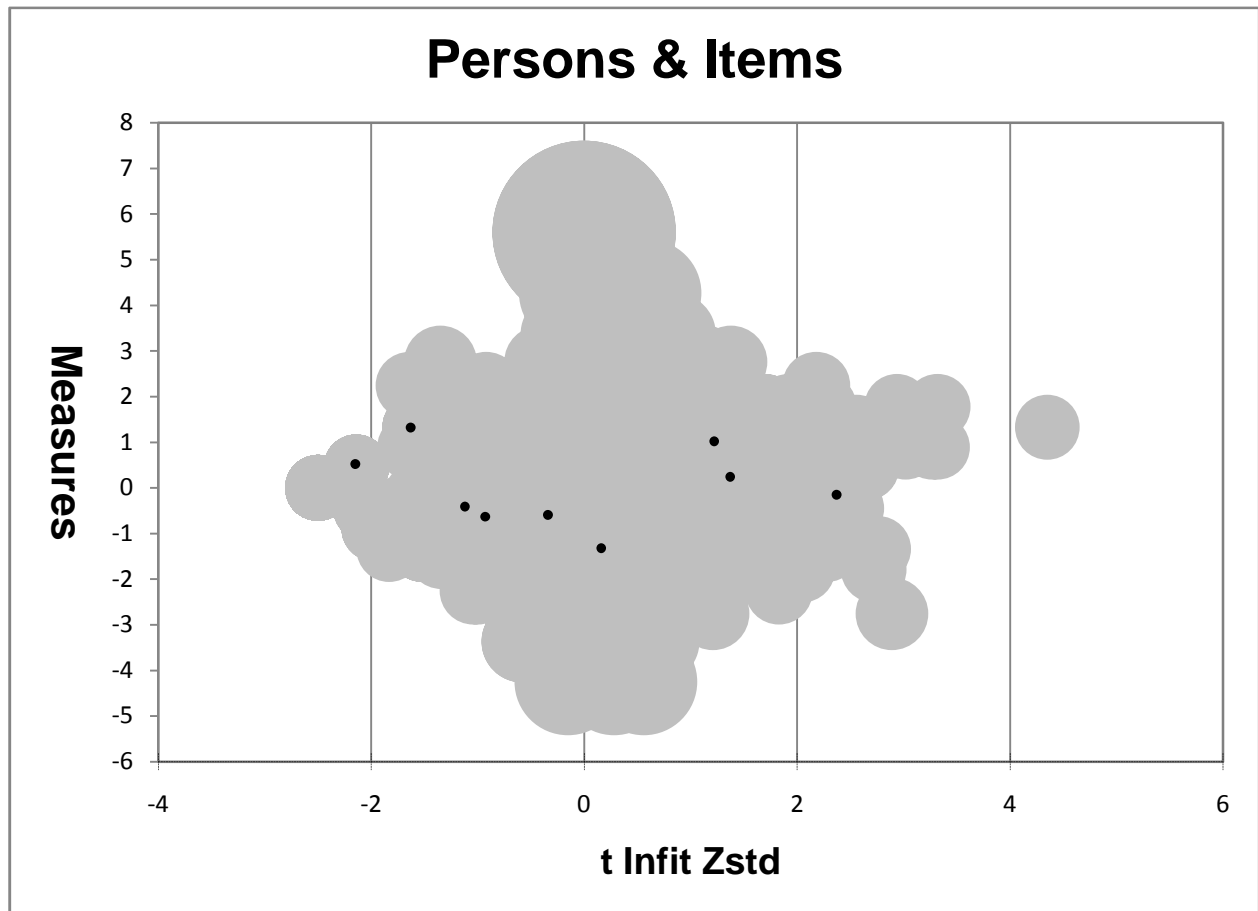Dynamic self-regulation items loading on the second possible sub-dimension indicated by principle components factor analysis (PCA).

Figure 11

```
     Persons MAP OF Items
            <more>|<rare>
   5            .  +
                  |
                  |
                  |
                  |
                  |
   4              +
                  |
                  |
                  |
                  |
                  |
   3            .  +
                  |
              #   |
                  |
               T|
            .##  |T D07PRE  suspensful movie stuying
   2         .#  +  D02PRE  morning or middle of night insight
             .   |
           .###  |
                  |
          .####  |
          .###  S|
   1              +S
         .#####   |  D10PRE  studying a chore
                  |
        .######   |
       .#######   |  D16PRE  hard time concentrating
                  |  D08PRE  stop frequently
   0 .############  +M D15PRE  not enough meaning
            .  M|  D13PRE  detecting inconsistencies
                  |  D21PRE  insights pleasantly surprising
       .#######   |  D04PRE  curiosity and interest
                  |  D14PRE  re-reading without remembering
                  |  D20PRE  remember insights
      .########   |
                  |
        .#####    |
  -1          .   +S D03PRE  ask teacher or students
       .#######   |
        .####   S|  D17PRE  satisfied by resolving inconsistencies
                  |
        .####     |
                  |  D05PRE  picture achieving goal
  -2         .#   +
                |T
            .#   |
             T|
            .#   |
                  |
  -3          .   +
                  |
              .   |
              .   |
                  |
                  |
  -4              +
            <less>|<frequ>
  EACH '#' IS 4.
```
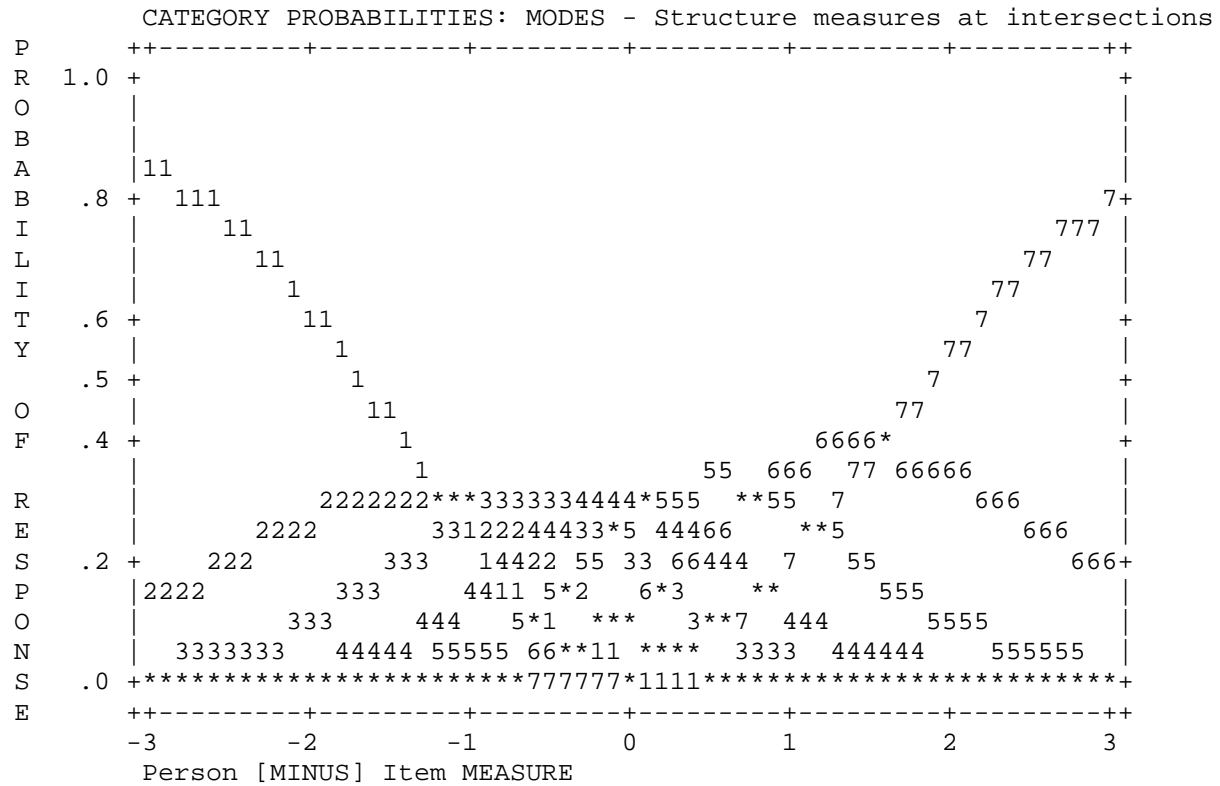
Person-item map for fourteen dynamic self-regulation items showing the relative locations of persons (left) and items (right) on the same hypothetical scale (*y*-axis).

Figure 12

Persons & Items

Measures (y-axis)

t Infit Zstd (x-axis)

Bubble chart representing the relative locations of 372 persons (grey) and fourteen dynamic self-regulation items (black) on the same hypothetical scale (*y*-axis), in terms of their infit standardized fit statistics (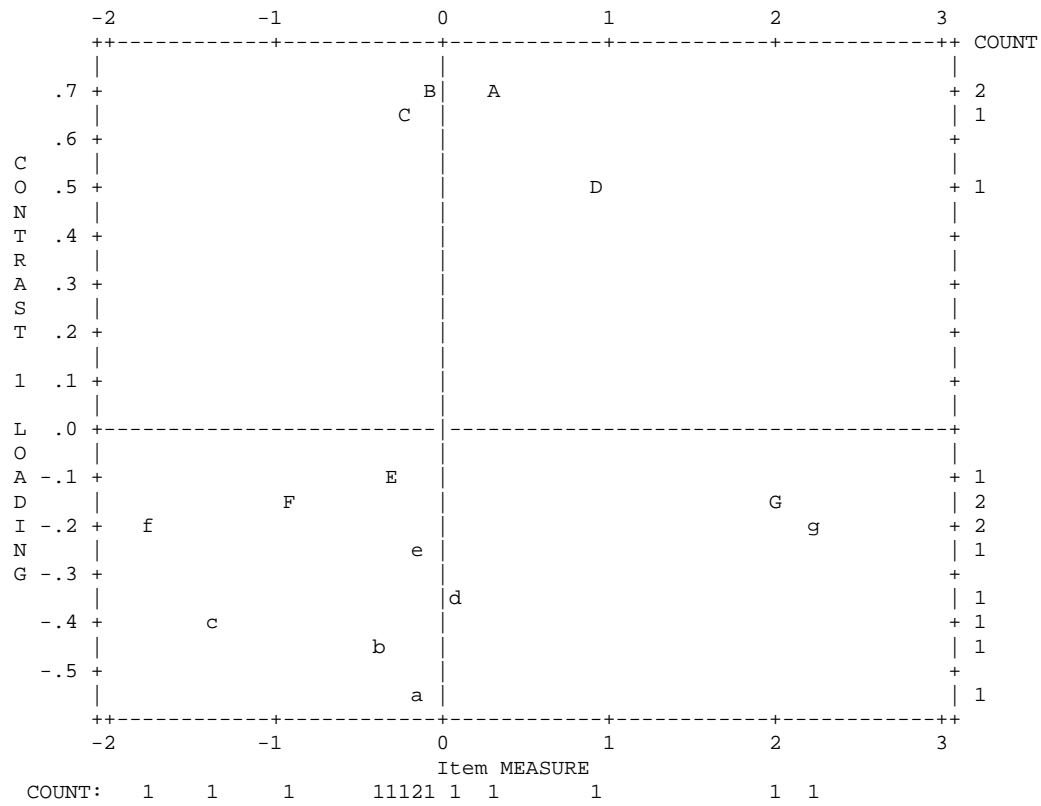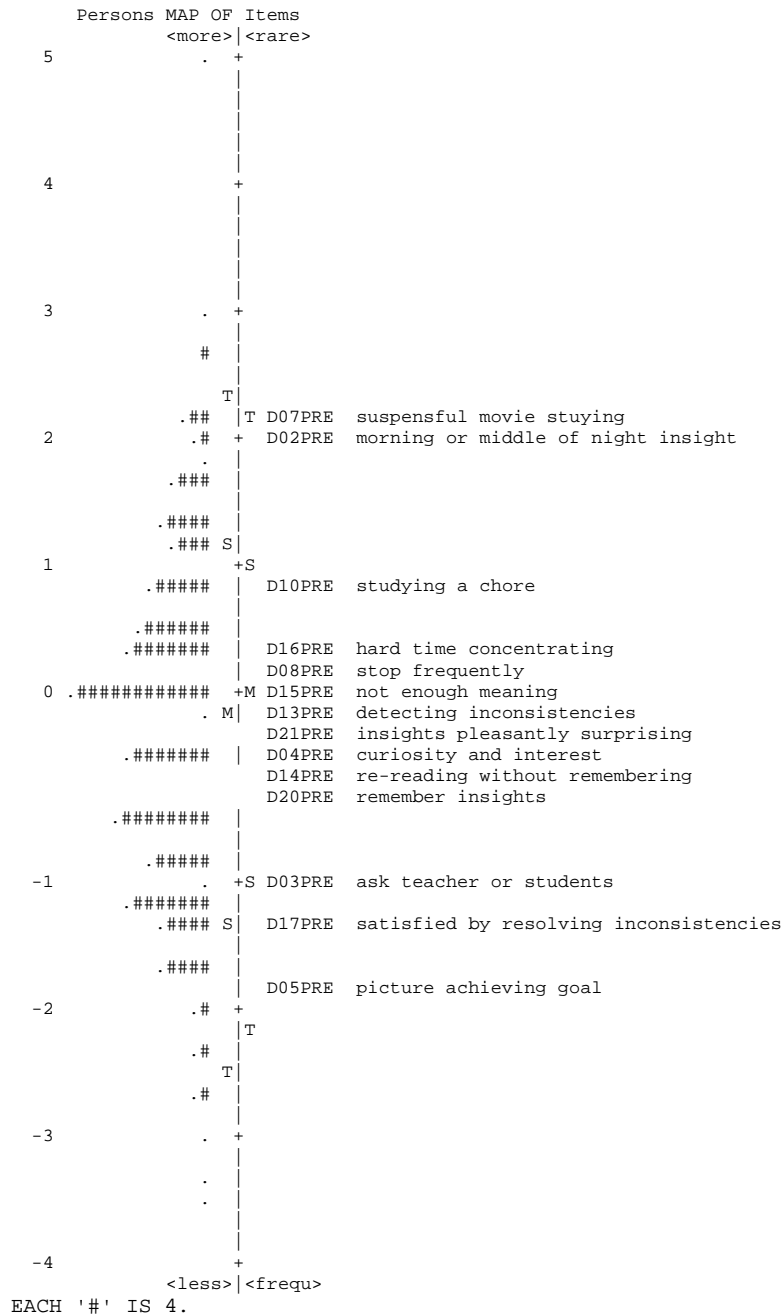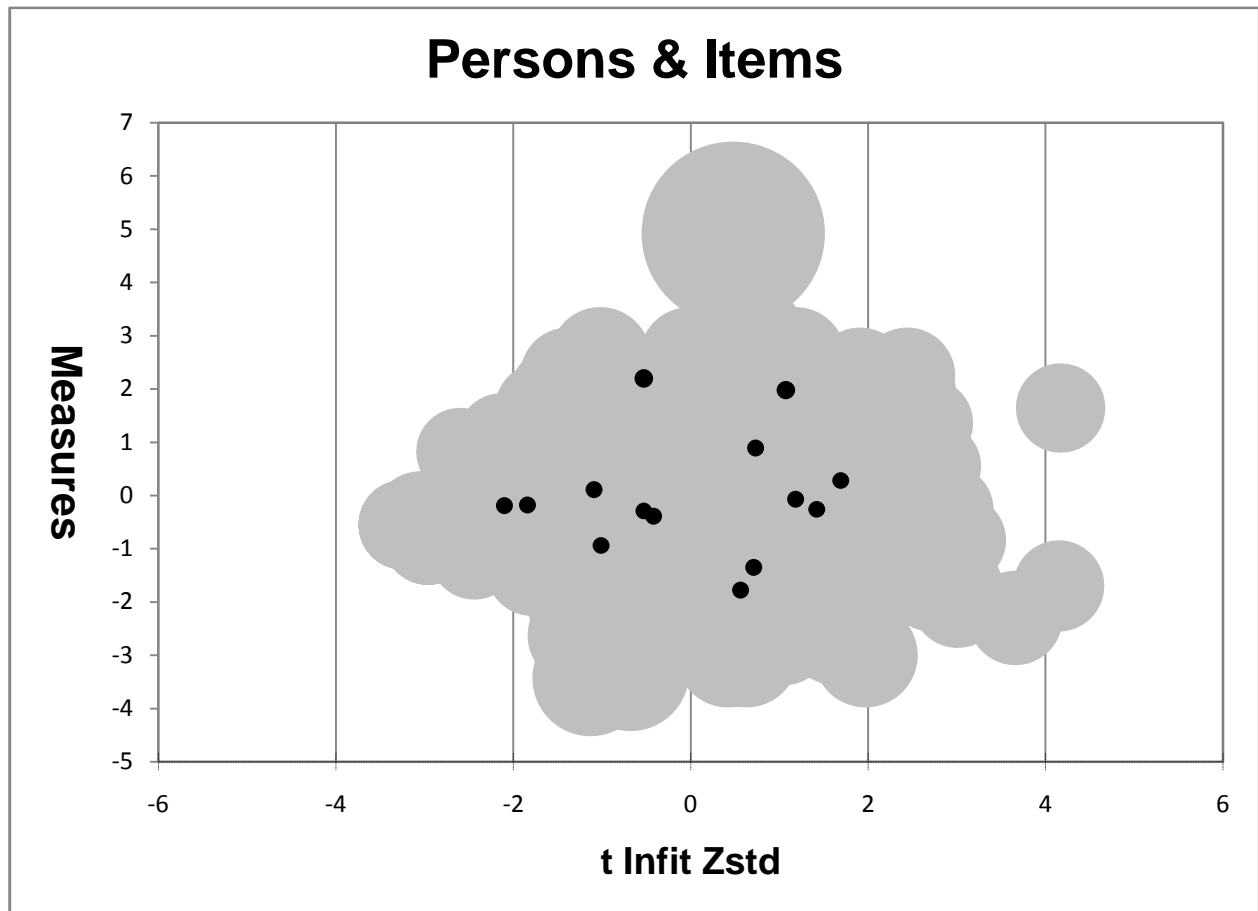*x*-axis) and standard errors (size).