UNDERSTANDING THE INTERACTION BETWEEN STUDENTS' THEORIES OF INTELLIGENCE AND LEARNING ACTIVITIES

by

Soniya Gadgil

B. Arch, University of Pune, 2000

M.S., University of Pittsburgh, 2009

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Soniya Gadgil

It was defended on

July 30, 2014

and approved by

Christian Schunn, PhD, Professor, Department of Psychology, University of Pittsburgh

John Levine, PhD, Professor, Department of Psychology, University of Pittsburgh

Vincent Aleven, PhD, Associate Professor, Human-Computer Interaction Institute, Carnegie Mellon University

Dissertation Advisor: Timothy Nokes-Malach, PhD, Associate Professor, Department of Psychology, University of Pittsburgh

UNDERSTANDING THE INTERACTION BETWEEN STUDENTS' THEORIES OF INTELLIGENCE AND LEARNING ACTIVITIES

Soniya Gadgil, PhD

University of Pittsburgh, 2014

Understanding the Interaction Between Students' Theories of Intelligence and Learning Activities

Soniya Gadgil, PhD

University of Pittsburgh, 2014

Understanding the interaction between students' motivation and instructional factors is critical for extending current cognitively based frameworks of learning, and can have important practical applications. Two laboratory experiments were conducted to explore how students' implicit theories of intelligence interact with different types of learning activities. The ICAP framework by Chi (2009) organizes learning activities into passive, active, constructive, and interactive activities representing an increasing order of effectiveness. In Experiment 1, participants' theories of intelligence were manipulated to be either entity or incremental, and the learning activity — inventing a formula to calculate variability, was manipulated to be constructive (inventing individually) or interactive (inventing collaboratively). It was predicted that individuals would learn procedurally simple aspects of the task better than collaborators regardless of their theory of intelligence, consistent with theories of collaboration and cognitive load. In contrast, while all collaborators were predicted to learn more conceptual knowledge than individuals, students with incremental theories were predicted to benefit more from collaboration than those with entity theories. Results showed that while individuals learned more than collaborators on procedural problems, the predicted interaction between collaboration and theories of intelligence on conceptual problems was not supported. Experiment 2 tested whether different types of constructive activities interacted with students' theories of intelligence to affect learning outcomes. In this experiment, students' theory of intelligence was manipulated to be either incremental or entity, and the type of constructive activity was manipulated to be either

tell-and-practice instruction or invention. Two competing interaction hypotheses were proposed. Hypothesis one was that if invention activities led to more constructive processing, entity theorists would learn more from invention than from tell-and-practice instruction, but incremental theorists would learn equally well from either type of instruction. Hypothesis two was that if invention activities cause off-task behavior and impose excessive cognitive load, then tell-and-practice instruction would lead to better learning for entity theorists, however, both types of instruction would be equally effective for incremental theorists. Bayesian model selection provided some support for hypothesis one. Results of the two experiments are discussed in terms of their theoretical and practical significance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I am writing this section on the eve of “Guru Purnima” a day in the Hindu calendar dedicated to honoring spiritual and academic teachers. I am deeply indebted to all those who have been my “guru” for the past few years — my advisor Tim Nokes-Malach, and my committee members, Chris Schunn, John Levine, and Vincent Aleven, who gave me their invaluable feedback and advice during multiple stages of this dissertation. Thanks to past and present members of the Higher Order Cognitive Collective (HOCC) at LRDC for doing inspiring research, for offering thoughtful and constructive comments on different stages of this work, and for being awesome colleagues in general. Thanks also to undergraduate research assistants — Julianna Sincavage, Kenny Holstein, Jonas Kerner, and Alex Mastrolonardo for their help with various tasks, including data collection, data entry, coding, reliability testing, and many others.

I couldn’t have possibly completed this dissertation without the help of various people who became the proverbial village to help me care for my two children during this time. Thanks go to my mother and my mother-in-law who made journeys of 8000 miles so that I could focus on my work, with the assurance that my children were in safe and loving hands. Thanks also to my father, my brother Suyog, sister-in-law Anagha, and members of my extended family for their unwavering encouragement and emotional support. Thanks to Rachel Inghram, nanny extraordinaire to my two children, who brings tremendous commitment and integrity to every task she undertakes. And thanks to my friends Mala, Samay, Kunal, Yulia, Mark, Karina, and

Srini, who have become my family in Pittsburgh, and to countless other friends whom I could always count on.

I am incredibly lucky to be married to Vinod, who has always stood by me like a rock. He encouraged me to keep going even when things seemed really bleak. Thank you for giving me your unwavering support, for being there for me through thick and thin, and for being the most amazing “baba” (dad) to our two children. And finally, my two munchkins Vismay and Shlok — your big smiles and cuddly hugs have helped me get through many a difficult day. I love you!

1.0 GENERAL INTRODUCTION

Over the past few decades, research in the learning sciences has been successful in identifying a number of instructional principles and strategies that promote learning (Winne & Nesbit, 2010). However, the translation of these principles into classroom instruction has not had the same level of success. For example, in a recent review, Dunlosky, Rawson, Marsh, Nathan, & Willingham (2013) examined ten instructional techniques derived from basic research, which were expected to improve learning outcomes. All of these techniques had ample evidence in their favor from laboratory studies. However, only one of the ten techniques was found to be consistently effective when used in educational contexts. Three others led to positive learning outcomes only under particular circumstances, five had insufficient evidence in their favor, and one was negatively related to learning. While conditions can be carefully controlled in laboratories to isolate individual variables and to test their effects on learning, conditions in classrooms are often "noisy," in that they involve several contextual factors and individual difference factors that may interact with cognitive factors. Therefore, in order to develop models of learning that can effectively generalize to classroom environments, it is important that such models incorporate the effect of contextual and individual difference factors, and their interactions with cognitive factors (Pintrich, 2004).

In this dissertation, I will focus on the individual difference variable of student motivation. According to the socio-cognitive theory of achievement motivation (Dweck, 2000;

Dweck & Leggett, 1988), students hold one of two implicit theories of intelligence — an entity theory or an incremental theory. Students who have an entity theory (also called fixed theory or fixed belief) hold that intelligence is a static trait, which remains constant throughout a person's lifetime. Conversely, students who have an incremental theory (also called malleable/growth theory or belief) hold that intelligence is a malleable trait that can be improved through effort and practice. While prior work has examined the relationship between students' theories of intelligence and learning outcomes, not much is known about how they interact with different types of instructional activities. For example, the effect of students' motivational beliefs may be strong enough to influence learning outcomes under different types of learning activities. Alternatively, certain types of learning activities may diminish or override the effects of theories of intelligence, and thereby influence learning outcomes more so than students' motivational beliefs. In order to tease apart the effects of motivational and cognitive factors, it is important to empirically test the competing hypotheses, in order to make more specific claims about applying these theories to educational practice (Nokes-Malach & Belenky, 2011).

The ICAP framework (formerly known as the Active-Constructive-Interactive framework, Chi 2009) provides a taxonomy of learning activities based on students' overt behaviors as categorized into one of four modes: Interactive, Constructive, Active, and Passive. *Active* learning activities such as taking notes during a lecture lead to better learning outcomes than do *passive* activities such as simply listening to a lecture. *Constructive* learning activities such as self-explanation produce better learning compared to *active* activities. *Interactive* activities such as learning collaboratively with a partner or interacting with an intelligent tutoring system are better than both *active* and *constructive* activities. While the ICAP framework does a good job of classifying learning activities and predicting which ones are likely to be effective, it

takes into account only cognitive factors, and neglects to consider motivational factors during learning. Within the ICAP framework, motivational beliefs such as theories of intelligence can potentially interact with instructional activities, which might impact what students from those activities.

As an example, incremental theorists are more likely than entity theorists to engage in productive interactions such as seeking help (Shih, 2007), offering help (Dweck & Bempechat, 1983), and reacting to negative feedback in a constructive manner (Hong, Chiu, Dweck, Lin, & Wan, 1999). Therefore, the prediction of the ICAP framework that interactive activities are better than constructive activities is more likely to be true for students with incremental theories, but students with entity theories may not necessarily benefit as much from interactive activities, because they would be less likely to engage in productive interactions. Presently, no empirical studies have tested how the instructional activities as described in the ICAP framework interact with motivational variables. Testing the predictions of the ICAP framework in relationship with motivational factors will strengthen our understanding of the generalizability of these predictions and understand important boundary conditions.

In this dissertation, I test two predictions of the ICAP framework in relationship with students' motivational factors. In Experiment 1, I compare an interactive activity (collaboratively inventing a formula for calculating variability) with a constructive activity (individually inventing a formula for calculating variability), while manipulating students' theories of intelligence to be either entity or incremental. The goal of the experiment is to test whether interactive activities are uniformly better than constructive activities, or whether they are more effective for students with incremental theories, compared to those with entity theories. In Experiment 2, I compare two kinds of constructive activities — inventing a procedure versus

learning from tell-and-practice instruction, while again manipulating students' theories of intelligence to be either entity or incremental. In this experiment, I test whether certain types of constructive activities differentially benefit students with entity theories and incremental theories. To situate the work, I will first review the ICAP framework and its predictions, followed by a review of existing work on implicit theories of intelligence and their relationship with learning.

1.1 ICAP FRAMEWORK

The ICAP framework proposed by Chi (2009; Chi & Wylie, 2014) makes predictions for the effectiveness of different kinds of learning activities. According to this framework, learning activities can be classified into four hierarchical categories: passive, active, constructive, and interactive activities. Active learning activities are defined as those in which learners are actively engaging in some activity while learning. For example, merely listening to a lecture is a passive activity whereas taking notes while doing so is an *active* activity. Instructors encourage being active as a means to increase engagement with the learning materials. The key difference between passive activities and active activities is that in active activities, learners engage with the materials in a more direct manner compared to in passive activities. However, because active activities do not involve creation of new knowledge through generating inferences and restructuring prior knowledge, they reflect only surface level processing.

Constructive activities are learning activities in which learners engage with learning materials to generate outputs that go beyond the content provided in the materials. As an example, when students self-explain text while reading, they actively construct new knowledge

by relating the text to their prior knowledge, discovering interrelationships between parts of the text, and drawing inferences (Chi, de Leeuw, Chiu, & LaVancher, 1994). Other examples of constructive activities include asking questions (Graesser & Person, 1994), constructing diagrams or concept maps (Horton et al., 1993), comparing and contrasting cases (Gentner, Loewenstein, & Thompson, 2003), making analogies (Novick & Holyoak, 1991), among others. In order to be constructive, a learner has to first actively engage with the materials; therefore, constructive activities necessarily subsume active activities. Because constructive activities lead to creation of new knowledge through transformation of existing knowledge, they require deep processing and engagement with the learning materials.

The next level in ICAP framework is that of *interactive* activities. While engaging in an interactive activity, a learner interacts with another entity, which could be a peer, a teacher, a tutor, or an intelligent tutoring system. Collaborative learning is an example of an interactive activity in which a learner collaborates with a peer during learning. In order to be interactive, learners first need to be active to communicate with their partners. They also need to engage in constructive activities such as explanation, elaboration, justification, question-asking, help-seeking, and so on. Therefore, interactive activities subsume both the active and constructive categories. The key difference between being constructive and interactive is that the goal of interaction is to arrive at a shared understanding of the material, or a "shared mental model" of the situation (Roschelle, 1992). According to the ICAP framework, interactive activities provide students with opportunities to create shared representations, and therefore lead to better learning outcomes compared to active and constructive activities.

While the ICAP framework provides a useful taxonomy for differentiating learning activities, and makes predictions for their effectiveness, it relies largely only on cognitive and

socio-cognitive factors to make these predictions. Given the complex interactions between cognition and motivation, it is critical to take into account factors such as motivation that go beyond "cold cognition" (Pintrich, Marx, & Boyle, 1993). For example, students with entity theories of intelligence may benefit from certain kinds of learning activities, whereas students with incremental theories may benefit from others. A second issue to consider is that each of the three levels of the ICAP framework may consist of subtypes of learning activities. For example, there are a multitude of activities that could fall under the umbrella of "constructive activities," each of which may be differentially effective for learners with different motivational beliefs. For example, learners with entity theories may benefit from one kind of learning activity such as invention, whereas those with incremental theories may benefit from another kind of constructive activity such as learning from worked example. Therefore, to make more specific and fine-grained predictions about learning, it is important to understand how the different levels of the ICAP framework interact with motivational factors.

1.2 IMPLICIT THEORIES OF INTELLIGENCE AND THEIR RELATIONSHIP WITH LEARNING

Consider two college students, Emily and Isabella who are taking an advanced statistics class. They were both straight A students in high school, and had maintained high GPAs up to this point in college. A few weeks into the semester, their instructor handed back their midterm exams, on which they had both struggled. Upon getting a C on the exam, Emily had the following reaction: "I give up! I am just not smart enough for statistics. Maybe I should consider dropping this course?" Isabella also received a C, but she had a different reaction. She thought,

“This is a really hard class; what do I need to do to get better at it? Perhaps I should look into using different studying techniques and strategies.”

The two vignettes described above characterize two “mindsets” or “theories of intelligence” that students can hold. Emily is said to have an entity theory of intelligence, while Isabella is said to have an incremental theory. Theories of intelligence have been an influential construct in research on motivation and learning. According to the socio-cognitive model of achievement motivation (Dweck, 2000; Dweck & Leggett, 1988), students who have an entity theory of intelligence (also called fixed theory or fixed belief) hold that intelligence is a static trait, which remains constant throughout a person’s lifetime. Conversely, students who have an incremental theory (also called malleable/growth theory or belief) hold that intelligence is a malleable trait that can be improved through effort and practice.

The difference between students with entity theories and incremental theories becomes most apparent in the face of a challenge, even when they do not differ on actual intellectual ability. Learners who have an entity theory of intelligence attribute success to inherent traits of intelligence, so when they face a difficulty, they view it as a reflection of their own inferior intellectual abilities and are discouraged by failure (Dweck, 2000; El-Alayli & Baumgardner, 2003). Conversely, learners who have an incremental theory attribute success to effort, so they view challenge as a learning opportunity, work harder on the task, and seek out opportunities to improve their performance (Hong et al., 1999).

Incremental theories of intelligence have been shown to be associated with various adaptive processes and outcomes (see Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013 for a review). Several studies have found that incremental beliefs are associated with high academic achievement (Blackwell, Trzesniewski, & Dweck, 2007; Greene, Costa, Robertson, Pan, &

Deekens, 2010; Jones, Wilkins, Long, & Wang, 2012; Stipek & Gralinski, 1996). This effect has been found to be sustained during academic transition (Blackwell et al., 2007; Henderson & Dweck, 1990) when coursework typically gets more challenging, and has been noted in a variety of domains such as science, math, engineering, physical education among others. It has also been observed in various cultures, for example, Korean (Lim, Plucker, & Im, 2002) and Hispanic cultures (Nichols, White, & Price, 2006).

Students' theories of intelligence predict the kinds of cognitive and behavioral strategies they use during learning. Entity theorists are less likely to use elaboration and critical thinking strategies (Dahl, Bals, & Turi, 2005), metacognitive regulation strategies such as planning and monitoring (Miele & Molden, 2010), and integrating across multiple sources of information (Braasch, Bråten, Strømsø, & Anmarkrud, 2014), while incremental theorists are more likely to use the aforementioned strategies. Entity theorists are also less likely to engage even in surface level processing strategies such as rehearsal (Paulsen & Feldman, 2007), and are more likely to procrastinate (Howell & Buro, 2009). Incremental theorists cope better with stressful situations while entity theorists show disengagement and less adaptive coping behaviors (Doron, Stephan, Boiché, & Scanff, 2009). While incremental theories are associated with a host of adaptive cognitive, affective, and behavioral processes and outcomes, and entity theories are associated with maladaptive ones, not much is known about how they interact with instructional factors during learning.

Although much of the early work on implicit theories of intelligence was conducted with K-12 age populations, there is some evidence that they influence learning processes and outcomes in adults as well. For example, one study found that although entity theorists entered college with higher SAT scores compared to incremental theorists, this did not translate into

higher achievement for entity theorists (Robins & Pals, 2002). Conversely, incremental theories and performance were found to be positively associated (Greene et al., 2010). However, some recent work shows mixed findings with respect to advantages of incremental theories in older adults, in that incremental theories were not associated with better learning in some studies (e.g., Plaks & Chasteen, 2013; B. Simon et al., 2008).

Most studies investigating the relationship between theories of intelligence and learning have been correlational in nature, but some studies have manipulated theories of intelligence to investigate causal relationships (e.g., Bergen, 1991; Hong et al., 1999 study 3). Intervention studies designed to promote incremental beliefs of intelligence have often led to positive outcomes. For example, in a study by Blackwell et al., (2007) with students beginning junior high school, students participated in an intervention that involved reading and participating in discussions about either incremental beliefs or an unrelated topic. Analyses of learning trajectories showed that those who participated in the intervention showed upward trajectories on math achievement, whereas those in the comparison group showed declining trajectories. In another intervention study with undergraduates conducted by Aronson, Fried, and Good (2002), students were asked to write a letter to a younger "pen-pal" advocating the incremental nature of intelligence. Post-intervention, letter writers made significant gains in GPA, reported greater enjoyment of the academic process, and showed greater academic engagement, compared to those who wrote letters unrelated to beliefs of intelligence, or did not write letters.

Not all intervention studies, however, have found the predicted benefits for adopting incremental beliefs. For example, one study investigated the effectiveness of a computer program called Brainology designed to encourage the adoption of incremental theories of intelligence through various activities and quizzes. Although participants were significantly likely to adopt

incremental beliefs upon completion of the intervention, this effect was not found to be sustained at follow-up three months later (Donohoe, Topping, & Hannah, 2012). Another study with college-age students in the domain of programming found no pre to post differences on a programming test performance for students adopting either theory (Simon et al., 2008). In sum, more research is needed to better understand the conditions under which theories of intelligence affect learning. To gain a better understanding of how theories of intelligence affect learning, I will review a process model proposed by Dweck and colleagues (e.g., Dweck & Legett, 1988).

1.2.1 Process model

In earlier conceptualizations of the socio-cognitive model of achievement motivation, Dweck offered the following model:

Table 1. Theories of intelligence, goals, and behavior (from Dweck & Legett, 1988)

Theory of Intelligence	Goal Orientation	Confidence in present ability	Behavior pattern
Entity theory (Intelligence is fixed)	Performance goal (Goal is to gain positive judgments/ avoid negative judgments of competence)	If high →	Mastery-oriented Seek challenge High persistence
		But	
		If low →	Helpless Avoid challenge Low persistence
Incremental theory (Intelligence is malleable)	Learning goal (Goal is to increase competence)	If high →	Mastery-oriented Seek challenge (that fosters learning) High persistence
		Or low ↗	

As seen in table 1, the relationship between theories of intelligence and achievement behaviors was thought to be mediated through the goals that students are likely to adopt in a learning situation (Roedel & Schraw, 1995). Entity theorists are more likely to adopt

performance goals, such that they seek to gain positive judgments or avoid negative judgments of competence (Stipek & Gralinski, 1996). Conversely, incremental theorists are more likely to adopt learning goals such that they seek to increase their own competence regardless of an external point of reference (Dweck & Leggett, 1988; Mangels, Butterfield, Lamb, Good, & Dweck, 2006). Performance goals translate into helpless or self-handicapping behaviors (characterized by avoidance of challenge and low persistence) only when a students' confidence in his or her abilities is low. When confidence is high, even entity theorists show mastery-oriented behaviors (characterized by seeking of challenge and high persistence), which are typical of incremental theorists. Incremental theorists engage in mastery-oriented behaviors regardless of whether they have high or low confidence in their abilities (Dweck, 1986; Elliott & Dweck, 1988).

More recent work has not found consistent relationships between students' implicit theories and goals (see Dupeyrat & Mariné, 2005b for a review) but Dweck and colleagues maintain that entity theories engender performance goals and incremental theories engender mastery goals which lead to differing learning behaviors, and subsequently lead to different learning outcomes (Dweck, 2000; Dweck & Molden, 2005). However, in a departure from the original model (Dweck & Leggettt, 1988), Dweck and colleagues no longer claim that high confidence in abilities can lead to entity theorists adopting mastery goals (Hong, Chiu, & Dweck, 1995). Thus, the revised model can be stated as follows:

Table 2. Revised model of implicit theories and goals

Theory of Intelligence	Goal Orientation		Behavior pattern
Entity theory (Intelligence is fixed)	Performance goal (Goal is to gain positive judgments/ avoid negative judgments of competence)	→	Helpless Avoid challenge Low persistence
Incremental theory (Intelligence is malleable)	Learning goal (Goal is to increase competence)	→	Mastery-oriented Seek challenge (that fosters learning) High persistence

Although work by Dweck and colleagues suggests that incremental theories are associated with optimal learning processes and outcomes, there is some evidence that does not bear out this prediction. In addition to the two aforementioned intervention studies that showed a lack of effect of incremental theories (Donohoe, Topping, & Hannah, 2012 and Simon et al., 2008), a few other studies have reported similar findings. For example, Furnham, Chamorro-Premuzic, and McDougal (2002), in a study with British undergraduates found that theories of intelligence were unrelated to academic performance. A similar lack of effect was reported by Stump, Husman, and Chung (2009), in the context of engineering education. Another study by Niiya, Brook, and Crocker (2010), found that people with incremental theories were not immune to self-handicapping, particularly when a threat to self-esteem was apparent. Some other studies that used path models to understand the relationship between theories of intelligence and achievement outcomes have found no direct relationship between the two variables (e.g., Dupeyrat & Mariné, 2005b; Leondari & Gialamas, 2002). These results suggest that adopting of incremental theories may not be universally adaptive, and more research replicating prior research and defining boundary conditions is necessary.

In sum, although prior work on theories of intelligence suggests numerous advantages for incremental theories of intelligence, it also has several limitations. Most studies have been correlational in nature, and the few that have manipulated theories of intelligence show mixed outcomes. Most of the studies have investigated theories of intelligence in younger (K-12) populations, and the effect has not been found to be as robust in college-age students. Some studies have failed to find the purported benefit for incremental theories. Finally, not much work has looked at how theories of intelligence interact with instructional activities. While students' theories of intelligence have been shown to be associated with learning outcomes, certain types of learning activities may diminish or override the effects of theories of intelligence and influence learning outcomes more so than students' motivational beliefs.

To address some of the limitations of past work on theories of intelligence and achievement, I seek to answer the following questions:

1. Do students with entity theories and incremental theories benefit equally from constructive activities and interactive activities? In Experiment 1, I manipulate participants' theories of intelligence to be either entity or incremental, and the learning activity — inventing a formula to calculate mean deviation to be constructive (inventing individually) or interactive (inventing collaboratively). I predict that on simple tasks such as procedural problems, individuals would learn better than collaborators for students with either theory of intelligence. In contrast, on complex tasks, collaborators would learn more than individuals, however, students with incremental theories would benefit more from collaboration compared to those with entity theories.

2. Do students with entity theories and incremental theories benefit equally from different types of constructive activities? In Experiment 2, I will explore the interaction between students' theories of intelligence and two types of constructive learning activities. Specifically, I will compare student learning under one of two conditions — a tell-and-practice instruction condition, in which students will be given a worked example and asked to solve similar practice problems, and an invention condition in which they will be asked to come up with a solution for an open-ended problem, followed by the worked example. As in Experiment 1, I expect no differences on procedural problems. As for performance on measures requiring deep, conceptual knowledge, I test two competing hypotheses. Hypothesis one is that invention would be more beneficial to entity theorists, because it would encourage them to engage in constructive activities. Incremental theorists are likely to be constructive regardless of condition, so they will learn equally well under either condition. Hypothesis two is that invention activities would lead to impasses that would cause entity theorists to abandon their efforts. They would therefore benefit more from tell-and-practice instruction. Incremental theorists are not likely to be deterred by impasses during invention, so they will learn equally well under either condition.

2.0 INTRODUCTION TO STUDY 1: THEORIES OF INTELLIGENCE IN COLLABORATIVE VERSUS INDIVIDUAL LEARNING

Collaborative learning is "a situation in which two or more people interact to learn or attempt to learn something together" (Dillenbourg, 1999). The collaborating partners are of equal status and there is no explicit assignment of roles such a tutor and a tutee (Cohen, Kulik, & Kulik, 1982). The aim of the interaction is to learn from the collaboration, and learning is assessed in some form of a subsequent posttest. Proponents of collaborative learning view it as the "educational psychology success story of the twentieth century" (D. W. Johnson & Johnson, 2009). However, this enthusiasm for collaborative learning is not universal, and critics of the approach claim that its benefits have been overstated, and that the research on collaboration has not been carefully controlled enough to warrant the claims of its efficacy (Anderson, Reder, & Simon, 1996; Druckman & Bjork, 1994).

Nevertheless, collaborative learning has found an important place in classrooms for its purported cognitive and educational benefits, and several large-scale collaborative learning programs have been implemented in school districts across the United States to improve student learning (Johnson & Johnson, 1994). Instructors believe that encouraging students to learn in groups will lead to better retention and understanding of materials (Lumpe, Haney, & Czerniak, 1998), and improve student motivation (Bossert, 1988). In recent years, research on computer-supported collaborative learning has burgeoned (e.g. Dillenbourg, 1999), in part because

working in a dyad or a group appears to have a distinct efficiency advantage compared to working alone in that it requires less time and fewer instructional resources (Arthur, Day, Bennett Jr, McNelly, & Jordan, 1997).

The ICAP framework by Chi (2009), classifies collaborative learning as an interactive learning activity, and predicts that students who learn with a partner, and engage in constructive interactions would learn better than students who engage in constructive activities individually. However, research on collaborative learning has shown mixed results (see F. Kirschner, Paas, & Kirschner, 2009a for a review), such that some studies have found that groups outperform individuals (e.g., Azmitia, 1988; D. W. Johnson, Johnson, & Smith, 2007), whereas others have found that they perform the same as (Crooks, Klein, Savenye, & Leader, 1998), or in some cases, even worse than individuals (e.g., Leidner & Fuller, 1997; Yetter et al., 2006). These results suggest that advantages of collaboration may depend on various moderating factors, and therefore, more research is necessary to identify the conditions under which collaboration can lead to better learning outcomes compared to learning individually. Next, I will discuss some reasons for the mixed outcomes on collaborative learning.

One of the limitations of current work on collaborative learning is that in many of the studies comparing collaborators and individuals, the outcome measures are not learning specific. The definition of “learning” is often unclear Many studies compare group performance with individual performance on the collaborative task, and show an advantage of collaboration, but future individual performance or learning is not measured (e.g., David W. Johnson, Johnson, & Stanne, 1989). When future individual performance is measured via a posttest, individuals who worked in groups prior to the posttest are sometimes found to perform no differently from those who worked individually (Pociask & Rajaram, 2014). Thus, conflating of learning measures and

performance measures may have led to overstating of the benefits of collaboration, particularly in meta-analytic reviews.

Another issue to consider is that task complexity may play a role in whether collaboration could lead to better outcomes compared to individual learning. Research based on the cognitive load theory suggests that the communication and coordination activities during collaboration impose extra cognitive load in addition to the cognitive load of the learning task itself (F. Kirschner, 2009). For simple tasks, the cognitive resources of an individual are sufficient to complete the task, so the communication and coordination processes of collaboration create a cognitive overhead. However, for complex tasks, the same communication and coordination processes constitute what is called a “germane load” because they are necessary for carrying out the learning task, which means that, an individual could not succeed alone at the task, and so the cost of collaboration may be necessary to potentially achieve success. Thus, according to the cognitive load theory, when the learning task is a complex one that requires integration and synthesis of multiple knowledge components as opposed to retaining simple facts or procedures, it necessarily imposes a high cognitive load on learners, and is therefore learned better collaboratively.

When group members collaborate on a complex learning task, they are able to develop higher quality knowledge representations or schemas by distributing the cognitive load across group members. Such high quality schemas facilitate performance on post-collaboration transfer measures even when tested individually. By contrast, individuals working on the same complex learning task by themselves would spend a majority of their cognitive resources simply memorizing relevant information, and will not have the benefit of extra processing capabilities that collaborators have, which are required for deeper conceptual understanding. As a

consequence, on post-collaboration transfer tests, individuals would perform worse than collaborators, on but on simpler tasks like retention, they would show high performance, because they would have effectively retained information about simple concepts and procedures. Collaborators, however, do not need to remember all information elements individually, because such information can be distributed across multiple working memories, and this can hamper their performance on simple retention tasks. Studies comparing individual performance to group performance when participants had to recall as many information elements as possible after studying them for a certain amount of time support this prediction. Although groups outperform individuals on the number of items recalled, when group performance is compared to the sum of individual scores (i.e., the nominal score), in most cases group performance is inferior to that of the nominal group (Andersson & Rönnberg, 1995; Meudell, Hitch, & Kirby, 1992; Weldon & Bellinger, 1997). In other words, when working together in a group to recall information, individuals recall less than when they work alone.

One study by Kirschner and colleagues (F. Kirschner, Paas, & Kirschner, 2009b) using high school Biology students tested the prediction made by the cognitive load theory that after engaging in a complex learning activity, individuals would perform better than collaborators on retention measures, but collaborators would perform better than individuals on transfer measures. This interaction prediction was confirmed. Kirschner et al. posit that group members were able to deeply process the materials, and interrelate the information elements to construct higher quality schema, leading to higher performance on transfer tasks. Conversely, individuals showed greater efficiency in retaining relevant information, and therefore performed better on retention measures. Similar findings were reported by Gadgil and Nokes-Malach (2012) with undergraduate students on a writing task. In this study, students were provided error-ridden

summaries of journal articles, and they worked either individually or with a partner to detect the errors, and revise the summaries. On a homework assignment where students had to write their own summaries, a significant interaction was found for the type of error (surface versus structural) and collaboration driven largely by the fact that collaborators made significantly fewer structural errors compared to individuals. These results are consistent with the ICAP framework, which predicts that insofar as collaborators are engaging in productive interactions, they would perform better on a future test, even when taking the test individually.

This brings us to the third limitation on current work on collaboration. The extent to which collaborators engage in productive interactions is often dependent on what kinds of motivational beliefs they hold. Very little work has examined the role of motivation in collaborative learning. Much of the prior research that has investigated the relationship between students' theories of intelligence and learning has been in the context of individual learning. Relatively less work has focused on the role of motivation in collaborative learning groups (Pintrich, Conley, & Kempler, 2003; Senko, Hulleman, & Harackiewicz, 2011; Winne & Nesbit, 2010). However, we do know that during collaborative learning, students with entity theories are less likely to seek help from teachers and peers (Shih, 2007), as well as provide help to others (Dweck & Bempechat, 1983). During conflicts, they are likely to voice their displeasure with others openly and constructively compared to incremental theorists (Kammrath & Dweck, 2006). Compared to entity theorists, incremental theorists report higher use of collaborative learning strategies (Stump et al., 2009) and believe that collaboration is an important aspect of learning (Cotton & Cook, 1982).

A few studies have investigated the role of a related motivational construct – achievement goals in the context of collaborative learning. Performance goals, associated with

entity theories appear to prompt a more critical view of teammates, and students pursuing performance goals are more likely to show favoritism in partner choices (Levy, Kaplan, & Patrick, 2004) and have less tolerance for disagreements with partners (Darnon, Muller, Schrager, Pannuzzo, & Butera, 2006). When students have mastery goals (associated with incremental theories), they are more likely to openly share and welcome all ideas, whether weak or strong, whereas those who have performance goals give guarded opinions (Poortvliet, Janssen, Van Yperen, & Van de Vliert, 2007), summarily dismiss weak ideas (Darnon, Harackiewicz, Butera, Mugny, & Quiamzade, 2007), but welcome strong ideas which may benefit their own success.

In the present work, I aim to address the three limitations of the current work on collaborative learning discussed above. I compare two instructional conditions — a collaborative condition in which students learn with a partner and an individual condition in which students learn individually. Upon completing the learning activity, participants will complete a posttest individually. This will allow us to understand whether the effects of collaboration (if any) are sustained when participants are tested individually. I will test whether students would benefit differently from collaboration, by manipulating students' theories of intelligence to be either entity or incremental. Because manipulating students' theories of intelligence can potentially affect other motivational variables as well, I will also collect data on students' achievement goals and expectancy values as ancillary measures.

To address the limitation related to task complexity, I will use different outcome measures on the posttest. Procedural knowledge, which is relatively less complex, will be measured by testing whether students retain relevant information and procedures to solve isomorphic problems, very similar to those encountered during learning. Transfer, a more

complex form of learning, will be measured by performance on problems in which students would need to apply what they learned to a related problem, requiring a deeper conceptual understanding. On the simpler, isomorphic problems, I expect individuals to perform better than collaborators, which is consistent with predictions of the cognitive load theory. Further, on simple isomorphic problems, students would not experience a challenge, so I do not expect to see a difference between entity theorists and incremental theorists.

On transfer problems that require deeper conceptual processing, I expect to see collaborators perform better than individuals, consistent with the predictions by the ICAP framework, and also the predictions of the cognitive load theory. In terms of motivational beliefs, the incremental theorists are likely to be more constructive during the invention activity, more likely to persist even when they fail to invent a formula, and therefore, would be better prepared to learn from the subsequent worked example, extracting deeper features, leading to better transfer. Conversely, entity theorists are likely to engage in shallow processing, and will be more likely to be discouraged during invention when they hit impasses. Consequently, they will be less prepared to learn deeply from the subsequent worked example and will perform less well on transfer problems.

2.1 HYPOTHESES

The following set of hypotheses stated in terms of the ordering of means are tested (see Fig. 1 for graphical representation).

H_1: On procedural knowledge problems, collaborators will perform worse than individuals. Incremental Singletons (μ_2) will perform better than Incremental Dyads (μ_4) and

Entity Singletons (μ_1) will perform better than the Entity Dyads (μ_3). However, because procedural knowledge problems will not be challenging after having studied worked examples, and theories of intelligence typically affect learning only under challenging situations, I do not predict a significant difference between entity theorists and incremental theorists on procedural problems.

The above hypotheses can be expressed in a single model as:

M_1: $\mu_3 < \mu_1$; $\mu_4 < \mu_2$

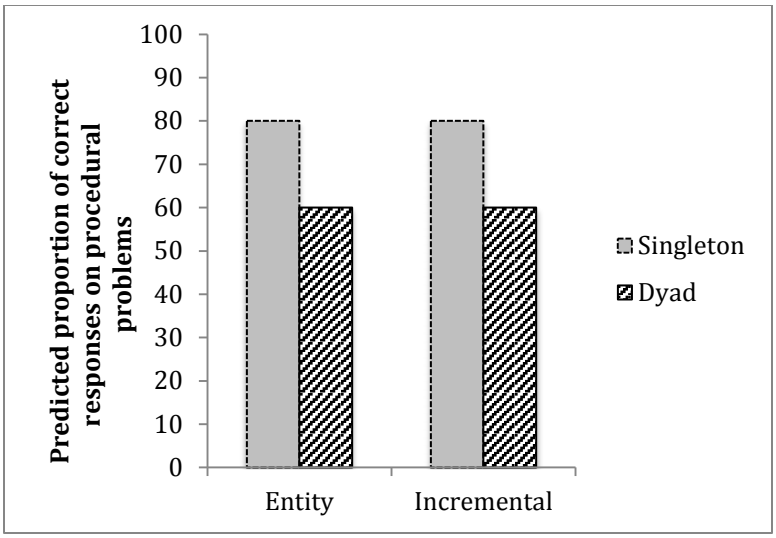
Thus, I predict a main effect such that individuals learn more than collaborators, no significant difference between incremental theorists and entity theorists, and no significant interaction on the procedural problems.

H_2: On transfer problems, collaborators will outperform individuals. Entity Dyads (μ_3) will outperform Entity Singletons (μ_1). Incremental Dyads (μ_4) will outperform Incremental Singletons (μ_2). The difference between the Entity Singletons (μ_1) and Entity Dyads will be less than the difference between Incremental Singletons (μ_2) and Incremental Dyads (μ_4), that is, students with incremental theories would benefit more from collaboration, compared to those with entity theories.

The above hypotheses can be expressed in a single model as:

M_1: $\mu_1 < \mu_3$; $\mu_2 < \mu_4$; $\mu_3 - \mu_1 < \mu_4 - \mu_2$

Thus, I expect a main effect such that incremental theorists learn better than entity theorists, a main effect such that collaborators learn more than individuals, and an interaction effect such that the difference between singletons and dyads will be more for incremental theorists.



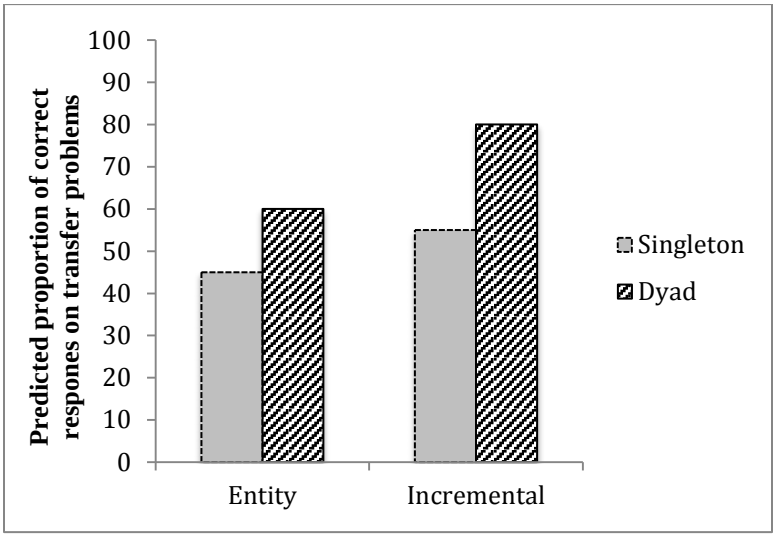


Figure 1. Means plots depicting expected pattern of results for procedural problems and transfer problems

3.0 METHOD

3.1 PARTICIPANTS

Participants were 163 undergraduates (83 female, 80 male) from the University of Pittsburgh, who participated in the experiment through the psychology subject pool. They received partial course credit for the course ‘’Introduction to Psychology’’ in return for their time. 122 were freshmen, 25 sophomores, 7 juniors, 2 seniors and two others noted their year in college as “other”. The average age of participants was 18.8 (*SD* = 1.65) years. As part of a demographic questionnaire, participants were asked to report whether they were currently taking or had taken in the past two years any college level mathematics and/ or statistics courses, including AP courses. The average number of courses taken by participants was 1.69 (*SD* = 1.17).

Prior research has shown that people interact differently with people from the same sex as themselves versus the opposite sex. Males have been shown to be more active and influential, and engage in more agentic activities in mixed-sex dyads compared to females (Levine & Moreland, 1990). These differences in interaction patterns could potentially cause men and women to learn differently from the interaction. To avoid this source of extra variance, dyads were restricted to same-sex dyads in this experiment.

3.2 DESIGN

The experiment had a 2 X 2 between subjects design. The first factor was the manipulated theory of intelligence. Participants were randomly assigned to adopt either an entity theory or an incremental theory of intelligence by having them read a fabricated "scientific" article that advocated either theory (see Materials for a full description). The second factor was learning condition, in which participants completed the learning session either individually or collaboratively. Thus, there were four experimental conditions: entity singleton (ES), incremental singleton (IS), entity dyad (ED), and incremental dyad (ID).

During the learning session, participants learned a novel statistics task, which involved calculating mean deviation as a measure of variability and calculating a standardized score to compare two sets of means (see Materials for a full description). After the learning session, all participants individually completed a post-test, which consisted of problems similar to the ones encountered during learning (isomorphic problems) and transfer problems.

3.3 MATERIALS

3.3.1 Materials used to induce theories of intelligence

Some prior studies have experimentally manipulated students' theories of intelligence (e.g., Bergen, 1991; Hong et al., 1999; Miele & Molden, 2010) using a paradigm in which students are asked to read a fabricated "scientific" article advocating either an entity theory or an incremental theory. In prior studies, manipulation checks showed that these manipulations were successful in

that students were significantly likely to endorse the entity view or incremental view, consistent with article they had initially read. A version of the articles used in prior studies was used in the current experiment (See Appendix A1 and Appendix A2 for complete articles).

The articles were formatted to resemble an article in a popular Psychology journal, such as "Psychology Today." Both articles were titled "The Origins of Intelligence: Is the Nature–Nurture Controversy Resolved?" Each article was formatted to match a magazine's layout with attention to detail such as font used, margins, and column width, complete with an advertisement. The three opening paragraphs of the two articles were identical, describing an eighteen-month-old precocious child. The subsequent paragraphs differed based on whether the article was advocating an entity view or an incremental view. The entity article offered a hereditary cause for the toddler's superior abilities, whereas the incremental article offered an environmental one. Each article contained approximately 1200 words, and was two pages long.

TOI questionnaire. In order to strengthen the manipulation, participants were asked to answer three open-ended questions after reading the article. These questions asked them to "summarize the main point of the article in one sentence," "describe the evidence from the article that you found most convincing," and "describe an example from your own experiences that fits with the main point of the article." These questions also served as a manipulation check.

3.3.2 Pretest

A pretest was administered to determine whether students had the right amount of prior knowledge in order to learn effectively from the learning materials. Participants who had extremely high or low prior knowledge coming into the experiment were excluded from further data analysis.

Materials for the pretest were adapted from a prior study by Schwartz and Martin (2004). The pretest consisted of three problems. The first problem tested procedural knowledge. Students were asked to calculate the mean, mode, median, and mean deviation for a set of eight numbers. This problem was scored out of four points, with one point for each of the measures. The second problem required graphical representation and calculation of variability to determine which of two football teams had a better record based on their number of wins for twelve consecutive seasons. This problem was scored out of three points — one point for the correct graphical representation showing a histogram, one point for the correct reasoning, and one point for the correct final answer. The third problem asked students to reason qualitatively about choosing the correct measure of central tendency for a given dataset. They were given a set of numbers representing the electricity bills of eleven families, based on which they had to determine whether a mean or a median would be the more appropriate measure of central tendency, and provide a reasoning for their choice. This problem was scored out of three points — one point for correct calculation of the mean and the median, and one point for the correct final answer along with the reasoning. If they provided the correct answer but did not provide any reasoning, they did not get the last point. Thus, the total score on the pretest was ten points.

Prior research suggests that students' theories of intelligence are more likely to come into play when a challenging situation is encountered (Dweck & Bempechat, 1983). Therefore, to make the challenge more salient, participants were allotted only 12 minutes to complete the pretest, even though pilot testing had indicated that they needed approximately 15 minutes to solve all the problems.

3.3.3 Learning materials

The learning materials were also adapted from Schwartz and Martin (2004). These included an invention task which involved inventing a formula for calculating mean deviation, instruction for calculating mean deviation that included a worked example, and a second invention task that required calculating standardized scores to compare two sets of means.

3.3.3.1 Inventing a formula for mean deviation

Students were first asked to invent a method to calculate variability for four sets of numbers. Each of the four grids seen in Figure 2 shows the result of a test using a different baseball-pitching machine. The diamonds represent where a pitch landed when aimed at the target X. Students had to devise a procedure for computing a quantity that expressed the variability for each of the pitching machines and decide which one was the most reliable. Given that students were novices in the domain, inventing a procedure for calculating mean deviation was a difficult task for most students, and was included to create conditions for failure. This was an important feature, because failure or a facing a challenging situation is important to invoke students' theories of intelligence. The second reason for choosing the invention task was that such tasks are likely to promote productive interactions such as asking questions, explaining, accepting and rejecting good and bad ideas, etc. among collaborators (Sears, 2006).

The invention problem carried one point for the correct answer and one point for correct reasoning. If participants correctly stated which pitching machine was the most reliable, they received one point. If they correctly stated the reasoning (that it has the least spread or variability), they received one point. If they stated no reasoning or an incorrect reasoning, for example, "it has the lowest average, so it is the most reliable," they received a zero.

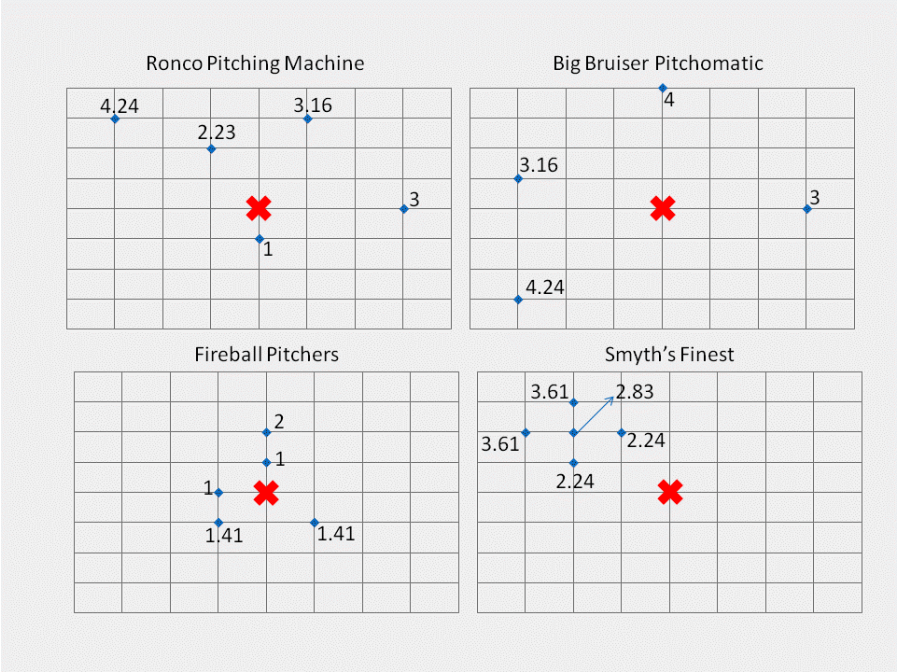


Figure 2. Inventing a formula for mean deviation (Schwartz & Martin, 2004)

3.3.3.2 Instruction on calculating mean deviation

The invention problem was followed by a one-page instruction on calculating mean deviation. This included a definition of mean deviation and an explanation of how it is calculated, followed by a worked example.

3.3.3.3. Inventing a procedure for standardization

After the instruction on mean deviation, students were given two new invention problems (problem 2 and problem 3). Problem 2 required participants to compare the records of two track stars across different sports, and devise a procedure to compare their performances, which required calculating standardized scores.

Problem 2 also carried one point for the answer and one point for reasoning. For correctly stating which track star had a more impressive record, participants received one point. For

correctly stating the reasoning for their choice, they received another point. An example of correct reasoning was as follows:

> "*Joe's record is more impressive because he scored more than two deviations away from the mean.*"

If they stated no reasoning or an incorrect reasoning, they received a zero. For example, if they calculated mean deviations and directly compared them without standardizing, they received a zero.

The third invention problem asked students to determine a student's grade on a curve by comparing it to scores of other people in the class. This problem required visually representing the scores on a histogram, calculating mean deviation, and plotting them on the histogram. Problem 3 carried one point for the answer and one point for graphical reasoning. If they correctly calculated the student's grade on each test, they received one point. If they correctly drew the histogram and plotted the student's score, the mean scores of the two classes and mean deviations, they received one point.

3.3.4 Test materials

The posttest was out of a total of 16 points, and consisted of one section with four problems testing procedural knowledge, and another section with two transfer problems. Each of these sections carried eight points.

3.3.4.1 Procedural knowledge problems

Four problems on the posttest tested procedural knowledge. The first three were isomorphic problems, which were closely related to the problems that students practiced during the learning

session. Students were required to calculate mean and mean deviation for three sets of numbers. These problems were solved by directly applying the formula for mean deviation that students learned during the learning activity. On each of the three problems, participants received one point each, for correctly calculating mean and mean deviation. For every incorrect answer, they received a zero. The fourth problem testing procedural knowledge also required the calculation of mean deviation, the only difference being that it was in the form of a word problem. It was worded as follows:

> "*Twenty students took a midterm in their science class, and they had an average score of 75. Five of them scored 70, five students scored 65, five students scored 80, and five students scored 85. What is the mean deviation?*

This problem acted as a distracter between the worked example on standardization (see section 3.3.4.2) and the transfer problems that followed. Participants could receive a score of 1 or 0 on this problem, depending on whether they calculated the mean deviation correctly.

Embedded worked example

After the first three problems in the procedural knowledge category, students received a worked example showing them how to calculate standardized scores. The embedded worked example was followed by a practice problem on which participants were required to calculate standardized scores and compare them. Participants received one point for each correct answer.

The isomorphic problems, the word problem, and the practice problem in the embedded worked example were together scored as a category of *procedural knowledge problems* out of a total of 8 points.

3.3.4.2 Transfer problems

There were two transfer problems, each of which required participants to calculate standardized scores as demonstrated in the embedded worked example. The first problem required participants to compare the performance of two students who took Biology tests with two different instructors. The second problem required them to compare the home runs of two baseball players during two different years. For each of the problems, the person with the higher standardized score had a better performance.

Each transfer problem was scored out of 4 points, making the total score on the transfer test 8 points. Two points were allotted for correctly calculating standardized scores. One point was allotted for determining the final answer. Lastly, one point per problem was allotted for the providing the correct conceptual reasoning. Participants could score either a 0 or 1 depending on whether they gave an incorrect or correct reasoning. If they simply stated an answer without giving any reasoning, or if they stated an incorrect answer, they received a zero. For example,

> *"Because this # (1.16) is lower, Susan scored better on the test."*

If they gave correctly stated the reasoning they received one point. An example is as follows:

> *"Robin did better because he scored 1.5 standard deviations above the average, whereas Susan only scored 1.16 standard deviations above the average."*

3.3.5 Questionnaires

3.3.5.1 In-task goal questionnaire (AGQ-R)

In order to assess students' achievement goals during the learning activity, they were given an activity questionnaire after they had solved the first of the invention problems. This measure

consisted of twelve items, and was created based on the Achievement Goal Questionnaire-Revised (AGQ-R) by Elliot and Murayama, (2008). Participants rated each item on a 5-point Likert scale (1 = *strongly disagree*, 3 = *unsure*, 5 = *strongly agree*). See Appendix B for full questionnaire. Some prior research supports the hypothesis that students' implicit theories of intelligence operate through goals — entity theories lead to performance goals, and incremental theories lead to mastery goals. Performance goals lead to surface processing, and poor learning outcomes, whereas mastery goals lead to deeper processing and good learning outcomes (Elliot, McGregor, & Gable, 1999). However, other studies have failed to find evidence for the predicted relationships between theories of intelligence, achievement goals, and performance (e.g., Dupeyrat & Mariné, 2005a). In this experiment, this AGQ-R was given to see whether the experimental manipulations of theory of intelligence affected students' goals in a systematic manner.

3.3.5.2 Theory of intelligence scale

After completing the test phase of the experiment, participants individually completed the eight-item Theories of Intelligence Questionnaire. This questionnaire developed by Dweck (1999) measures a relative preference for an entity or incremental theory of intelligence, by asking participants to rate their level of agreement (on a 1–7 Likert scale) with statements such as "Intelligence is something basic about a person that cannot be changed" and "No matter how much intelligence you have, you can change it quite a bit." (Appendix D). Incremental items were reverse coded and a composite score ranging from 8 (most incremental) to 56 (most entity) was calculated for each participant. Further information regarding the reliability and validity of this measure, can be found in Dweck, Chiu, and Hong (1995), and Dweck (1999).

Although there may have been some advantage in administering the TOI scale at the beginning of the experiment in addition to at the end of the experiment, to measure students' existing theories of intelligence, it was only administered at the end in the current study. Prior research on stereotype threat in social psychology indicates that even a single item on a questionnaire administered before taking a test can be enough to introduce stereotype threat, and affect performance on the test (Steele & Aronson, 1995). In a similar vein, taking the TOI scale prior to learning may have primed students to adopt a certain theory of intelligence, which may have interfered with the manipulated theory of intelligence. To avoid such interference, the TOI scale was given only at the end of the experiment.

3.3.5.3 Expectancy value questionnaire

The expectancy value questionnaire consisting of eleven items on a five point Likert scale, and two additional open-ended items was adapted from Wigfield and Eccles (2000). The first construct measured was *expectancy beliefs*, measured by the first five items on the scale. The first three items under expectancy beliefs denote *ability beliefs*, which are defined as a person's perception of his or her current competence at a given activity. The next two denote *expectancies for success*, which are expectancies focused on the future. Because ability beliefs and expectancies are closely related, they are collapsed into a single construct of expectancy beliefs. The next construct measured was *attainment value*, which refers to how important it is for the person to learn in that domain, measured by two items on the scale.

The subsequent two items measure *intrinsic value*, which refers to the person's intrinsic interest in that domain. The last two measure *utility value*, which refers to usefulness of the knowledge in that domain to the person. See appendix E for the full questionnaire. The expectancy value questionnaire measures a motivational construct orthogonal to students'

theories of intelligence. Sometimes, a person may believe that intelligence is fixed, yet engage in cognitive processes that are more typical of incremental theorists. According to the expectancy value theory, a person is likely to invest time and resources in learning something he or she believes to be useful, independent of what theory of intelligence they hold. In this experiment, information about students' expectancy values was collected to see whether the theory of intelligence manipulation affected these in any systematic way.

3.3.5.4 Demographic questionnaire

Participants completed a demographic questionnaire that included standard demographic questions about age, gender, and education level. This questionnaire also asked participants to report their SAT Math scores, and list all the mathematics courses they had taken at the college level, including AP classes. See Appendix F for full questionnaire.

3.4 PROCEDURE

The experiment took approximately 100 minutes to complete, and consisted of a pretest, theory of intelligence manipulation, a learning section, a posttest, and questionnaires. Figure 3 illustrates step-by-step the procedure that participants followed during the experiment.

All participants first individually completed a pretest that consisted of problems based on calculating mean, mode, median, and mean deviation, for which they were allotted twelve minutes. Next, they read either the entity article or incremental article for seven minutes. After reading the article, they completed the TOI questionnaire, which consisted of three open-ended questions (as described in Materials), for which they had five minutes. Next, they were given

talk-aloud practice, by asking them to talk aloud while solving some simple arithmetic problems. After the talk-aloud practice, they began the learning section, which was videotaped. Participants in the individual conditions completed the learning section individually, whereas those in the dyadic conditions completed the learning section with a partner. Participants in the individual conditions were simply asked to complete the activities in the booklet, while those in the dyadic conditions were asked to complete them with their partner. The collaboration was open-ended, in that no specific instructions with respect to collaboration (such as a script) were given. Between section 2 and section 3 of the learning section, participants completed the in-task achievement goals questionnaire (AGQ_R). The learning section took approximately 35 minutes.

After the learning section, all participants completed the test section individually for which they had 17 minutes. Finally, they completed the following questionnaires: the TOI scale, the expectancy-value questionnaire, and the demographic questionnaire. Upon completing the questionnaires, participants were given a full debriefing, in which they were informed that the article that they had read at the beginning of the experiment was not a scientific article, but was created just for the sake of this study. Any questions they had about the procedure were answered, and they were requested not to share the details of the experiment with others.

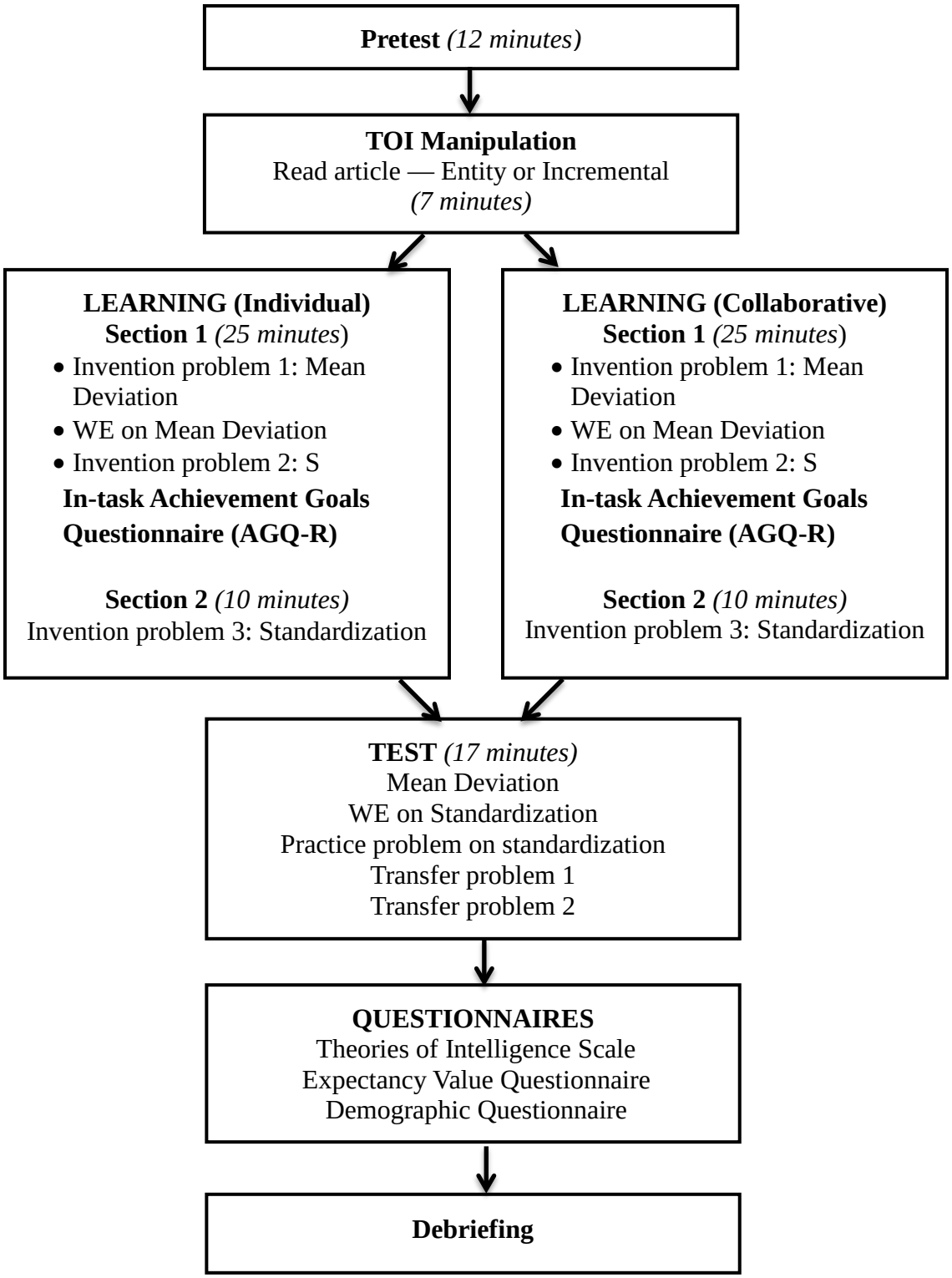


Figure 3. Flowchart of procedure

4.0 RESULTS

The results are presented in five sections. In the first section, I describe the results from two manipulation checks used to determine whether the manipulations used to induce theories of intelligence were successful. In the second section, I present the pretest results, followed by learning results in the third section. In the fourth section, I present the posttest results, in which I first describe overall posttest performance, followed performance on each type of problem - procedural knowledge and transfer. In the final section, I present the results on the motivational questionnaires, that is, the in-task AGQ-R and the expectancy value questionnaire. I set the alpha level at .05 for all main effects, interactions, and planned comparisons (Keppel, 1991). I calculated effect sizes (eta squared, $\eta_p{}^2$) for all significant main effects, interactions, and planned comparisons. I followed the guidelines by Cohen (1988) according to which effects are regarded as small when $\eta_p^2 < .06$, medium when $\eta_p^2 < .14$, and large when $\eta_p^2 > .14$.

To establish inter-rater reliability for qualitative portion of the learning problems and transfer problems, 25% of the problems were first scored by two independent raters. Disagreements were resolved through discussion. The resulting kappa was .89 across all problems.

4.1 SUCCESS OF MANIPULATION

Before testing our hypotheses regarding students' learning outcomes, it is important to first determine whether the manipulations used to induce theories of intelligence were successful, and whether students endorsed the theories of intelligence consistent with the article that they had read. I used two manipulation checks — the first was the TOI questionnaire that students completed immediately after reading the article (see section 3.3.1 for details). The second manipulation check was the TOI Scale (Dweck, 2000) that participants completed towards the end of the experiment (see section 3.3.5.2 for details).

4.1.1 TOI Questionnaire

The TOI questionnaire that was given immediately after participants completed reading either article served as the first manipulation check. The questionnaire consisted of three open-ended questions that asked participants to "summarize the main point of the article in one sentence," "describe the evidence from the article that you found most convincing," and "describe an example from your own experiences that fits with the main point of the article." The answer to each of these questions was coded 0 or 1, depending on whether it was consistent with an entity theory or an incremental theory respectively. Thus, participants could have scores ranging from 0 to 3 on the questionnaire. A score of 0 or 1 indicated an entity theory, whereas a score of 2 or 3 indicated an incremental theory. If participants who read the incremental article scored 0 or 1 on the questionnaire, their answers were considered to be inconsistent with their manipulated TOI. Similarly if participants who read the entity article scored 2 or 3 on the questionnaire, their answers were considered to be inconsistent with their manipulated TOI. Two independent raters

scored 25% of the questionnaires. A kappa of .96 was obtained on the first pass, so the first rater went ahead and scored the rest of the questionnaires.

Out of 158 participants, only three participants gave answers inconsistent with the article that they had read, and all three were in the entity condition. Thus, most participants answered the TOI questionnaire consistently with their manipulated theory of intelligence.

4.1.2 TOI Scale

For the second manipulation check, I analyzed participants' responses on the TOI scale that they completed towards the end of the experiment. The TOI scale was a Likert scale of 1-7 (with some items reversed), and scores could range from 8-56, with 8 indicating an extremely incremental view and 56 indicating an extremely entity view. Cronbach's alpha for the TOI scale was .95, indicating high internal consistency. The mean TOI score of participants who read the entity essay was 35.26 (SD = 10.66), whereas that of participants who read the incremental essay was 22.52 (SD = 9.01). Students who read the entity essay scored closer to the median (i.e., 32), compared to those who read the incremental essay. A two-tailed t-test indicated that students who read the entity essay and incremental essay responded significantly differently on the TOI scale, $t(161) = 8.24$, $p < .001$.

The TOI scale was also scored dichotomously by performing a median split, such that participants in the incremental condition who got a score between 8-31 were coded 1 for consistent, and those above 31 were coded 0 for inconsistent. Participants in the entity condition who obtained a score between 33-56 were coded 1 for consistent, and those below 33 were coded 0 for consistent. All participants whose score was 32 were coded as inconsistent. This conservative coding yielded 114 participants (72%) who endorsed a TOI consistent with their

manipulation. Of these, 29 were entity singletons, 32 incremental dyads, 17 entity dyads, and 33 incremental singletons. A chi-square test indicated that the likelihood of endorsing a TOI consistent with the manipulation was significantly different by condition, $\chi^2(3, N = 158) = 12.44$, $p = .006$. Entity dyads were least likely to endorse the TOI consistent with the manipulation.

4.2 PRETEST RESULTS

As described in the materials, participants were given more problems than they could reasonably solve in the allotted time on the pretest, to make the pretest more challenging. Accordingly, out of 163 participants, approximately 31% could not complete the last problem, whereas most could complete problems 1 and 2. Therefore, only scores on the first two problems were considered. Thus, the new total possible score on the pretest was 7 points. The mean proportion of correct responses on the pretest was 54 % (SD = 20), a relatively high proportion, considering that students were novices in the domain. Two participants got a score of 100% on the pretest, and one participant got a score of 0%. These three participants, being more than two standard deviations away form the mean were considered outliers and were excluded from subsequent data analyses. Two further participants were excluded due to missing data and problems with materials. Thus, the final number of participants was 158. Of these, 40 were in the Entity Singleton (ES) condition, 40 were Incremental Singleton (IS) condition, 36 were in the Entity Dyad (ED; 18 dyads) condition, and 42 were in the Incremental Dyad (ID; 21 dyads) condition. Upon eliminating the outliers, the range of scores on the pretest was 14% to 86%. Cronbach's alpha for the pretest was .6 suggesting weak internal consistency (likely due to the low number of problems in the pretest).

A two-way ANOVA was conducted on pretest scores to see whether participants differed by condition at pretest. There was no difference between entity theorists and incremental theorists, $F(1,154) = .567$, $p = .457$, $\eta_p^2 = .004$. There was no difference between singletons and dyads, $F(1,154) = .229$, $p = .63$, $\eta_p^2 = .001$. There was no significant interaction, $F(1,154) = .078$, $p = .78$, $\eta_p^2 = .001$. See Table 3 for means and standard deviations. This result suggests that participants were not different from each other at the outset. However, given that the range of scores of the pretest was relatively wide, the pretest scores were added as a covariate in further analyses.

Table 3. Means and standard deviations on pretest by condition

Condition	Mean	SD	n
Entity Singleton	0.52	0.18	40
Entity Dyad	0.54	0.20	36
Incremental Singleton	0.55	0.20	40
Incremental Dyad	0.55	0.20	42

4.3 LEARNING RESULTS

During the learning section, participants completed three invention problems. The first invention problem gave participants data from four pitching machines and asked them to determine which one was the most reliable. 116 out of 122 participants (the *n* is different from the pretest for the learning results, because each dyad is considered a single unit for these analyses) could correctly identify which pitching machine was the most reliable. However, 100 out of the 116 used incorrect reasoning to arrive at the answer. They calculated the average of all pitches, and took

the lowest one to be the most reliable. Thus, most participants did not solve the first invention problem successfully. A Chi-square test for final answer on the first invention problem showed no difference between conditions for the final answer $\chi^2(3, N = 122) = 1.905$, $p = .592$. A Chi-square test for conceptual reasoning component of the first invention problem was significant $\chi^2(3, N = 122) = 8.03$, $p = .045$. See table 4 for cell frequencies.

Follow-up Chi square tests for all six possible comparisons were conducted for the significant Chi-square omnibus test for the conceptual reasoning component of the first invention problem. The comparison between entity dyads and incremental dyads was significant, $\chi^2(1, N = 42) = 7.00$, $p = .021$. Incremental dyads were more likely to correctly state the conceptual reasoning for problem 1 compared to entity dyads. The comparison between entity singletons and incremental dyads was marginally significant, $\chi^2(1, N = 61) = 3.465$, $p = .079$. Incremental dyads were more likely to correctly state the conceptual reasoning for problem 1 compared to entity singletons. The comparison between incremental singletons and entity dyads was marginally significant, $\chi^2(1, N = 61) = 3.494$, $p = .085$. Incremental singletons were more likely to correctly state the conceptual reasoning for problem 1 compared to entity dyads.

On problem 2, participants were asked to compare the records of two track stars across different sports. Just before attempting problem 2, they had received instruction on calculating mean deviation. 85 out of 122 participants arrived at the correct answer. However, only nine of the 85 were able to provide a correct reasoning that involved standardizing of the scores. Most participants simply calculated the mean deviation for each sportsperson, and compared them without standardizing. A Chi-square test for final answer on the second invention problem showed no difference between conditions $\chi^2(3, N = 122) = .125$, $p = .989$. A Chi-square test for

conceptual reasoning of the second invention problem was also not significant $\chi^2(3, N = 122) = 2.009$, $p = .571$. See table 4 for cell frequencies.

The third invention problem required graphical reasoning, and was significantly more challenging compared to the first two problems. Only four participants got the correct answer, and none of the four gave the correct reasoning for their answer. Nearly 50% of the participants could not complete this problem, so it was not analyzed further.

Table 4. Frequencies of correct and incorrect answers on learning problems

		Entity Singleton	Incremental Singleton	Entity Dyad	Incremental Dyad
Problem 1 Answer	Incorrect	3	6	4	3
	Correct	37	34	17	18
Problem 1 Reasoning	Incorrect	36	34	21	15
	Correct	4	6	0	6
Problem 2 Answer	Incorrect	12	12	6	7
	Correct	28	28	15	14
Problem 2 Reasoning	Incorrect	38	37	18	20
	Correct	2	3	3	1

4.4 POSTTEST RESULTS

The posttest consisted of two types of problems — problems testing procedural knowledge, and transfer problems. I will first report performance on both types of problems collapsed together, followed by performance on each type of problem.

4.4.1 Overall posttest scores

Overall posttest scores ranged from 0% correct to 100% correct. Posttest performance was controlled for pretest performance by including the pretest score as a covariate. The effect of the covariate was significant, $F(1,153) = 5.375$, $p = .022$, $\eta_p^2 = .034$. Cronbach's alpha for the overall posttest was .806, indicating high internal consistency.

A two-way ANCOVA revealed that there was no significant difference between participants in the entity condition and participants in the incremental condition, $F(1,153) = .626$, $p = .430$, $\eta_p^2 = .004$. However, there was significant difference between singletons and dyads. $F(1,153) = 4.041$, $p = .046$, $\eta_p^2 = .026$, with singletons performing better than dyads. There was no significant interaction, $F(1,153) = .272$, $p = .603$, $\eta_p^2 = .002$. See Fig. 4 for means and standard errors.

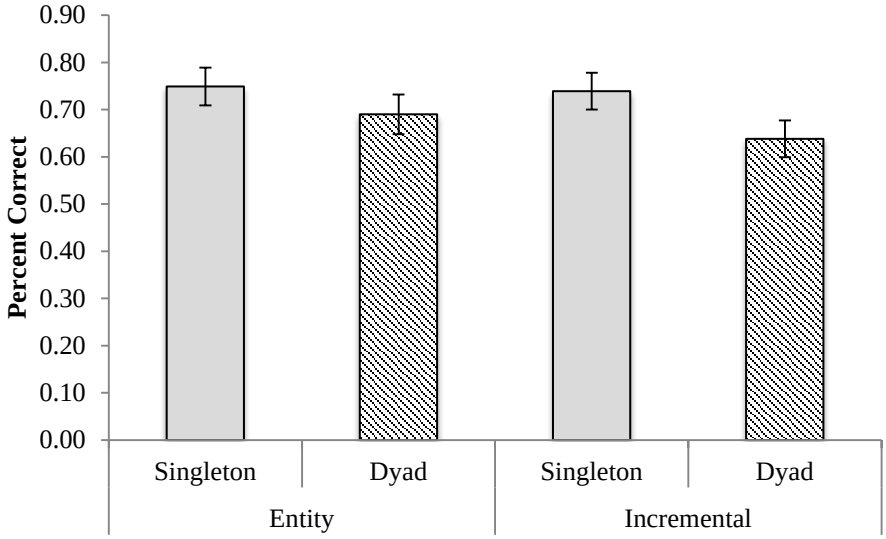


Figure 4. Posttest scores for all problem types adjusted for pretest scores

4.4.2 Procedural knowledge problems

Procedural problems accounted for a total of 8 points on the posttest. The proportion of correct responses ranged from 0% to 100%. A two-way ANCOVA with pretest percent correct as a covariate was used to test differences between conditions. The effect of the covariate was not significant, $F(1,153) = .835$, $p = .362$, $\eta_p^2 = .005$. Internal consistency for the procedural knowledge problems was weak, Cronbach's $\alpha = .612$.

No significant difference was found between participants in the entity article and participants in the incremental condition, $F(1,153) = .077$, $p = .782$, $\eta_p^2 = .001$. However, there was a significant difference between singletons and dyads, favoring singletons, $F(1,153) = 6.359$, $p = .013$, $\eta_p^2 = .040$, which was consistent with my prediction. There was no significant interaction, $F(1,153) = .322$, $p = .571$, $\eta_p^2 = .002$, again as predicted. See Fig. 5 for means and standard errors.

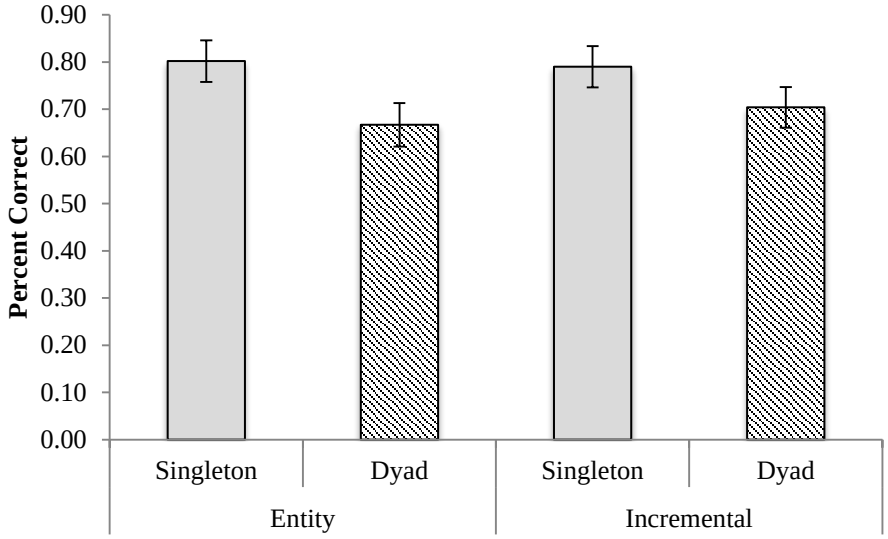


Figure 5. Posttest scores for procedural problems adjusted for pretest scores

4.4.3 Transfer problems

Transfer problems accounted for a total of 8 points on the posttest. Internal consistency for the procedural knowledge problems was high, Cronbach's α = .917. The proportion of correct responses ranged from 0% to 100%. A two-way ANCOVA with pretest percent correct as a covariate was used to test differences between conditions. The effect of the covariate was significant, $F(1,153) = 6.809$, $p = .010$, $\eta_p^2 = .043$.

No significant difference was found between participants in the entity condition and participants in the incremental condition, $F(1,153) = 1.847$, $p = .176$, $\eta_p^2 = .012$, which was contrary to the original prediction. There was no significant difference between singletons and dyads, $F(1,153) = .790$, $p = .376$, $\eta_p^2 = .005$, again contrary to the original prediction. Finally, there was no significant interaction. $F(1,153) = .232$, $p = .571$, $\eta_p^2 = .009$, again contrary to the original prediction. See Fig. 6 for means and standard errors.

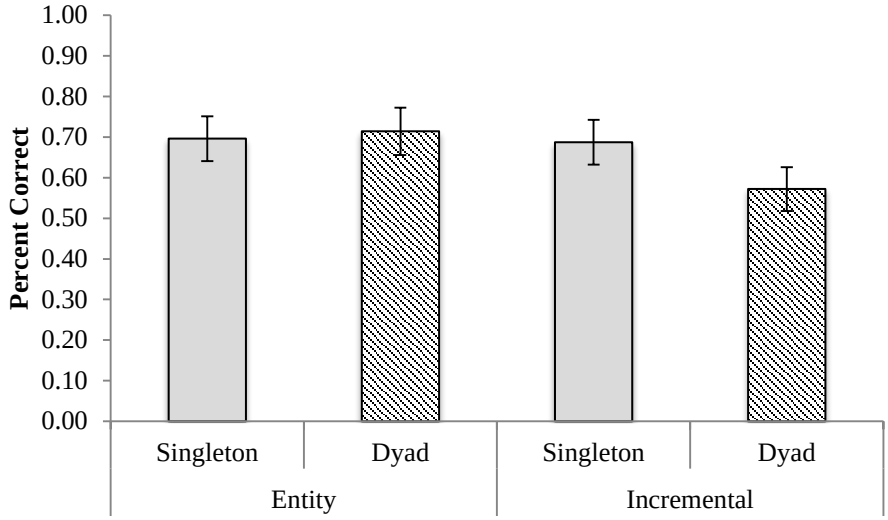


Figure 6. Posttest scores for transfer problems adjusted for pretest scores

4.4.4 Performance of participants who endorse a TOI consistent with their manipulation

Given that the predicted effect of theory of intelligence were not found on transfer problems, I analyzed the performance of only those participants who responded with a TOI consistent with their manipulated TOI on the scale given at the end of the experiment. For this analysis, I used the more conservative measure of the dichotomously scored scale, according to which 114 participants responded consistently with their TOI.

4.4.4.1 Overall posttest scores

A two-way ANCOVA tested whether participants differed by condition on overall posttest scores, using pretest scores as a covariate. The effect of the covariate was marginally significant, $F(1,109) = 3.024$, $p = .085$, $\eta_p^2 = .027$. There was no difference between those who read the entity article and those who read the incremental article, $F(1,109) = .036$, $p = .85$, $\eta_p^2 = .000$. There was a significant difference between singletons and dyads, $F(1,109) = 4.451$, $p = .037$, $\eta_p^2 = .039$, with singletons performing better than dyads. There was no significant interaction, $F(1,109) = .421$, $p = .518$, $\eta_p^2 = .004$.

4.4.4.2 Procedural knowledge problems

A two-way ANCOVA with pretest percent correct as a covariate was used to test differences between conditions on procedural knowledge problems. The effect of the covariate was not significant, $F(1,109) = .214$, $p = .644$, $\eta_p^2 = .002$. No significant difference was found between participants in the entity condition and participants in the incremental condition, $F(1,109) = 1.414$, $p = .237$, $\eta_p^2 = .013$. However, there was a significant difference between singletons and

dyads, favoring singletons, $F(1,109) = 6.845$, $p = .010$, $\eta_p^2 = .059$. There was no significant interaction, $F(1,109) = .316$, $p = .575$, $\eta_p^2 = .003$.

4.4.4.3 Transfer problems

A two-way ANCOVA with pretest percent correct as a covariate was used to test differences between conditions on transfer problems. The effect of the covariate was significant, $F(1,109) = 4.593$, $p = .034$, $\eta_p^2 = .040$. No significant difference was found between participants in the entity condition and participants in the incremental condition, $F(1,109) = .423$, $p = .517$, $\eta_p^2 = .004$. There was no significant difference between singletons and dyads, $F(1,109) = 1.007$, $p = .318$, $\eta_p^2 = .009$. There was no significant interaction, $F(1,109) = 1.878$, $p = .173$, $\eta_p^2 = .017$.

4.5 QUESTIONNAIRE DATA

4.5.1.1 In-task achievement goal questionnaire (AGQ-R)

The AGQ-R consisted of twelve items, three for each goal (mastery approach, mastery avoidance, performance approach, and performance avoidance). Participants rated them on a 5-point Likert scale (1 = *strongly disagree,* 3 = *unsure,* 5 = *strongly agree*). Scores for each goal were computed by aggregating ratings on three items representing that goal. The total possible score for each goal was 21. Cronbach's alphas calculated for each scale were as follows:

Mastery Approach: $\alpha = .798$; Mastery Avoidance: $\alpha = .679$; Performance Approach: $\alpha = .882$; Performance Avoidance: $\alpha = .812$, suggesting moderate to high internal consistency.

The means and standard deviations (in parentheses) for each goal by condition can be seen in table 5.

Table 5. Means and standard deviations on AGQ-R

	Entity Singleton	Incremental Singleton	Entity Dyad	Incremental Dyad
Mastery approach	15.65 (2.82)	16.13 (3.69)	15.94 (3.73)	16.05 (3.66)
Mastery avoidance	13.68 (3.78)	13.75 (4.03)	12.14 (3.73)	13.48 (4.53)
Performance approach	14.90 (3.71)	15.75 (4.10)	11.89 (5.12)	13.24 (4.39)
Performance avoidance	13.80 (4.40)	15.33 (4.01)	12.67 (4.91)	13.50 (4.88)

Next, separate two-way ANOVAs were conducted on each of the four goals. Table 6 shows the results from the two-way ANOVAs.

Table 6. ANOVA results for AGQ-R

Condition	TOI	Collaboration	Interaction
Mastery approach	*ns*	ns	ns
Mastery avoidance	*ns*	ns	ns
Performance approach	*ns*	$F(1,154) = 15.96$; $p < .001$**	ns
Performance avoidance	*ns*	$F(1,154) = 4.15$; $p = .043$**	ns

** denotes a statistically significant effect at $p = .05$

Singletons were found to endorse both performance goals more compared to dyads. There was no evidence for entity theorists endorsing more performance goals and incremental theorists endorsing more mastery goals, contrary to Dweck's process model.

4.5.1.2 Expectancy value questionnaire

This questionnaire consisted of eleven items on a five point Likert scale, and two additional open-ended items. The first five items measured expectancy beliefs, and the next six items measured attainment value, intrinsic value, and utility value with two items for each construct. Cronbach's alphas calculated for each construct were as follows: Expectancy Beliefs: $\alpha = .895$; Attainment Value: $\alpha = .817$; Intrinsic Value: $\alpha = .943$; Utility Value: $\alpha = .85$. Scores for each construct were computed by aggregating ratings on all items representing that construct. Table 7 shows the means and standard deviations for each construct.

Table 7. Means and standard deviations on the expectancy-value questionnaire

	Entity Singleton	Incremental Singleton	Entity Dyad	Incremental Dyad
Expectancy Beliefs (Total possible 25)	13.60 (3.52)	14.18 (4.34)	13.89 (3.79)	13.86 (4.00)
Attainment Value (Total possible 10)	5.70 (2.29)	6.55 (2.22)	5.83 (1.90)	5.55 (1.89)
Intrinsic Value (Total possible 10)	5.35 (2.05)	5.73 (2.15)	4.72 (2.19)	4.74 (2.04)
Utility Value (Total possible 10)	6.63 (2.02)	7.15 (2.08)	6.53 (1.54)	6.57 (1.93)

Next, separate two-way ANOVAs were conducted on each of the four constructs. Table 8 shows the results from the two-way ANOVAs.

Table 8. ANOVA results for expectancy value questionnaire

Condition	TOI	Collaboration	Interaction
Expectancy Beliefs	*ns*	*ns*	*ns*
Attainment Value	*ns*	ns	F (1,154) = 2.92, p = .09*
Intrinsic Value	*ns*	F (1,154) =5.809, p = .017**	*ns*
Utility Value	*ns*	*ns*	*ns*

** denotes a statistically significant effect at p = .05, * denotes marginal significance

Results suggest that collaborators placed less intrinsic value on learning statistics compared to individuals, across both motivational conditions. There was also a marginal interaction effect for attainment value such that incremental theorists showed high attainment value compared to entity theorists when learning individually, but lower attainment value than entity theorists when learning collaboratively. Follow-up t-tests were conducted on each of the six possible comparisons. The difference between incremental singletons and incremental dyads was significant, $t(80) = 2.206$, $p = .03$, favoring incremental singletons. The difference between entity singletons and incremental singletons was marginally significant, $t(78) = 1.686$, $p = .096$, favoring incremental singletons. No other comparison was significant.

5.0 DISCUSSION

This experiment investigated how students' theories of intelligence interact with different types of learning activities. According to the ICAP framework by Chi (2009), engaging in interactive learning activities such as collaboration leads to better learning compared to engaging in constructive activities individually. In this experiment, I investigated whether students' implicit theories of intelligence (entity versus incremental) interact with constructive and interactive learning activities.

Prior research suggests that theories of intelligence are activated only when a person is facing a challenge. Therefore, on relatively simple problems that tested procedural knowledge, I predicted that there would be no effect of theories of intelligence on learning. In terms of collaboration, I expected that for simple problems, collaboration would actually be worse than learning individually. Simple problems can be solved effectively by individuals, and therefore collaboration was not expected to provide additional benefit. In fact, it would hinder learning because of the extra cognitive load imposed by collaboration. On problems testing deep conceptual understanding, I predicted a different pattern of results. I predicted that collaboration would lead to better outcomes compared to individual learning, but students with entity theories would benefit less from collaboration compared to students with incremental theories, because students with incremental theories will be more likely to engage in productive interactions.

Data from Experiment 1 provided moderate support for the hypotheses on procedural knowledge measures. On the procedural knowledge problems, a small effect (η_p^2 = .04) of collaboration was observed. Consistent with my prediction, singletons learned significantly more than collaborators. Also consistent with my prediction, there was no effect of theories of intelligence. There was also no significant interaction between the variables, as predicted. On transfer problems that tested deeper conceptual understanding, contrary to my prediction, no significant difference was found between collaborators and individuals. Also contrary to my prediction, there was no significant difference between entity theorists and incremental theorists, and no significant interaction. The hypotheses were tested again using a more stringent manipulation check, to see whether there were effects for those participants who endorse the same theory of intelligence on the TOI scale as the article that they had read. However, even after eliminating data from participants whose TOI as not consistent with the manipulation, an effect for TOI was not observed on transfer problems.

5.1 THEORIES OF INTELLIGENCE AND LEARNING

In this section, I will describe some reasons theories of intelligence may not have had the predicted effect on learning in the case of transfer problems. Some prior studies have noted that college students, in general are more likely to endorse incremental theories over entity theories (Duda & Nicholls, 1992). In the present experiment as well, participants who received the entity manipulation scored much closer to median on the TOI scale compared to incremental theorists. This suggests that because they were more likely to endorse incremental beliefs at outset, participants moved towards the middle of the scale by the entity manipulation. This may have led

to most participants behaving in an incremental-like fashion. Unfortunately, we did not have a measure of participants' incoming TOI. Having students answer the TOI scale in the beginning would have primed them with a particular TOI, and interfered with the manipulated TOI. Ideally, it would have been desirable to obtain a measure of incoming TOI via administering the questionnaire a few weeks prior to the experiment, so participants would not connect it with the manipulation during the experiment. However, in the present study, it was not possible for practical reasons.

A second possible reason for not seeing an effect of TOI was that during the experiment, materials were presented such that participants first took the pretest, and then read TOI manipulation articles. Most participants scored an average of 50% correct on the pretest. Prior research shows that being challenged with a difficult task is an important precondition for implicit theories of intelligence to affect students' behavior and cognition. Participants may not have felt sufficiently challenged by the pretest, and this lack of challenge would have prevented them from connecting the message from the manipulation article to their own personal experience. Three steps would be taken in the follow-up experiment to address these concerns. First, the pretest would be made significantly more challenging. Participants would be asked to solve problems that would go beyond the prior knowledge of statistics for most students in that population. Second, the manipulation would be presented before the pretest rather than after, so that when students are solving the challenging problems, they would be more likely to think about the message presented in the article. Finally, students will receive failure feedback on their pretest indicating that their overall score was low and that they performed less well compared to other students on the pretest, regardless of how they actually performed. This was intended to introduce interpersonal comparison, and potentially strengthen the effect of the manipulation.

A third reason for the theory of intelligence manipulation having less of an impact is that the article they read was fairly general in nature. People have been shown to hold different implicit theories for different domains (Dweck, Chiu, & Hong, 1995). For example, some people may have entity theories about mathematics, but incremental theories about verbal abilities. Given that the manipulation article did not specifically talk about entity or incremental theories in the domain of mathematics or quantitative abilities, it may have seemed disconnected from the learning context to participants. Had the article been better integrated with the learning context, it may have had more impact in changing students' theories of intelligence. To address this concern in the follow-up experiment, the article would include some content that would connect directly with quantitative abilities. Giving participants a concrete example using vignettes that describe mathematics to be either an innate ability or a learned one should potentially help participants connect the article with the learning task that they complete later in the experiment.

A fourth reason for not seeing an impact of TOI was that the post-test was probably not discriminative enough. The transfer problem was placed too close to the worked example, so students would have easily made a connection between the embedded worked example and the transfer problems. Indeed, about 75% of the participants solved it correctly. Thus, participants may not have felt adequately challenged by the transfer problems, and their theories of intelligence would not have been activated, leading to similar cognitive processes and outcomes for entity theorists and incremental theorists. With better measures and tests that require conceptual thinking and reasoning at a deeper level, we may have observed an effect of students' theories of intelligence. This shortcoming will be remedied in the follow-up experiment by including better measures of procedural and conceptual knowledge, and also placing the embedded worked example and target transfer problem further apart.

A final possibility is that theories of intelligence have very little effect on learning outcomes in this context, and it is the learning activities in which the students engage that drive learning. The learning materials in this experiment were invention-based which may have led to more mastery like behaviors across the board. In a prior study by Belenky and Nokes-Malach (2013), task structure was found to be more effective in shaping students' motivation, than the instructions to adopt particular goals. In that experiment, students who participated in an invention-based learning activity tended to show mastery-like behaviors even though they were instructed to adopt performance goals. In the present experiment as well, the kinds of learning activities that students engaged in during the invention task could have been more powerful than the TOI manipulation, and impacted learning outcomes overriding students' theories of intelligence. In a subsequent follow-up study, I will attempt to answer the question of whether it was the manipulation that was not strong enough or whether theories of intelligence really do not affect learning as predicted.

5.2 COLLABORATIVE VERSUS INDIVIDUAL LEARNING

On procedural knowledge problems, consistent with our prediction, a main effect was found for collaboration such that individuals performed better than collaborators on the overall posttest scores. However, on transfer problems, there was no difference among conditions, which was not consistent with our predictions, and with past research.

According to the cognitive load theory, collaboration is likely to produce better learning outcomes compared to individual learning only when material demands cognitive resources of more than one person (Kirschner et al., 2009). Collaboration imposes its own costs, e.g.

transaction costs, which are offset only when the material to be learned is sufficiently challenging. In the present experiment, most participants scored about 50% on the pretest. Their performance on the transfer problems was similarly high, with nearly 80% of the participants getting at least one problem out of two correct on the posttest. The materials adapted for the current study were used primarily with high school students in prior research (e.g., Kapur & Bielaczyc, 2012; Schwartz & Martin, 2004); although a few studies have used them with college-age students as well (Belenky & Nokes-Malach, 2012; Wiedmann, Leach, Rummel, & Wiley, 2012). It is possible that the present materials may not have been challenging enough to require joint cognitive resources of two college-age students, whereby the communication and coordination costs imposed by the collaborative activity were not germane to the learning. The participants had also taken at least one college level course on mathematics or statistics on average, so they were not entirely novices in the domain. Thus, future studies need to examine this interaction by using tasks that are more difficult, and do require the cognitive resources of more than one person. Future studies should also test the same hypotheses with a younger population, or with students at a less selective institution, so that they are less likely to have high prior knowledge of the concepts and procedures to be learned.

5.3 RESULTS IN THE CONTEXT OF THE ICAP FRAMEWORK

The ICAP framework predicts that collaboration being an interactive activity would lead to better learning compared to learning individually. Chi (2009) provides a caveat that being interactive is better than being constructive when partners are being truly interactive. Certain types of interaction do not afford joint construction of knowledge. For example, if participants

simply divide the work among themselves and only share the final answer, they are not jointly creating knowledge that goes beyond the learning materials. In such cases, collaboration cannot be expected to lead to better learning compared to learning individually. Future studies need to understand what patterns of interaction lead to better collaboration by analyzing protocol data from collaboration. Such analyses would also help identify patterns of productive collaboration, and help scaffold better collaborative interactions in classrooms and other settings.

Another boundary condition of the ICAP framework may be that certain activities do not require joint construction of knowledge. If participants have the requisite prior knowledge, and are simply learning rote procedures, or relatively simple knowledge, there would not be much of a benefit to learning with a partner. This would also be consistent with the cognitive load theory, which predicts that if the task does not demand joint resources of more than one participant, collaboration is more likely to harm than help. The results from the present study provide some evidence in support of this claim — on procedural problems, singletons performed better than collaborators, which was also consistent with some past studies (e.g., Gadgil & Nokes-Malach, 2012; Kirschner et al., 2011). Therefore, a thorough cognitive task analysis may be beneficial in deciding whether the content would be learned more efficiently under a collaborative or an individual condition. Future studies should also include a measure of task difficulty as reported by students.

Finally, the ICAP framework also needs to consider the dimension of student motivation. Although the current study did not find an effect of students' theories of intelligence on learning, it is possible that other motivational factors such as goals or expectancy beliefs interact with cognitive factors, and lead to different outcomes than those predicted by the ICAP framework. Future studies should test the hypotheses tested in this study using other motivational measures.

6.0 STUDY 2: INTERACTION OF TOI WITH TWO TYPES OF CONSTRUCTIVE ACTIVITIES — TELL-AND-PRACTICE INSTRUCTION AND INVENTION

In the ICAP framework by Chi (2009), *active* activities are defined as ones in which students are characterized as "doing something" while learning. These activities are more perceptual than cognitive, and involve engaging activities such as looking, gesturing, selecting, repeating, in which the learner engages with the learning materials but does not typically generate any output that goes beyond the learning materials. *Constructive* activities are defined as ones that involve self-construction, such as explanation, elaboration, constructing a knowledge-map, in which the learner is creating new knowledge when engaging with the learning material. Finally, *interactive* activities are activities in which participants interact with another entity such as a peer, a tutor, or an intelligent tutoring system to create a joint understanding of the material to be learned. The ICAP framework predicts that interactive activities lead to better learning than constructive activities, which in turn are better than active activities.

Several studies have demonstrated that when the learners engage in constructive learning activities such as self-explanation (Chi et al., 1994), comparing across examples (Gadgil, Nokes-Malach, & Chi, 2012), or creating knowledge maps (Nesbit & Adesope, 2006), they learn more and retain what they have learned for longer periods of time. However, given the wide range of learning activities that can be termed "constructive", this also leads to some new questions. For example, do certain types of constructive activities work better than others? Does

the type of constructive activity interact with learner factors such as motivation, and do certain types of constructive activities work better for students with particular motivational beliefs? This experiment is designed to extend the work on constructive learning activities and test their interaction with students' theories of intelligence.

The debate about whether instruction should be open-ended and discovery oriented or whether it should be in the form of explicit, direct instruction is long-standing in the cognitive and educational literature (Lee & Anderson, 2013). Researchers are not in agreement about what amount of assistance during learning leads to the most optimal learning outcomes, and this debate has been termed the "assistance dilemma" (Koedinger & Aleven, 2007). The objective of this experiment is to understand the interaction between students' theories of intelligence and type of instruction (invention versus tell-and-practice), and whether one type of instruction may be suited for students with particular theories of intelligence over another.

On one end of the continuum, proponents of direct instruction argue that "direct" or "explicit" instruction produces robust learning and transfer, and that minimal guidance just does not work (e.g., P. A. Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004; Mayer, 2004). On the other end of the continuum, proponents of discovery learning methods (also called inquiry-based methods, problem-based methods, experiential methods, constructivist methods, or invention) argue that direct instruction produces only shallow learning and little to no transfer (e.g., Dean Jr & Kuhn, 2007), and that constructivist methods are better suited to achieving robust learning and transfer. It should be noted, however, that the definitions of the terms "direct instruction" and all instructional techniques under the umbrella of "discovery learning" are often vague and inconsistent. The two are often defined in relative terms, that is, the condition receiving less instruction is referred to as the "discovery condition" and the condition receiving

more is referred to as the “direct instruction” condition (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011).

More recently, the debate has shifted from whether one type of instruction is better than other, to what sequence of instruction would produce the most robust learning gains. For example, some studies have found that when students engaged in invention activities before receiving direct instruction, they showed a high degree of learning and transfer, as opposed to simply receiving direct instruction. For example, invention activities as operationalized by Schwartz and colleagues (see Schwartz & Martin, 2004) involve learning by attempting to invent a procedure for a solving a given problem, before being presented with the canonical procedure. Such activities were especially helpful in preparing students to learn from future instruction. Students are given a worked example embedded in the posttest, and then asked to solve a transfer problem similar to the worked example later in the test. Students who engage in invention activities prior to receiving the worked example are much more likely to solve the transfer problem correctly. This result suggests that withholding assistance early on in the instruction and providing it later can help students transfer better.

A similar paradigm has been used by Kapur (2008, 2012) who has demonstrated across several studies that even though students fail to generate the correct solution procedure during invention, they learn more from subsequent instruction, compared to being directly told the correct procedure. Kapur terms the failure to generate a solution “productive failure” because it helps students extract important principles from subsequent instruction, which they might otherwise overlook. During invention activities, students engage in constructive processes such as case comparison, schema extraction, and error correction, therefore, invention is classified as a “constructive activity” in Chi’s ICAP framework.

Within the ICAP framework, tell-and-practice would also be defined as a constructive activity. While the term tell-and-practice can be thought of as a form of “direct instruction” wherein of students learn concepts and procedures by direct transmission of knowledge rather than discovering it themselves, it can nevertheless afford plenty of opportunities for constructive learning behaviors (Chi, 2009). For example, in Klahr and Nigam (2004), students learned the control of variables strategy under either discovery-based or direct instruction. Students in both conditions had the opportunity to engage in exploration before being engaging in more discovery tasks or direct instruction. Further, students in the direct instruction asked by the instructor *“whether or not they thought the design would allow them to "tell for sure" whether a variable had an effect on the outcome.”* Thus, students getting direct instruction were far from being passive receivers of knowledge. They actively engaged with the materials and were also given the opportunity to be constructive through instructor-guided questions. As another example, tell-and-practice instruction in problem-solving often involves the use of worked examples (Sweller & Cooper, 1985). In order to effectively learn from a worked example, students need engage in constructive activities such as self-explanation (Chi, Bassok, Lewis, Reimann, & Glaser, 1989) or analogical comparison across multiple examples (Gentner et al., 2003). Thus, students need to be constructive in order to learn from worked exampples, therefore, tell-and-practice activities would fall under constructive activities under Chi’s ICAP framework.

Given that tell-and-practice and invention are both constructive activities, Chi’s framework would predict that both types of instruction would be equally effective. However, there is very little agreement among theorists regarding the effectiveness of tell-and-practice instruction and invention-based activities. Another limitation of the Active-Constuctive-Interactive framework is that it does not take into account how individual difference factors such

as motivation may interact with instruction. Next, I will review some arguments in favor of each type of instruction, and discuss how motivation can be an important moderating factor.

6.1 ARGUMENTS IN FAVOR OF INVENTION

The primary argument in favor of instruction involving discovery is that it promotes more robust knowledge acquisition. Proponents of invention-based instruction claim that direct instruction leads to only inert, rote knowledge, which cannot transfer easily outside the context of instruction (Dean Jr & Kuhn, 2007; McDaniel & Schlager, 1990). They argue that discovery tasks such as invention encourage students to be constructive, rather than merely be recipients of transmitted information.

Another proposed benefit of invention is that students benefit from learning from errors. When grappling with an invention task, learners are more likely to make errors and face impasses, which prompts them to delve deeper into the content to resolve these impasses. For example, research on impasse-driven learning during problem-solving suggests that when students reach an impasse, they learn better compared to when they don't (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). Even when students received the exact same explanations from a tutor, they did not learn as well when these explanations were not in response to an impasse. This finding suggests that the opportunities to make errors and learn from them, which are present in discovery learning situations, makes it superior to direct instruction, where fewer such opportunities are available.

Finally, advocates of discovery-based instruction argue that activities such as invention offer motivational benefits (Williams, 1993). When students are asked to invent a procedure or

discover a rule rather than being told directly, they have a greater control over their learning environment, which is shown to be beneficial to learning. It also promotes intrinsic interest, which translates into learning or mastery goals, which have (generally) been shown to lead to better outcomes compared to performance goals (Lepper & Chabay, 1985).

6.2 ARGUMENTS IN FAVOR OF TELL-AND-PRACTICE INSTRUCTION

According to the cognitive load theory, open-ended discovery-based tasks impose large costs on the inherently limited of human working memory (Kirschner et al., 2006). In discovery tasks, learners are required to search for a solution to a problem in a large, unstructured problem-space with minimal guidance, which taxes their cognitive resources, which then cannot be devoted to learning. This is particularly true of novice learners who lack the schemas in which to integrate the new knowledge. For example, a worked example (a form of direct instruction) eliminates the necessity to search, and directs the learners' attention to the essential problem-solving steps (Tarmizi & Sweller, 1988). Solving problems without the requisite prior knowledge is decidedly worse than solving problems after studying worked examples (e.g., Sweller & Cooper, 1985), which illustrates the superiority of direct instruction according to cognitive load theorists.

Another proposed benefit of direct instruction is its efficiency. After engaging an open-ended discovery task, students may eventually discover a principle or concept by themselves, but this is a significantly less efficient approach compared to being told a principle or concept via direct instruction. Given the open-ended nature of invention learning tasks, educators are often faced with a dilemma of whether to devote more time to invention activities, or to cover the required content prescribed by the syllabus in the given time frame (Hammer, 1997). Some

studies have found that discovery learning led to successful learning outcomes only when combined with high levels of practice. For example, Brunstein and colleagues (Brunstein, Betts, & Anderson, 2009) studied how students learn Algebra under increasing levels of guidance in the context of an intelligent tutoring system. Students were given no guidance (discovery condition), verbal directions, direct demonstration, or both (direct instruction). Students receiving direct instruction accomplished the task in shorter amounts of time and learned just as well as those who engaged in discovery tasks as measured by immediate, delayed, and transfer, tests. Thus, students receiving direct instruction learned more efficiently than those who engaged in discovery. If left to their own devices, students often experience floundering, and excessive floundering not only increases time on task, but also causes students to forget what they have just learned (Lewis & Anderson, 1985). Direct instruction reduces floundering, thereby increasing the effectiveness and efficiency of instruction.

Finally, advocates of direct instruction argue that direct instruction is more motivating than discovery-based instruction. Novice learners often do not have the prior knowledge necessary to successfully discover a principle or solve an invention problem during an inquiry-based task, which causes floundering. Floundering can lead to boredom and frustration, and lead to maladaptive behaviors (H. A. Simon, 2000). Failing to correctly solve a discovery problem can also lead to negative judgments of performance (e.g., Reiser, Copen, Ranney, Hamid, & Kimberg, 1994).

In the present study, I test the interaction of students' motivational beliefs and the type of instructional activity. Students' theory of intelligence will be manipulated to be entity or incremental, and they will participate in either an invention activity or tell-and-practice instruction. As in Experiment 1, participants in this study will complete a pretest, a learning

session, a posttest, and several questionnaires measuring motivational and demographic variables. In the next section, I will outline hypotheses for main effects and interactions.

6.3 HYPOTHESES

Prior research comparing tell-and-practice and invention has found mixed outcomes with respect to learning. I propose two competing hypotheses for how the two learning activities might interact with students' motivational beliefs.

H_1: If invention activities prompt students to be more constructive, leading to better learning, we can expect entity theorists to benefit more from invention activities. Incremental theorists are likely to be constructive regardless of type of instruction, so they will not show significant differences in learning under the two instructional conditions. This is consistent with Chi's ICAP framework, under which both invention and tell-and-practice instruction are both classified as constructive activities.

If Entity Tell-and-Practice = $\mu 1$, Incremental Tell-and-Practice = $\mu 2$ Entity Invention = $\mu 3$ & Incremental Invention = $\mu 4$, then in terms of mean differences, the above hypothesis can be stated in the form of a model as:

M_1: $\mu_1 < \mu_3$; $\mu_2 = \mu_3$; $\mu_3 = \mu_4$

In order words, we expect a main effect such that incremental theorists learn better than entity theorists, a main effect such that participants in the invention condition learn more than those in the direct instruction condition, and this main effect will be qualified by an interaction effect such that entity theorists will benefit more from invention, whereas incremental theorists would perform equally well under both instructional conditions. The advantage of invention

activities for entity theorists would be more prominent for problems requiring a deep conceptual understanding. For isomorphic problems requiring routing application of procedures, entity theorists and incremental theorists are not expected to be significantly different.

H_2: If invention activities hurt learning by causing off-task behavior and imposing excessive cognitive load in comparison with tell-and-practice instruction, then a different pattern of results can be expected. In such a scenario, tell-and-practice instruction should be more beneficial to entity theorists. However, incremental theorists will be less likely to be affected by the type of instruction, since they are likely to engage in constructive regardless. Therefore, they will not show significant differences in learning under the two instructional conditions. Again, this prediction is consistent with Chi's ICAP framework, under which both invention and tell-and-practice instruction are both classified as constructive activities.

In terms of mean differences, the above hypothesis can be stated in the form of a model as:

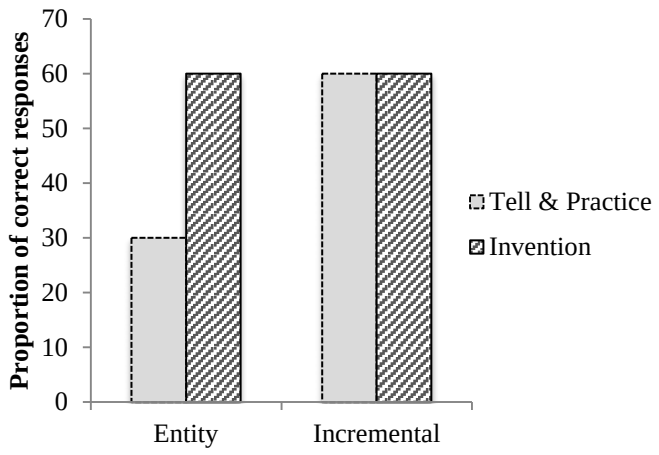
M_2: $\mu_3 < \mu_1$; $\mu_1 = \mu_2$; $\mu_1 = \mu_4$

In order words, we expect a main effect such that incremental theorists learn better than entity theorists, a main effect such that participants in the tell-and-practice condition learn more than those in the invention condition, and this will be qualified by an interaction effect such that entity theorists will benefit more from tell-and-practice instruction, whereas incremental theorists would perform equally well under both instructional conditions. The advantage of tell-and-practice instruction for entity theorists would be more prominent for problems requiring a deep conceptual understanding. For isomorphic problems requiring routing application of procedures, entity theorists and incremental theorists are not expected to be significantly different.

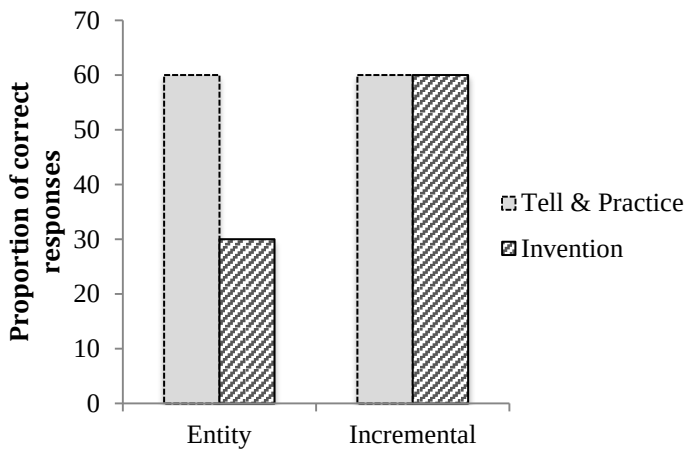
The models described above based on informative hypotheses will be tested against the unconstrained model Mo, such that there are no relationships between μ_1, μ_2, μ_3, & μ_4

M_o: μ_1 μ_2 μ_3 μ_4.

Figures 7 shows the predicted ordering of means for models M_1 and M_2.



Model M_1



Model M_2

Figure 7. Means plots depicting expected pattern of results for hypothesis 1 and hypothesis 2 for conceptual problems

7.0 METHOD

7.1 PARTICIPANTS

Participants were 100 undergraduate students from University of Pittsburgh, who participated in the experiment through the psychology subject pool. They received partial course credit for "Introduction to Psychology" in return for their time. All except three were freshmen. The average age of participants was 18.2 (*SD* = .63) years. As part of a demographic questionnaire, participants were asked to report whether they had taken in the past two years or were currently taking any college level mathematics and/ or statistics courses, including AP courses. The average number of courses taken by participants was 1.72 (*SD* = 1.24).

7.2 DESIGN

The experiment was a 2 X 2 between subjects design. The first factor was the manipulated theory of intelligence. Similar to Experiment 1, participants were randomly assigned to adopt either an entity theory or an incremental theory of intelligence, by having them read a fabricated

“scientific” article that advocates either theory (see Materials for a full description). The second factor was instructional condition, with two levels — tell-and-practice instruction and invention. In the tell-and-practice instruction condition, participants were shown the procedure to calculate mean deviation via a worked example and then given several practice problems. In the invention condition, participants were asked to invent a procedure to calculate mean deviation and then shown the correct procedure via a worked example, followed by a few practice problems. Thus, there were four conditions — entity tell-and-practice, entity invention, incremental tell-and-practice, and incremental invention. There were 25 participants in each condition.

During the learning session, participants learned to calculate mean deviation as a measure of variability. After the learning section, they completed a test section, which included an embedded worked example that showed how to calculate a standardized score for two sets of means (see Materials for a full description).

7.3 MATERIALS

Materials were adapted from prior research on theories of intelligence (Dweck, 2000) and research on preparation for future learning (Schwartz & Martin, 2004), and were similar to ones used in Experiment 1, with a few modifications as described later in this section. Materials consisted of articles used to induce theories of intelligence, a pretest, learning activities, a post-test and several questionnaires, described in more detail in the subsequent sections.

7.3.1 Materials used to induce theory of intelligence

Similar to Experiment 1, participants' theories of intelligence were manipulated by having them read an article that argued in favor of either the entity theory or the incremental theory. Both articles contained approximately 1200 words, and were two pages long.

The articles were very similar to the ones used in Experiment 1 (see Appendix A), except for a few changes. The most notable change was to include a paragraph in the article that directly tied it to mathematical abilities. This change was made because one possible reason for the lack of effect of TOI in experiment 1 could be that although participants largely appeared to adopt the theory of intelligence espoused by the article they had read, they may not have necessarily connected it to the activities that they completed later in the experiment. To make the connection more salient, the following paragraph was included in the incremental article:

> *"While past research has largely focused on intelligence as a general construct, newer work has begun to address whether people's abilities in specific domains are dominated more by their genes or their environments. For example, some people seem to have a gift for mathematics – no matter how complex a mathematical procedure, they quickly master it. Dr. Marissa Feng at Stanford University has focused on this very issue for the past six years. Over a series of experiments, she put participants of varying mathematical abilities through rigorous training sessions on calculus operations. She repeatedly found that upon completion of training, all participants made astounding gains in their problem-solving skills, even those who seemed to enter the experiment with a complete lack of a "gift" for mathematics. This evidence led her to conclude that the concept of "innate talent for mathematics" is largely a myth and people can improve their abilities with the right kind and amount of training and practice."*

Correspondingly, the following paragraph was included in the entity article:

> *"While past research has largely focused on intelligence as a general construct, newer work has begun to address whether people's abilities in specific domains are dominated more by their genes or their environments. For example, some people seem to have a gift for mathematics – no matter how complex a mathematical procedure, they quickly master it. Dr. Marissa Feng at Stanford University has focused on this very issue for the past six years. Over a series of*

experiments, she put participants of varying mathematical abilities through rigorous training sessions on calculus operations. She repeatedly found that despite the amount of training, the only participants who excelled at the task were ones who had superior mathematical skills to begin with. This evidence led her to conclude that people either have an innate talent for mathematics or they do not."

7.3.1.1 TOI questionnaire

After reading the article, participants completed an open-ended questionnaire with three questions designed to strengthen the manipulation. This questionnaire was the same as used in Experiment 1. Please see section 3.3.1 for details.

7.3.2 Pretest

The pretest consisted of four problems and the maximum score that a participant could attain was 10 points. The first two problems asked to calculate the mean and mean deviation for a set of numbers, each for possible score of two points. These two problems sought to determine participants' procedural knowledge based on their prior knowledge of these two concepts. The third problem required them to calculate a standardized score for two sets of data comparing different things. Each participant got either version A or version B of this problem, counterbalanced with the posttest. The possible score on this problem was three points.

Problem version A was as follows:

"Two people were arguing whether Joe Smith or Mike Brown had more power for hitting home runs. Joe Smith's longest homerun was 540 ft. That year, the mean homerun among all players was 420-ft long, and the average deviation was 70 ft. Mike Brown's longest homerun was 590 ft. That year, the mean homerun was 450 ft, and the average deviation was 90 ft. Who do you think showed more power for his biggest homerun, Joe Smith or Mike Brown? Use math to help back up your opinion."

Problem version B was as follows:

> *"Susan and Robin are two teenagers who both just took their state driver's license road test. They are arguing about who got a better score on their test, which is scored out of 100 possible points. Susan got an 88 taking the driving test with Mr. Wheelie. The mean score Mr. Wheelie gave out that day was a 74, and the average deviation was 12 points. The average deviation indicates how close all the people taking the test were to the average. Robin earned an 82 on Mrs. Axel's driving test. On that day, the mean score Mrs. Axel gave out was a 76, and the average deviation was 4 points. Both Mr. Wheelie and Mrs. Axel tested one hundred teenagers that day. Who do you think did better, Susan or Robin? Use math to help back up your opinion. Please use scrap paper if you need additional space for your calculations or graphs."*

This problem sought to differentiate students who relied on intuitive knowledge to provide reasoning from those who used a conceptually accurate reasoning based on standardized scores (even though they may not state the terminology correctly).

Finally, the last problem on the pretest required participants to determine where a student's grade of 120 points fell on a curve for each of two tests, while being given the number of students that fell in each range of scores. This was a difficult problem, and required more time to solve, compared to the other two problems. Participants were (falsely) told by the experimenter after scoring the pretest that the average score on the pretest was 8 points (see the Procedure section for more details), therefore, it was important to preclude most participants from scoring 8 or close to 8 points. The difficult problem was included in the pretest to make it very difficult for participants to get a high score on the pretest.

7.3.3 Learning materials

The materials for the invention condition and the tell-and-practice condition were kept informationally equivalent to the highest degree possible. The time spent by participants completing the activities was the same for the two conditions at 35 minutes.

7.3.3.1 Invention condition

The materials for the invention condition were very similar to ones used in Experiment 1. They consisted of three sections: The first section consisted of an invention problem, which required participants to invent a formula for calculating mean deviation. The second section consisted of a worked example on mean deviation followed by a practice problem. The third section consisted of two more invention problems requiring the calculation of a standardized score.

Section 1 - Inventing a formula for mean deviation

During the learning session, students were first asked to solve a problem based on the mean deviation formula. This problem was the same as used in Experiment 1 (See section 3.3.3.1 for a description).

Section 2 - Instruction on calculating mean deviation

In this section, students were given a one-page instruction on calculating mean deviation using a worked example. Again, this was the same as was used in Experiment 1.

Section 3 - Inventing a procedure for standardization.

The first problem in this section (invention problem 2) asked to compare two scores from different distributions to one another. This problem was about two students in different classes

who want to know who did better on a test, which may have been graded differently by their respective teachers. Students were provided with means and mean deviations and a histogram for each of the classes. However, they were not shown how to map the information on the histogram and how this can help determine which student did better. Instead, students were expected to come up with the procedure themselves, and provide reasoning for who they thought did better. This problem was intended to move them one step closer to the procedure for calculating standardized scores.

The second problem in this section (invention problem 3) was another invention problem ("Track Stars"), in which students are asked to decide which of two players (Bill and Joe) from different events (high jump and long jump) shattered a record more. Students were given a set of scores from each of the two events, and two values that represented the performance of Bill and Joe. They were asked to come up with a procedure that would help them decide who shattered the record more vis-a-vis the other performances in their category. This problem required the calculation of standardized scores in order to compare the scores from two different datasets to each other.

7.3.3.2 Tell-and-practice condition

The materials for the tell-and-practice condition consisted of two sections:

Section 1 – Worked example and practice problems on mean deviation

The first section consisted of a worked example on mean deviation followed by several practice problems. For the first practice problem, students had to determine which of two high school football teams had a better record based on their number of wins for twelve consecutive seasons.

This problem carried one point for the correct final answer, and one point for using the correct strategy (i.e., calculating mean deviation).

The first practice problem was followed by four other practice problems, which included four sets of data and required students to calculate mean deviation for each. The data provided were identical to the pitching machine data used in the invention condition, so that materials in the invention condition and the tell-and-practice condition could be informationally equivalent to the greatest degree possible.

Section 2 – Worked example and practice problems on standardization

The same problem that was used in the invention condition (section 2), about students graded differently on two tests was used. However, in the tell-and-practice condition, instead of requiring students to invent a procedure, they were given the correct solution to the problem. Specifically, they were shown how to mark mean and mean deviation on the histogram, and were given an explanation of what each deviation indicates. The worked example was followed by another practice problem, which was the same as the third invention problem "Track Stars" used in the invention condition.

7.3.4 Posttest materials

The posttest consisted of eight problems that tested three components of students' understanding — procedural knowledge, conceptual understanding, and qualitative reasoning. Several of the problems tested a combination of these three compoenents. Table 9 shows the problem components assessed by each problem. Additionally, there was an embedded worked example

that demonstrated the procedure of calculating standardized scores, followed by a worked example.

Table 9. Problems and problem components

	Procedural	Conceptual	Reasoning
Problem 1	✓		
Problem 2	✓		
Problem 3		✓	✓
Problem 4		✓	✓
Problem 5	✓		✓
Problem 6			✓
Problem 7	✓	✓	✓
Problem 8		✓	

Problems 1 and 2

Problems 1 and 2 tested procedural knowledge by asking students to calculate the mean deviation for two sets of numbers. Each of these problems was scored out of one point. Participants received one point for each of the problems correctly calculating mean deviation. If they had correctly stated the formula for mean deviation, but gotten the calculations incorrect, they still received one point.

Embedded worked example

Students received a worked example showing them how to calculate standardized scores. Participants were simply required to study the worked example, and there were no points associated with this section.

Problems 3 and 4

Problems 3 and 4 tested both **conceptual understanding** and **qualitative reasoning** skills. Problem 3 asked them to look at histograms of four datasets and determine which one had the least mean deviation. They were required to provide reasoning for their answer. Participants received 0 or 1 on the conceptual understanding portion depending on whether they correctly identified the dataset with the least mean deviation. They could receive 0, 1, or 2 on the reasoning portion, which was scored as follows:

They received a zero if their reasoning was completely unrelated to why the dataset had the least mean deviation. For example, one participant stated,

"Team B because there are fewer than two players representing each data point"

The reasoning offered is a clear example of an incorrect reasoning, because the number of entries representing each data point has no bearing on whether the mean deviation would be high or low.

An incomplete or partially correct reasoning received one point. As an example,

"Only three points 72,74,and 76 are represented."

Although this participant correctly notes that the dataset with the least mean deviation has only the three points mentioned above, he or she neglects to mention that the points are in a close range, and that this indicates low variability.

A completely stated and correct reasoning received two points. For example,

> *"By simply looking at the histograms, I would think that histogram "C" would have the least mean deviation because the numbers are closest together and are more consistent in value."*

This participant correctly states that there is less spread in the data, and more consistency in the value.

Problem 4 described a procedure to calculate mean deviation that consisted of two flaws, described as used by a fictitious student. One of the flaws in the procedure was a surface level one, which was a simple error in the calculation. The other flaw was a deeper one, in that the sum of deviation was divided by the mean instead of the number of data points. Participants had to determine whether he arrived at a correct answer for the conceptual understanding portion of this problem, and explain their reasoning by pointing out the flaws in the procedure for the qualitative reasoning portion of the problem. Participants received 0 or 1 on the conceptual understanding portion depending on whether they correctly stated whether the fictitious student used a correct procedure. They could receive 0, 1, 2, or 3 on the reasoning portion, which was scored as follows:

If both the surface and deep level flaws were mentioned, the participant received a 3. If only the deep level flaw was mentioned, the participant received a 2. If only the surface level flaw was mentioned, the participant received a 1. If any flaw other than the two mentioned above was stated, or if no reasoning was given, the participant received a 0.

Problem 5

Problem 5 asked to calculate mean deviation for a set of numbers with a value “55” included and excluded. They had to then explain how that value affected the mean deviation. The number 55 was an outlier, and therefore caused the mean deviation to increase greatly. This problem tested **procedural knowledge** as well as **qualitative reasoning.**

For the procedural knowledge component, participants received one point each for correctly calculating mean deviation with and without the value “55”. For the qualitative reasoning component, they received a 0 if they provided an incomplete or an incorrect reasoning, and 1 if they provided a correct reasoning. An example of incorrect reasoning was as follows:

"The mean deviation increases dramatically with the inclusion of 55 in the data set. The mean deviation shoots up from 2 to 5.33 with the inclusion of 55."

This statement makes no reference to 55 being an outlier. It simply restates that the mean deviation increases upon including 55, which can be easily discerned by looking at the two mean deviations.

An example of correct reasoning would be as follows:

"Including 55 greatly increases the mean deviation because it is very far off from the rest of the data."

Problem 6

Problem 6 provided participants with two data sets representing the numbers of races won by two horses in five seasons. They were told that one of the horses was a better bet because it had a better average. They had to determine the flaw in the reasoning. The flaw was that consistency was not taken into account and that mean deviation would be a better measure of evaluating the two horses. Thus, this problem tested **qualitative reasoning.** Participants could score a 0,1, or 2 on this problem, which was scored as follows:

If the participants gave no reasoning, or gave an incorrect reasoning, or mentioned an unrelated construct, they received a 0. For example,

"The flaw is that even though Marmalade has a better average than Supernova, her standardized score may not be as good as Supernova's."

Standardized score does not matter in this example, because the two quantities being compared (i.e., the number of races one) are the same scale.

An incomplete reasoning received a score of 1. If participants simply state that she should have considered mean deviation without explaining why mean deviation would be a better

measure of consistency than the mean, they received one point. For example,

> *"The flaw in Clara's reasoning is that she did not take into account the deviation of the mean."*

A completely stated and correct reasoning received two points. For example,

> *"So her flaw is that she did not consider how consistent the horse was. Marmalade may have won more races, but Supernova is more consistent with her wins and losses."*

Problem 7

Problem 7 was a transfer problem, and participants got either version A or version B of the problem, counterbalanced with the problem they had gotten at pretest. Version A required them to compare the home runs of two baseball players during two different years. Version B required them to compare the performance of two teenagers who took a driving test with two different instructors. On each of the versions, the person with the higher standardized score had a better performance. Students were required to apply the standardization procedure that they had learned in the embedded worked example. The problems in between the embedded worked example and the transfer problem were to ensure that students did not notice an immediate connection between the two. The transfer problem had all three components – **procedural knowledge, conceptual understanding**, and **qualitative reasoning,** which were scored as follows.

For the **procedural knowledge** component, they received one point for each correctly calculated standardized score. Thus, they could score a 0, 1, or 2. For the **conceptual understanding** component, they had to state who performed better, and received a zero or a one for each incorrect or correct answer respectively. For the qualitative reasoning component, they had to demonstrate an understanding of what the standardized score meant, that is, the higher standardized score on the test meant a better performance. If no reasoning or incorrect reasoning

was provided, participants received a zero. For example,

> *"Because this # (1.16) is lower, Susan scored better on the test."*

For a correctly stated reasoning, they received one point. For example,

> *"Robin did better because he scored 1.5 standard deviations above the average, whereas Susan only scored 1.16 standard deviations above the average."*

Problem 8

Problem 8 was another **conceptual understanding** problem in which participants were asked to create two sets of data such that the mean deviation of Set A was less than the mean deviation of Set B, and the mean of Set A was greater than the mean of Set B. They received 1 point if they created a dataset that satisfied both conditions, and 0 if no condition or only one of the conditions was satisfied.

7.3.5 Questionnaires

Participants completed several questionnaires during the course of the experiment.

7.3.5.1 In-task goal questionnaire

This questionnaire was the same as used in Experiment 1. See section 3.3.5.1 for details. .

7.3.5.2 Theories of intelligence scale

This questionnaire was the same as used in Experiment 1. See section 3.3.5.2 for details. .

7.3.5.3 Expectancy value questionnaire

This test was adapted from Wigfield and Eccles (2000) and was the same as used in Experiment 1. See section 3.3.5.3 for details. .

7.3.5.4 Demographic questionnaire

This was the same questionnaire as used in Experiment 1. See section 3.3.5.4 for details. .

7.4 PROCEDURE

Figure 8 illustrates the procedure that participants followed during the experiment. The experiment took approximately 112 minutes to complete, and consisted of the TOI manipulation, a pretest, a learning section, and a posttest, and some questionnaires.

All participants first read either the entity article or the incremental article, and then completed the TOI questionnaire. They then completed a pretest in the next ten minutes. In order increase the strength of the TOI manipulation, it was necessary for participants to experience a challenge and feel like their performance was inadequate. Therefore, once the participant completed the pretest, the experimenter scored it and told every participant that they got 20% correct on the pretest and that other participants in the experiment got 80% correct. It was expected that upon receiving failure feedback, participants would connect it to the article that they had read, that is, those who read the entity article would attribute their failure to innate abilities, while those who read the incremental article would attribute it to external factors, such as the test being difficult.

The pretest was followed by talk-aloud practice with simple arithmetic problems to familiarize participants with talking aloud as they solved problems. Next, in the learning section participants completed the activities from either the invention booklet or the tell-and-practice booklet depending on the condition to which they were assigned. The learning section was videotaped and participants were told to talk aloud and say what they were thinking as they solved they problems. If they fell silent for more than a few seconds, the experimenter reminded them to keep talking aloud.

The learning materials for the invention condition were divided into three sections. The first section consisted of one invention problem on mean deviation, for which participants had 10 minutes to solve. The second section consisted of a worked example demonstrating the procedure for calculating mean deviation, followed by a practice problem. Participants spent 10 minutes of the second section. After completing the second section, participants completed the in-task goal questionnaire, which asked them to report their achievement goals during the learning activity. The third section consisted of two more invention problems – one based on standardization and another based on variability. Participants spent fifteen minutes to complete the third section.

For the tell-and-practice condition, the learning materials were divided into two sections. The first section consisted of a worked example on mean deviation followed by four practice problems. Participants spent twenty minutes on the first section. After completing the first section, participants completed the in-task goal questionnaire, which asked them to report their achievement goals during the learning activity. The second section consisted of a worked example on standardization followed by a practice problem on standardization. Participants spent 15 minutes on the second section.

After the learning section, all participants completed the test section. They had thirty minutes to complete the test section, after which they completed the following questionnaires: the TOI scale, the expectancy-value questionnaire, and the demographic questionnaire. After completing the questionnaires, participants were given a full debriefing, in which they were informed that the article that they had read at the beginning of the experiment was not a scientific article, but was created just for the sake of this study. They were also told that the initial feedback they had received about pretest performance was also false, and that the average performance on the pretest was about 20% correct, and not 80% as they had been told. Any questions or concerns that they had about the procedure were answered, and they were requested not to share the details of the experiment with others.

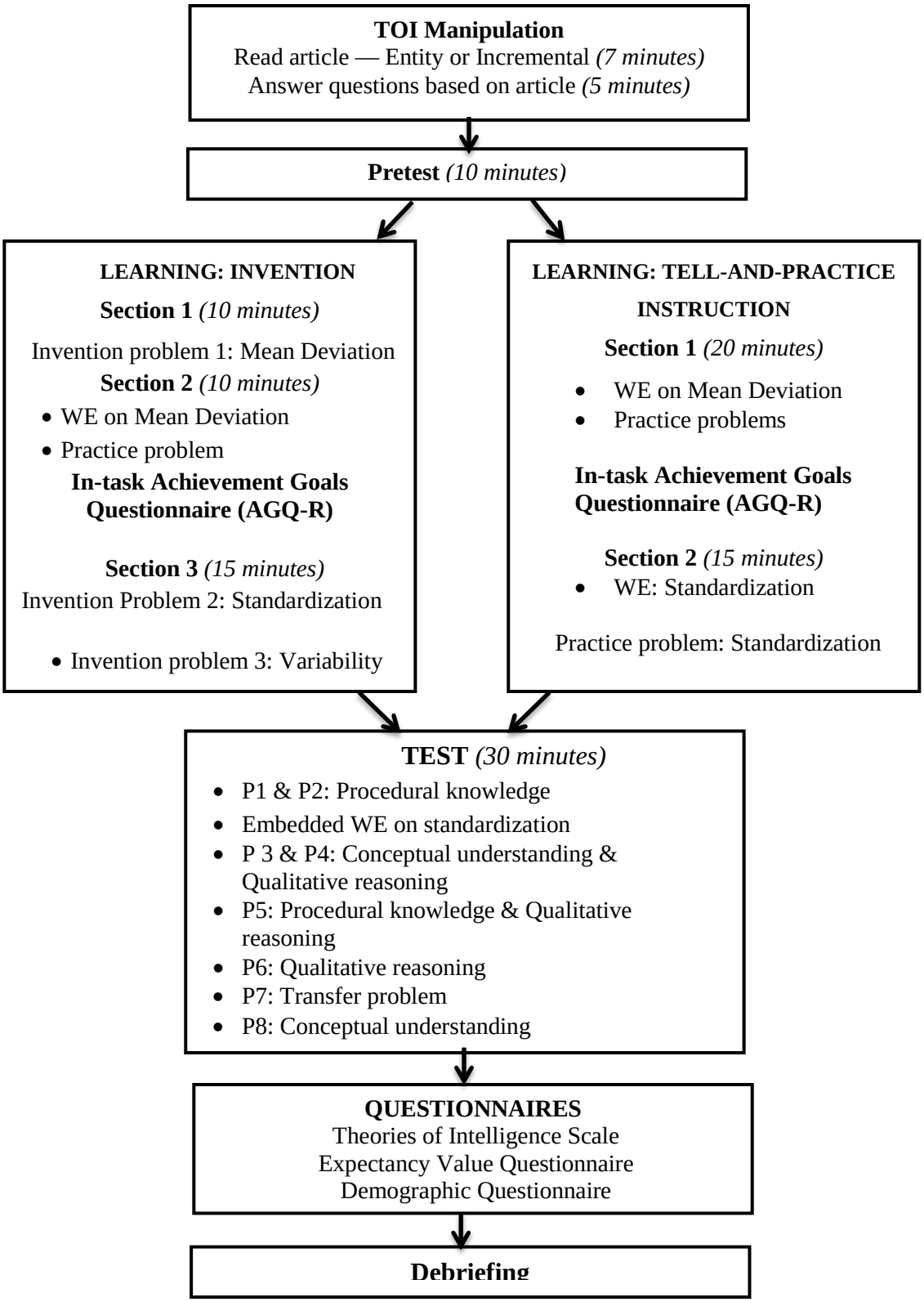


Figure 8. Flowchart of procedure

8.0 RESULTS

Results are presented in six sections. In the first section, I describe the results from two manipulations checks- the TOI questionnaire and the TOI scale. In the second and third sections, I present the pretest results and learning results respectively. In the fourth section, I present the posttest results, within which, I first describe overall posttest performance, followed by performance on each type of problem. In the fifth section, I present the results on the motivational questionnaires, that is, the in-task AGQ-R, the expectancy value questionnaire. In the sixth section, I describe posttest results using a Bayesian model selection approach for evaluating the two competing interaction hypotheses. For null hypothesis significance testing, I set the alpha level at .05 for all main effects, interactions, and planned comparisons (Keppel, 1991). I calculated effect sizes (eta squared, η_p^2) for all significant main effects, interactions, and planned comparisons. I followed the guidelines by Cohen (1988) according to which effects are regarded as small when $\eta_p^2 < .06$, medium when $\eta_p^2 < .14$, and large when $\eta_p^2 > .14$.

To establish inter-rater reliability for problems that involved qualitative scoring, 25% of the problems were first scored by two independent raters. Disagreements were resolved through discussion, and the process was repeated if the resulting kappa was less than .8. None of the problems required more than two iterations of coding. I first describe whether the manipulation to get students to adopt particular theories of intelligence was successful.

8.1.1 TOI Questionnaire

The TOI questionnaire given immediately after participants completed reading either article served as a manipulation check. Participants could obtain scores ranging from 0 to 3 on the questionnaire, with a score of 0 or 1 indicating an entity theory, and a score of 2 or 3 indicating an incremental theory (same as in Experiment 1, see section 4.1.1 for details). Two independent raters scored 25% of the questionnaires. A kappa of .95 was obtained on the first pass, so the first rater went ahead and scored the rest of the questionnaires. Out of 100 participants, only two participants gave answers inconsistent with the article that they had read, and both were in the entity condition. This result indicates that most students answered the open-ended questions consistent with the manipulation article that they had read.

8.1.2 TOI scale

As a second manipulation check, I analyzed participants' responses on the TOI scale that they completed towards the end of the experiment. The same scale as used in Experiment 1 was used. A score of 8 on the scale indicated an extremely incremental view, while 56 indicated an extremely entity view. Cronbach's alpha for the TOI scale was .935, suggesting high internal consistency. The mean TOI score of participants in the entity condition was 32.52 (SD = 9.75), whereas that of participants in the incremental condition was 22.3 (SD = 9.09). Students in the entity condition scored closer to the median (i.e., 32), compared to in the incremental condition. A t-test indicated that students who read the entity essay answered significantly differently on the TOI scale compared to those who read the incremental essay, $t(98) = 5.42$, $p < .001$. The second

manipulation check offers further evidence that participants tended to internalize the message from the article that they had read.

The TOI scale was also scored dichotomously by performing a median split, such that participants in the incremental condition who got a score between 8-31 were coded 1 for consistent, and those above 31 were coded 0 for inconsistent. Participants in the entity condition who obtained a score between 33-56 were coded 1 for consistent, and those below 33 were coded 0 for consistent. All participants whose score was 32 were coded as inconsistent. This conservative coding yielded 66 participants out of 100 who endorsed a TOI consistent with their manipulation. Of these, 13 were in the entity tell-and-practice condition, 19 were in the incremental tell-and-practice condition, 12 were in the entity invention condition, and 22 were in the incremental invention condition.

8.2 PRETEST RESULTS

The mean proportion of correct responses on the pretest across conditions was 26 % (SD = .12). Scores ranged from 10% correct to 70% correct. Cronbach's alpha for the pretest was .74 suggesting moderate to high internal consistency.

A two-way ANOVA was conducted on the pretest scores to see whether participants differed by condition at pretest. There was no difference between participants in the entity condition and those in the incremental condition, $F(1,96) = .334$, $p = .565$. There was no difference between participants in the tell-and-practice condition and the invention condition, $F(1,96) = .824$, $p = .366$. There was no significant interaction; $F(1,96) = .007$, $p = .934$. See Table 10 for means and standard deviations. This result suggests that the conditions were not

different at outset. However, given the relatively wide range of scores on the pretest, the pretest score was used as a covariate in further analyses.

Table 10. Means and standard deviations on pretest

Condition	Mean	Std. Dev	n
Entity Tell-and-Practice	0.27	0.12	25
Entity Invention	0.25	0.09	25
Incremental Tell-and-Practice	0.28	0.15	25
Incremental Invention	0.26	0.10	25

8.3 LEARNING RESULTS

Given that the tell-and-practice condition and the invention condition had different materials, the learning results will be presented separately for each of two conditions. Each of the conditions had 50 participants.

8.3.1 Invention Condition

Data for two participants from this condition were missing, so the results described here are for the remaining 48 participants.

8.3.1.1 Invention problem 1: Pitching machine problem

This first invention problem gave participants data from four different pitching machines and asked them to find out which one as the most reliable. 37 out of 48 participants could correctly identify which pitching machine was the most reliable. However, 30 out of the 37 took the average of all pitches, and took the lowest one to be the most reliable, thus arriving at the correct final answer using incorrect conceptual reasoning.

A Chi-square test for final answer showed no difference between the entity and the incremental conditions, $\chi^2(2, N = 48) = 2.947$, $p = .229$. A Chi-square test for conceptual reasoning component was also not significant $\chi^2(2, N = 48) = 1.088$, $p = .580$.

8.3.1.2 Invention problem 2: Football teams

The first problem was a word problem that required students to calculate mean deviation to determine which of two football teams had a better winning record. The answers were coded as 0 for incorrect, 1 for correct, and 2 if they did not complete the problem. A vast majority of the participants (39 out of 48) did not finish the problem. Of the 8 who finished, 3 participants chose the correct team as the final answer and 6 chose the incorrect team. The solution strategies were coded as 0 for incorrect, 1 for correct, and 2 if no strategy was given. A vast majority of the participants (45 out of 48) used the correct strategy (calculating mean deviation), whereas 3 used an incorrect strategy.

A Chi-square test for final answer showed no difference between the entity and the incremental conditions, $\chi^2(2, N = 48) = .451$, $p = .601$. A Chi-square test for the conceptual reasoning component was significant $\chi^2(2, N = 48) = 1.088$, $p = .580$.

8.3.1.3 Invention problem 3: Track Stars

In the third invention problem, students were asked to decide which one of two players' records was more impressive. Given that they competed in different events, their scores needed to be standardized in order to be compared.

Answers were coded as 0 for incorrect, 1 for correct, and 2 if they did not complete the problem. Out of 48, 29 participants did not finish the problem. Of those who finished, 15 got the correct final answer, and 4 got it incorrect. A Chi-square test for final answer showed no difference between the entity and the incremental conditions, $\chi^2(2, N = 48) = 1.296$, $p = .523$. None of the participants used a correct strategy of calculating standardized scores.

8.3.2 Tell-and-Practice Condition

Data for two participants from this condition were missing, so the results described here is for the remaining 48 participants.

8.3.2.1 Section 1: Worked example and practice problems on mean deviation

Section 1 of the tell-and-practice condition contained a worked example on mean deviation, followed by five practice problems. The first problem was a word problem that required students to calculate mean deviation to determine which of two football teams had a better winning record. The subsequent four problems were four sets of numbers, and students were required to calculate mean deviation for each of the data sets.

On the first problem, the answers were coded as 0 for incorrect, 1 for correct, and 2 if they did not complete the problem. Out of 48 participants, 32 participants got the final answer incorrect, 8 got it correct, and 8 others did not complete the problem. 26 participants used the

correct strategy, and 22 used an incorrect one. Out of the 26 participants who used the correct strategy, 7 participants chose an incorrect final answer despite using the correct strategy.

The frequencies of correct answers, incorrect answers and incomplete problems for problems 2 through 5 are summarized in table 11. Chi square tests indicated that none of these frequencies were significantly different for the entity and incremental conditions at $p = .05$.

Table 11. Performance on practice problems for the tell-and-practice condition

	Correct	Incorrect	Did not finish	χ^2	p
Problem 2	24	24	2	2.848	0.241
Problem 3	15	29	6	3.81	0.149
Problem 4	12	26	12	1.333	0.513
Problem 5	9	21	20	0.364	0.834

These results suggest that over 50% of participants successfully applied the formula for mean deviation on practice problems. However, their relatively poor performance on the word problem suggests that although they may have learned the procedure to calculate mean deviation, they may have not gained a deeper conceptual understanding of variability.

8.3.2.2 Section 2: Worked example and practice problems on standardization

Section 2 of the tell-and-practice condition consisted of a worked example that gave a conceptual explanation for standardization, followed by a practice problem that required students to compare the records of two players on different sports by standardizing their scores. Answers were coded as 0 for incorrect, 1 for correct, and 2 if they did not complete the problem. Out of 48, 40 participants did not finish the practice problem. Of those who finished, 5 got the final answer correct, and 3 got it incorrect. A Chi-square test indicated no difference between the entity and the incremental conditions, $\chi^2(2, N = 48) = 1.296$, $p = .523$.

8.4 POSTTEST RESULTS

The posttest tested three aspects of students' understanding: procedural knowledge, conceptual understanding, and qualitative reasoning. First, I will report the scores as a percent correct score for all problem types taken together. The total possible score on the posttest was 19 points, and the proportion of correct responses ranged from 19% to 95%.

A two-way ANCOVA was conducted with the pretest score as a covariate. The effect of the covariate was significant, F (1,95) = 7.616, p = .007, η_p^2 = .074. The proportion of correct responses ranged from 20% to 100% correct. As seen in Fig. 9, there was no significant difference between participants in the entity condition and the incremental condition, F (1,95) = .322, p = .572, η_p^2 = .003. There was no significant difference between students in the tell-and-practice and invention conditions. F(1,95) = 1.20, p = .276; η_p^2 = .012. There was no significant interaction, F(1,95) = 1.646, p = .203, η_p^2 = .017.

Next, I analyzed each of the three components of the post-test, viz. procedural, conceptual, and reasoning problem components separately.

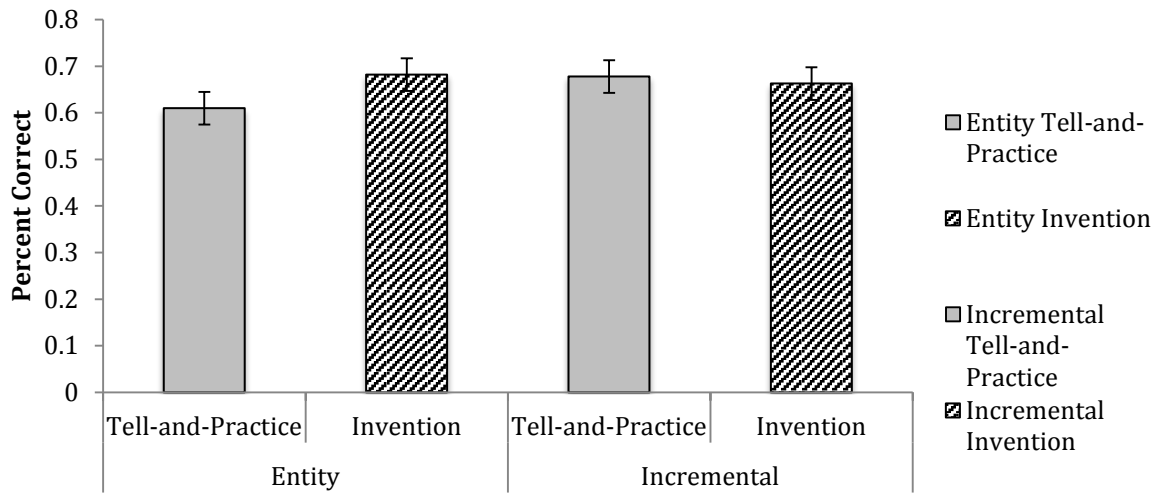


Figure 9. Posttest scores: All problem components taken together, adjusted for pretest score

8.4.1 Procedural knowledge

Next, I analyzed problems and problem components testing procedural knowledge. This included problem 1, problem 2, and the procedural component of problem 5 and problem 7 (see Materials for more details). The total possible score on the procedural knowledge measure was six points, and the proportion of correct responses ranged from 0 % to 100 % correct. Cronbach's alpha for the procedural knowledge problems was .68 indicating moderate internal consistency.

A two-way ANCOVA was conducted with pretest score as covariate. The effect of the covariate was significant, F (1,95) = 8.435, p = .005, η_p^2 = .082. As seen in Fig. 10, there was no significant difference between participants in the entity condition and those in the incremental condition, F(1,95) = .405; p = .526; η_p^2 = .004. There was no significant difference between students in the tell-and-practice and invention conditions, F(1,95) = 1.411; p = .238; η_p^2 = .015. There was no significant interaction, F(1,95) = 2.182; p = .143 ; η_p^2 = .022.

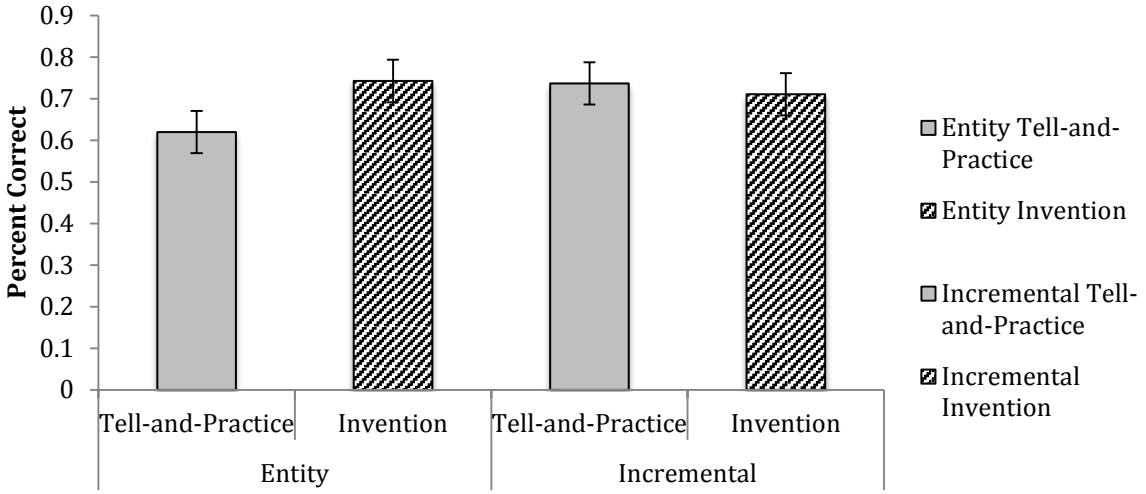


Figure 10. Posttest scores on procedural problems adjusted for pretest score

A planned comparison tested the difference between procedural scores of students in the invention condition and the tell-and-practice condition only for the entity condition, taking pretest percent correct as the covariate. The effect of the covariate was not significant, F (1,47) = .189, p = .666, η_p^2 = .004. The scores on the posttest procedural problems were marginally different, such that participants in the invention condition scored higher than those in the tell-and-practice condition, which is consistent with our hypothesis, F (1,47) = 3.096, p = .085, η_p^2 = .062.

8.4.2 Conceptual understanding problems

Next, I analyzed the scores on conceptual understanding problems. This measure included the conceptual understanding components of problems 3, 4, 7, and 8. The total possible score on the conceptual understanding measure was 4 points, and the proportion of correct responses ranged

from 25 % to 100 %. Cronbach's alpha for the conceptual understanding problems was .359 indicating low internal consistency.

A two-way ANCOVA was conducted with pretest score as covariate. The effect of the covariate was significant, F (1,97) = 5.208, p = .025, η_p^2 = .052. As seen in Fig. 11, there was no significant difference between participants in the entity condition and the incremental condition, F(1,95) = .508; p = .478; η_p^2 = .005. There was no significant difference between students in the tell-and-practice and invention conditions, F(1,95) = .400; p = .528; η_p^2 = .004. There was no significant interaction, F(1,95) = 1.571; p = .213 ; η_p^2 = .016.

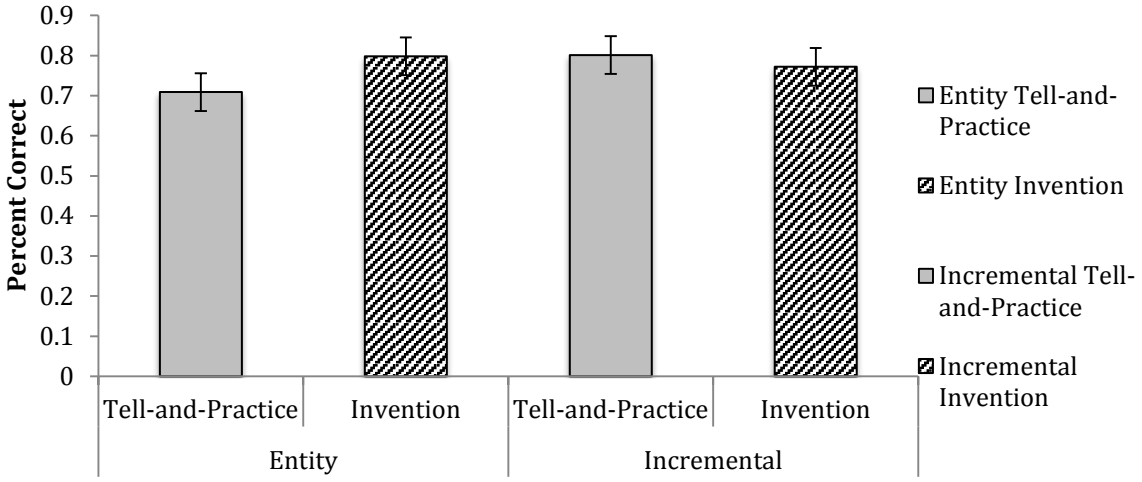


Figure 11. Posttest scores on conceptual problems adjusted for pretest score

A planned comparison tested the difference between conceptual understanding scores of participants in the invention condition and the tell-and-practice condition only for the entity condition, taking pretest percent correct as the covariate. The effect of the covariate was not significant, F (1,47) = 1.227, p = .274, η_p^2 = .025. The scores on the conceptual problems were not significantly different for participants in the invention and tell-and-practice conditions, F (1,47) = 1.311, p = .258, η_p^2 = .027.

8.4.3 Reasoning problems

Finally, I analyzed the mean scores on reasoning problems, adjusted for pretest scores. This measure included problem 6, and the reasoning components of problems 3, 4, 5, and 7. The total possible score on the reasoning problems was 9 points, and the proportion of correct responses ranged from 10 % to 100 %. Cronbach's alpha for the conceptual understanding problems was .159 indicating low internal consistency.

A two-way ANCOVA was conducted with pretest score as covariate. The effect of the covariate was not significant, F (1,97) = 2.278, p = .135, η_p^2 = .023. As seen in Fig. 10, there was no significant difference between participants in the entity condition and the incremental condition, F(1,95) = .007; p = .933; η_p^2 = .000. There was no significant difference between participants in the tell-and-practice and invention conditions, F(1,95) = .431; p = .513; η_p^2 = .005. Contrary to our hypothesis, there was no significant interaction, F(1,95) = .257; p = .613 ; η_p^2 = .003.

A planned comparison tested the difference between scores on reasoning problems for participants in the invention condition and the tell-and-practice condition only for the entity condition, taking pretest percent correct as the covariate. The effect of the covariate was not significant, F (1,47) = .685, p = .412, η_p^2 = .014. The scores on the reasoning problems were not significantly different for participants in the invention condition and the tell-and-practice condition, F (1,47) = .579, p = .451, η_p^2 = .012.

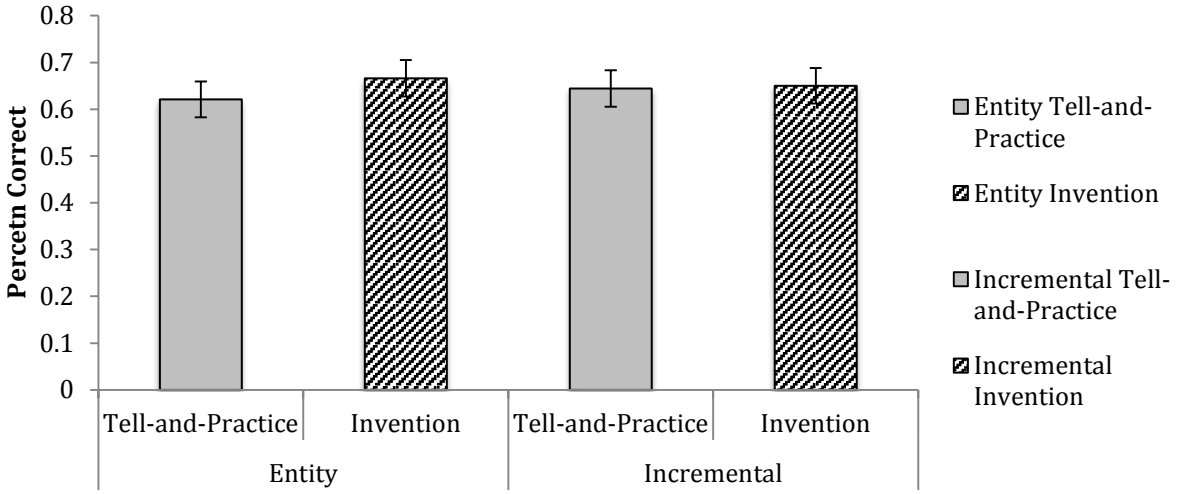


Figure 12. Posttest scores on reasoning problems adjusted for pretest score

Given that the expected effect of theory of intelligence were not found on transfer problems, I analyzed the performance of only those participants who responded with a TOI consistent with their manipulated TOI on the scale given at the end of the experiment. For this analysis, I used the dichotomously scored scale, according to which 66 participants responded consistently with their TOI.

8.4.3.1 Overall posttest scores

A two-way ANCOVA tested whether participants differed by condition on overall posttest scores, using pretest scores as a covariate. The effect of the covariate was significant, $F(1,61) = 5.213$, $p = .026$, $\eta_p^2 = .079$. There was no difference between entity theorists and incremental theorists, $F(1,61) = 1.688$, $p = .199$, $\eta_p^2 = .027$. There was no difference between those who received tell-and-practice instruction and those who invented, $F(1,61) = .181$, $p = .672$, $\eta_p^2 = .003$. There was no significant interaction, $F(1,61) = 1.223$, $p = .273$, $\eta_p^2 = .020$.

8.4.3.2 Procedural knowledge problems

A two-way ANCOVA tested whether participants differed by condition on procedural knowledge scores, using pretest scores as a covariate. The effect of the covariate was significant, $F(1,61) = 5.594$, $p = .021$, $\eta_p^2 = .084$. There was no difference between entity theorists and incremental theorists, $F(1,61) = 2.530$, $p = .117$, $\eta_p^2 = .040$. There was no difference between those who received tell-and-practice instruction and those who invented, $F(1,61) = .140$, $p = .709$, $\eta_p^2 = .002$. There was no significant interaction, $F(1,61) = 2.514$, $p = .118$, $\eta_p^2 = .040$.

8.4.3.3 Conceptual understanding problems

A two-way ANCOVA tested whether participants differed by condition on conceptual understanding, using pretest scores as a covariate. The effect of the covariate was marginally significant, $F(1,61) = 3.784$, $p = .056$, $\eta_p^2 = .058$. There was no difference between entity theorists and incremental theorists, $F(1,61) = 2.667$, $p = .108$, $\eta_p^2 = .042$. There was no difference between those who received tell-and-practice instruction and those who invented, $F(1,61) = .102$, $p = .750$, $\eta_p^2 = .002$. There was no significant interaction, $F(1,61) = .077$, $p = .782$, $\eta_p^2 = .001$.

8.4.3.4 Reasoning problems

A two-way ANCOVA tested whether participants differed by condition on reasoning, using pretest scores as a covariate. The effect of the covariate was not significant, $F(1,61) = 2.419$, $p = .125$, $\eta_p^2 = .038$. There was no difference between entity theorists and incremental theorists, $F(1,61) = .103$, $p = .749$, $\eta_p^2 = .002$. There was no difference between those who received tell-and-practice instruction and those who invented, $F(1,61) = .477$, $p = .492$, $\eta_p^2 = .008$. There was no significant interaction, $F(1,61) = .463$, $p = .499$, $\eta_p^2 = .008$.

8.5 QUESTIONNAIRE DATA

8.5.1.1 In-task achievement goal questionnaire

This questionnaire consisted of twelve items, three for each goal (mastery approach, mastery avoidance, performance approach, and performance avoidance). Participants rated them on a 5-point Likert scale (1 = *strongly disagree,* 3 = *unsure,* 5 = *strongly agree*). Scores for each goal were computed by aggregating ratings on three items representing that goal. The total possible score for each goal was 21.

Cronbach's alphas calculated for each scale were as follows: Mastery Approach: α = .822; Mastery Avoidance: α = .758; Performance Approach: α = .913; Performance Avoidance: α = .724, indicating high internal consistency. The means and standard deviations (in parentheses) for each goal by condition can be seen in table 12.

Table 12. Means and standard deviations on the AGQ-R

	Entity Tell-and-Practice	Incremental Tell-and-Practice	Entity Invention	Incremental Invention
Mastery approach	16.80 (2.96)	16.60 (3.69)	15.04 (3.84)	14.76 (3.35)
Mastery avoidance	14.08 (3.78)	13.96 (5.07)	13.60 (3.75)	14.16 (3.04)
Performance approach	14.80 (4.49)	15.16 (4.87)	14.64 (3.70)	13.28 (3.69)
Performance avoidance	14.84 (3.94)	16.16 (4.43)	15.40 (3.61)	13.96 (3.92)

Next, separate two-way ANOVAs were conducted on each of the four goals. Table 13 shows the results from the two-way ANOVAs.

Table 13. ANOVA results for AGQ-R

Condition	TOI	Instructional Condition	Interaction
Mastery approach	*ns*	$F(1,96) = 6.718$; $p = .011$**	*ns*
Mastery avoidance	*ns*	ns	*ns*
Performance approach	*ns*	*ns*	*ns*
Performance avoidance	*ns*	*ns*	$F(1,96) = 2.99$; $p = .087$*

** denotes a statistically significant effect at $p = .05$

Results indicate that students in the tell-and-practice condition reported significantly higher mastery approach goals compared to students in the invention condition. There was also a significant interaction, such that students in the tell-and-practice condition reported higher performance avoidance goals when they held incremental beliefs relative to when they held entity beliefs, while students in the invention condition reported lower performance avoidance goals when they held incremental beliefs, relative to when they held entity beliefs.

8.5.1.2 Expectancy value questionnaire

This questionnaire consisted of eleven items on a five point Likert scale, and two additional open-ended items. The first five items addressed expectancy beliefs, and the next six items addressed attainment value, intrinsic value, and utility value with two items for each construct. Scores for each construct were computed by aggregating ratings on all items representing that construct. Cronbach's alphas calculated for each construct were as follows:

Expectancy Beliefs: $\alpha = .842$; Attainment Value: $\alpha = .723$; Intrinsic Value: $\alpha = .937$; Utility Value: $\alpha = .834$. Table 14 shows the means and standard deviations for each construct.

Table 14. Means and standard deviations on the expectancy-value questionnaire

	Entity Tell-and-Practice	Incremental Tell-and-Practice	Entity Invention	Incremental Invention
Expectancy Beliefs (Total possible 25)	12.24 (4.01)	12.72 (3.08)	12.56 (3.27)	12.04 (2.53)
Attainment Value (Total possible 10)	5.20 (1.83)	6.44 (1.76)	5.48 (2.10)	5.56 (1.73)
Intrinsic Value (Total possible 10)	4.40 (2.43)	4.88 (1.86)	4.84 (1.99)	4.96 (2.28)
Utility Value (Total possible 10)	6.48 (1.74)	6.56 (1.96)	6.28 (1.65)	6.36 (1.73)

Next, separate two-way ANOVAs were conducted on each of the four constructs. Table 15 shows the results from the two-way ANOVAs.

Table 15. ANOVA results for the expectancy-value questionnaire

Condition	TOI	Instructional Condition	Interaction
Expectancy Beliefs	*ns*	*ns*	*ns*
Attainment Value	F (1,96) = 3.14, p = .079*	ns	*ns*
Intrinsic Value	*ns*	*ns*	*ns*
Utility Value	*ns*	*ns*	*ns*

** denotes a statistically significant effect at p = .05, * denotes marginal significance

Results indicate that there was a marginal main effect of theory of intelligence on attainment value, such that incremental theorists reported higher attainment value than did entity theorists.

8.6 BAYESIAN MODEL SELECTION

The conventional approach to testing differences between means is to use null hypothesis significance testing (NHST), also known as a frequentist approach. In a factorial ANOVA, the null and the alternative hypotheses for the main effect can be stated as follows:

H_0: there is no main effect,

H_a: there is a main effect,

Similarly, for the interaction effect, the null and the alternative hypotheses can be stated as:

H_0: there is no interaction effect,

H_a: there is an interaction effect,

However, when there are specific predictions about ordering of means among the four conditions, as well predicted relationships between mean differences, a Bayesian model selection approach comparing the fit of the hypotheses using a model selection criterion has been suggested as an alternative to NHST (Hoijtink, Klugkist, & Boelen, 2008; Klugkist, Laudy, & Hoijtink, 2005). The core idea behind the Bayesian approach is that a priori beliefs are updated with observed evidence and both are combined in a posterior distribution (Hoitjink et al., 2008). The hypotheses that make specific predictions for means have been called "informative hypotheses."

Informative hypotheses can be compared using the ratio of two marginal likelihood values, which is a measure for the degree of support for each hypothesis provided by the data (Klugkist et al., 2005). This ratio results in the Bayes factor, which represents the amount of evidence in favor of one hypothesis over another. The Bayes factor is composed of two

components: model fit and model complexity. A Bayes factor of 1 suggests that the hypothesis A and hypothesis B are equally supported by the data. A Bayes factor of 10 suggests that the support for hypothesis A is 10 times stronger than the support for hypothesis B. A Bayes factor of .25 suggests that the support for hypothesis B is 4 times stronger than the support for hypothesis A.

When Bayes factors for all hypotheses are calculated, they are converted into posterior model probabilities (PMPs). A posterior model probability represents the relative support for a hypothesis within a certain set of hypotheses. The relative support measure is not a real probability, but it can be loosely interpreted as the probability on a 0-1 scale that the hypothesis at hand is the best of a set of finite hypotheses after observing the data. A PMP is computed for each model under consideration, and this way an easy comparison of many models can be made. The relative fit of a hypothesis is computed by dividing its BF compared with the unconstrained hypothesis by the sum of all BFs.

In the current experiment, I tested two competing theories, derived from past research on the topic, which led to differing sets of hypotheses. By conducting a 2X 2 ANOVA, we can tell whether main effects or interactions exist, however, in the light of specific hypotheses proposed, (ordered means and interaction effects), NHST does not give us sufficient information to evaluate and compare which of the hypotheses best fit the data. Accordingly, it would be more informative to test the hypotheses using Bayesian model selection.

For analysis using the Bayesian approach, I used the BIEMS software developed by Hoijtink and colleagues (Mulder, Hoijtink, & de Leeuw, 2012). For evaluating Bayes' factor values, I used guidelines by Kass and Raftery (1995) as shown in table 16.

Table 16. Guldelines for evaluating Bayes factors

Bayes factor	Interpretation
1-3.2	Not worth more than a bare mention
3.3-10	Substantial
10-100	Strong
>100	Decisive

In the first analysis, I tested the informative hypotheses formulated in models M_1 and M_2. I first tested the hypotheses using the overall posttest scores using pretest percent correct as covariate. These hypotheses were evaluated against the unconstrained model M_0. For each hypothesis, the Bayes factor comparing the hypothesis with the unconstrained model M_0, shows if there is support in the data for the constraints (if BF > 1), or not (BF < 1). The results are presented in Table 17. Based on Kass and Raftery's (1995) guidelines, the data offer decisive evidence in favor of model M_1.

Table 17. Bayes factor values and posterior model probabilities for all problem types combined

Model	BF (against model M0)	PMP
M_0: μ_1 μ_2 μ_3 μ_4 (unconstrained)		0.01
M_1: $\mu_1 < \mu_2$; $\mu_2 = \mu_3$; $\mu_3 = \mu_4$	156.26	0.94
M_2: $\mu_2 < \mu_1$; $\mu_1 = \mu_3$; $\mu_3 = \mu_4$	8.90	0.05

Next, I tested the hypotheses using the scores on procedural problems using pretest percent correct as covariate. These hypotheses were evaluated against the unconstrained model

M0. The results are presented in Table 18. Again, based on Kass and Raftery's (1995) guidelines, the data offer decisive evidence in favor of model M1.

Table 18. Bayes factor values and posterior model probabilities for procedural problems

Model	BF (against model M0)	PMP
M0: μ_1 μ_2 μ_3 μ_4 (unconstrained)		0.01
M1: $\mu_1 < \mu_2$; $\mu_2 = \mu_3$; $\mu_3 = \mu_4$	166.44	0.96
M_2: $\mu_2 < \mu_1$; $\mu_1 = \mu_3$; $\mu_3 = \mu_4$	6.01	0.03

Next, I tested the hypotheses using the scores on conceptual problems using pretest percent correct as covariate. These hypotheses were evaluated against the unconstrained model M_0. The results are presented in Table 19. Again, based on Kass and Raftery's (1995) guidelines, the data offer decisive evidence in favor of model M_1.

Table 19. Bayes factor values and posterior model probabilities for conceptual problems

Model	BF (against model M0)	PMP
M_0: μ_1 μ_2 μ_3 μ_4 (unconstrained)		0.01
M_1: $\mu_1 < \mu_2$; $\mu_2 = \mu_3$; $\mu_3 = \mu_4$	145.20	0.89
M_2: $\mu_2 < \mu_1$; $\mu_1 = \mu_3$; $\mu_3 = \mu_4$	16.44	0.10

Finally, I tested the hypotheses using the scores on reasoning problems using pretest percent correct as covariate. These hypotheses were evaluated against the unconstrained model M0. The results are presented in Table 20. Once again, based on Kass and Raftery's (1995) guidelines, the data offer decisive evidence in favor of model M1.

Table 20. Bayes factor values and posterior model probabilities for reasoning problems

Model	BF (against model M0)	PMP
M_0: μ_1 μ_2 μ_3 μ_4 (unconstrained)		0.01
M_1: $\mu_1 < \mu_2$; $\mu_2 = \mu_3$; $\mu_3 = \mu_4$	118.88	0.77
M_2: $\mu_2 < \mu_1$; $\mu_1 = \mu_3$; $\mu_3 = \mu_4$	34.75	0.22

9.0 DISCUSSION

This experiment investigated whether students' theories of intelligence interact with the type of constructive learning activity — invention versus tell-and-practice instruction. Some prior research has shown benefits for tell-and-practice type of instruction over open-ended activities such as invention (e.g., Matlen & Klahr, 2013), whereas other work indicates that invention activities lead to deeper and more robust learning (e.g., Schwartz & Martin, 2004). Accordingly, two competing hypotheses about the interaction of type of constructive activity and theory of intelligence were tested. If invention led to better learning compared to tell-and-practice instruction, it was predicted that invention would be more beneficial to entity theorists, while incremental theorists would learn well under either instructional condition. If tell-and-practice instruction was better than invention, then again, it would benefit entity theorists more, because entity theorists are more likely to abandon an invention task in response to floundering, but incremental theorists are likely to persist regardless of the type of instruction.

9.1 EFFECT OF THEORIES OF INTELLIGENCE ON LEARNING

Just as in Experiment 1, manipulated theories of intelligence were not found to have an effect on learning. Several possible measures to strengthen the TOI manipulation were implemented in Experiment 2. Materials were modified to create conditions of challenge under which theories of

intelligence are most operative. First, the pretest made more difficult than in Experiment 1 by including a transfer problem. None of the participants could solve it correctly, and the average score on the pretest was close to 25% as opposed to 50% in Experiment 1. Second, the order of presentation of the manipulation article and pretest were changed, such that participants first read the article, and then took the pretest. Taking a difficult pretest and being challenged immediately after reading about entity theories or incremental theories was expected to underscore the connection between the pretest and the article, and strengthen the TOI manipulation. Third, failure feedback was included by telling participants they received a low score, and that most other participants scored much higher than they did. Furthermore, to better connect the manipulation article with the learning domain, a vignette regarding math performance that attributed math abilities to either innate abilities or amount of training and practice was added to the article. Finally, the posttest was made more discriminative by adding measures of procedural knowledge, conceptual understanding, and reasoning.

As in Experiment 1, both the manipulation checks (the theory of intelligence questionnaire consisting of open-ended items and the TOI scale towards the end of the experiment) suggested that students endorsed the theories of intelligence consistent with the manipulation article. Despite this, the hypothesized effect for theories of intelligence were not observed. One potential reason could be that college students are much more likely to be incremental theorists at the outset, and while their responses on the manipulation checks suggested that they adopted the manipulated theory of intelligence, the manipulation may not have been strong enough to override their original theory of intelligence. Another possible reason could be that theories of intelligence are perhaps not associated the same processes and outcomes for this population, as they are with younger age-groups. Variables such as how much

importance students place on the content to be learned or their expectancies for success on that task override the effects of theories of intelligence. In other words, even though college students may believe that they do not have an innate ability for a particular domain, they may be likely to invest time and effort in learning things that they believe to be important.

Finally, given that strengthening the motivation manipulation did not lead to the predicted effect on learning, it is possible that theories of intelligence do not affect learning outcomes in this context as predicted. When students engage in constructive activities, the variance caused by differences in beliefs are potentially minimized. Future research should examine whether TOI affect learning differently when students engage in passive activities and active activities as opposed to constructive activities. An implication for educational practice would be to ensure that students engage in constructive learning activities rather than attempt to modify their implicit theories of intelligence.

9.2 INTERACTION BETWEEN THEORIES OF INTELLIGENCE AND INSTRUCTIONAL TASK

Two competing hypotheses were tested with respect to the interaction between students' theories of intelligence and type of instruction. I predicted that if invention activities lead to better learning compared to tell-and-practice instruction, entity theorists would benefit more from them compared to incremental theorists. Incremental theorists being more likely to be constructive regardless of type of instruction were predicted to learn equally well under the two instructional conditions (Model M_1). In contrast, if direct instruction was better than invention, we would see a benefit for entity theorists, and incremental theorists would learn equally well under both

conditions (Model M_2). Both of these effects were expected to be stronger for conceptual understanding and reasoning tasks, compared to procedural tasks.

Null hypothesis significance testing revealed no differences between entity and incremental theorists, and no difference between tell-and-practice instruction and invention. No interaction was observed either. However, the Bayesian model selection approach provided some evidence for model M_2. The effect was particularly strong for procedural knowledge. The posterior model probability (PMP) was .97, indicating that model M_2 was more likely to fit the data compared to an unconstrained model that posited no relationship between the means. The PMP for reasoning problems was .77, and that for conceptual problems was .89.

If providing tell-and-practice instruction works just as well as having students engage in invention activities prior to instruction (Matlen & Klahr, 2013), why did entity students do worse under tell-and-practice instruction? Prior work indicates that entity theorists are less likely to engage in constructive processes during learning (e.g., Dahl, Bals, & Turi, 2005). When engaging in invention activities, entity theorists are encouraged to be more constructive via generation and testing of hypotheses, making predictions, making errors and trying to resolve them and so on. Incremental theorists, however, are likely to engage in such activities regardless of condition, because of which they learned well under either type of instruction.

9.3 RESULTS IN THE CONTEXT OF THE ICAP FRAMEWORK

Prior research has not yet sufficiently addressed whether certain characteristics of learners lend themselves to be suited to one type of instruction over another. For example, we know that students with incremental theories are more likely to engage in constructive activities and deeper

processing during learning (e.g., Dahl, Bals, & Turi, 2005; Stipek & Gralinski, 1996). Such students will potentially engage in constructive activities regardless of which activity is presented first. In contrast, students with entity theories will benefit more from first engaging in invention activities, because such activities would encourage them to engage in constructive activities, which they are not otherwise likely to do. When followed by tell-and-practice instruction, this will lead to a deeper conceptual understanding, because students would have had the opportunity to think constructively about these problems during the invention phase.

According to the ICAP framework, tell-and-practice instruction and invention are both constructive activities, and should lead to similar learning outcomes. However, because they are different kinds of constructive activities, students may potentially engage in different cognitive processes when learning with either type of activity. In the present experiment, students' theories of intelligence were found to have a minimal effect on learning outcomes. This suggests that as long as students are engaging in constructive processing, their TOI do not greatly affect learning at least among college-age populations. Future studies should test the effect of TOI in other types of activities – passive activities and active activities identified in the ICAP framework.

9.4 LIMITATIONS AND FUTURE DIRECTIONS

In the present study, a specific kind of invention activity was tested against tell-and-practice instruction. However, there are several other types of constructive activities that can help with invention, for example, simulations, using manipulatives, game-based discovery, and so forth. Future studies should also test the efficacy of different kinds of constructive activities in various domains.

The findings from this experiment also make a contribution towards resolving the constructive learning vs direct instruction debate. Researchers need to reevaluate whether the controversy between direct instruction and invention is a productive one. Direct instruction was found to be an effective form of instruction when learners were motivated, and had an incremental theory of intelligence. Such learners are more likely to be active, constructive learners regardless of instructional task. For example, when studying a worked example, incremental theorists are expected to self-explain, make connections to their prior knowledge, engage in analogical comparison, and such other processes that have been shown beneficial to learning. Future studies should look at process data to corroborate these expectations. Invention was beneficial to learning even entity as well as incremental students. However, the activity chosen here was not one of purely unguided discovery. Invention was supported by providing a worked example, which is a form of direct instruction. The impasses encountered during invention may have caused participants to think more deeply about the worked example, compared to simply studying the worked example. These findings suggest that the nature of processing by the learning is key to learning outcomes more than the instructional task. Instructional tasks should be designed in a way to maximize constructive processes by students, rather than focusing on labels such as “direct instruction” and “discovery learning.”

10.0 GENERAL DISCUSSION

Two experiments were designed to test whether students' theories of intelligence interact with instructional factors during learning. These experiments were designed to answer two research questions:

1. Do students with entity theories and incremental theories benefit equally from constructive and interactive activities? In Experiment 1, participants' theories of intelligence were manipulated to be either entity or incremental, and the learning activity — inventing a formula to calculate mean deviation was manipulated to be constructive (inventing individually) or interactive (inventing collaboratively). It was predicted that on procedurally simple tasks, individuals would learn better than collaborators for students with either theory of intelligence. In contrast, on complex tasks requiring deep conceptual understanding, collaborators would learn more than individuals, however, students with incremental theories would benefit more from collaboration compared to those with entity theories.

2. Do students with entity theories and incremental theories benefit equally from different kinds of constructive activities? Experiment 2 explored the interaction between students' theories of intelligence and two types of constructive learning activities. Specifically, student learning was compared under one of two conditions — a tell-and-

practice condition, in which students were given a worked example and asked to solve similar practice problems, and an invention condition in which they were asked to come up with a solution for an open-ended problem, which was followed by a worked example.

10.1 HOW DID THEORIES OF INTELLIGENCE AFFECT LEARNING?

The two experiments presented in this dissertation sought to address some of the limitations in the current literature on theories of intelligence and learning. In the next sections, I will discuss how the findings relate to and extend past work on theories of intelligence.

10.1.1 Manipulating students' theories of intelligence

First, theories of intelligence were manipulated rather than measured. In much of the prior work, theories of intelligence were measured at one or more time points, and their effect on learning outcomes was assessed via correlational measures (e.g., Dweck & Henderson, 1989; Robins & Pals, 2002). Several of these correlational studies found positive associations for incremental theories and learning outcomes. However, the causal relationship among these variables was not very clear. Studies in which path analyses were conducted showed conflicting evidence for the casual relationship between theories of intelligence and learning (e.g., Dupeyrat & Marine, 2005). Some studies found the relationship between theories of intelligence and achievement to be mediated through achievement goals (e.g., Dweck & Leggett, 1988) while others failed to find such a mediated relationship (e.g., Elliot, McGregor, & Gable, 1999). Furthermore, in studies that manipulated theories of intelligence, the predicted effect of

incremental theories on learning was not always found (e.g., Donohoe, Topping, & Hannah, 2012).

The two experiments described in this dissertation address the above limitations by conducting a carefully controlled manipulation of theories of intelligence. Across both studies, incremental theories did not lead to better learning outcomes over entity theories. This suggests that theories of intelligence perhaps do not affect learning outcomes as predicted, and it is the kinds of constructive activities in which the students engage that drive learning outcomes more so than their implicit theories of intelligence.

10.1.2 Interactions of theories of intelligence with instructional activities

A second limitation of prior work on theories of intelligence is that interactions with instructional activities are rarely tested. While prior research shows a benefit for incremental theories of intelligence for learning, certain instructional conditions may moderate the effect of theories of intelligence. For example, whether the task is performed individually or collaboratively may have a bearing on how theories of intelligence affect learning outcomes. Prior research has shown that students with incremental theories tend to show productive patterns of interaction with learning partners, compared to those with entity theories. Therefore, it was hypothesized that collaborators would stand to gain more from collaboration when they have incremental theories of intelligence.

The type of constructive activity may also affect how theories of intelligence impact learning. Certain tasks may be better suited to students with incremental theories, and others to students with entity theories. Accordingly, the right choice of learning task may offset the pitfalls

of entity theories of intelligence, whereas the wrong learning task may worsen it. However, interactions of this nature have not been tested in prior literature.

The interactions tested in the two experiments presented in this dissertation offer important insights on the effects on theories of intelligence on learning. In Experiment 1, an interaction effect was predicted such that on complex problems, dyads would learn more than singletons overall, however, dyads with incremental theories would learn significantly more than dyads with entity theories. This predicted interaction was not supported. There are three possible reasons for the lack of interaction effect. First, the predicted main effect of theories of intelligence was not observed. This may have been because the population for this study was college-age students, and for college-age students, theories of intelligence may not be the most instrumental motivational factor. It is conceivable that college students may engage in constructive learning activities when their perceived utility of the subject is high, even when they perceive their competence in the domain in terms of ability rather than effort. For example, a student may strive hard to increase her competence in statistics even if she may have an ability-based view of statistics competence, because she may place a high value on the utility of statistics knowledge.

A second reason for the lack of interaction could have been that although the problems were designed to be complex they were not as challenging as they were expected to be. Given that close to 70% of participants solved the transfer problem correctly, the complexity of the task was not sufficient to have the desired effect on performance. Thus, students' performance on these problems was similar to that on simpler problems, whereby the predicted main effects on collaboration and theories of intelligence were not observed. Indeed, incremental dyads scored the lowest on transfer problems. Consistent with the cognitive load theory, because the transfer

problems were not complex enough, collaboration could have actually hindered learning because of the communication and coordination costs imposed by collaboration.

Finally, a third possible reason for the lack of interaction effect is that theories of intelligence only minimally affect learning outcomes as predicted, and it is the type of instructional activity that determines learning outcomes more so that students' TOI. While TOI have been shown to affect learning outcomes in prior studies, many of these studies are correlational, and are conduced with younger populations. In the present context, however, TOI did not affect learning as predicted, and it is possible that they do not influence learning to the extent that instructional activities do. This possibility should be further tested in future research, by examining the interaction of TOi with other types of learning activities (for example passive activities and active activities as described in the ICAP framework).

Experiment 2 tested the interaction between theories of intelligence and different types of constructive learning activities. Two competing interaction hypotheses were proposed. Hypothesis 1 was that if invention activities led to better learning over tell-and-practice activities, entity theorists were expected to benefit more from them compared to with incremental theorists, who were expected to learn well under either instructional condition. Conversely, Hypothesis 2 was that if invention activities hurt learning by causing students to give up in the face of failing to invent a correct procedure, incremental theorists were expected to be affected less because they are more likely to be persistent and not give up in the face of failure to invent the correct solution to a problem.

Neither of the proposed hypotheses found strong support using the conventional data analysis method of null hypothesis significance testing. Weak support was found for a planned comparison in hypothesis 1, comparing the invention and tell-and-practice conditions for only

the entity condition, in which the invention condition learned marginally more than the tell-and-practice condition, only on procedural problems. A similar planned comparison on conceptual problems and reasoning problems showed no effect of instructional condition. A possible reason could be that the construct validity for the conceptual and reasoning problems was not particularly strong. Cronbach's alpha for the procedural problems was moderate at .68, for conceptual problems it was low at .359, and even lower for reasoning problems at .159. Procedural problems had been better validated in prior research (e.g., Schwartz & Martin, 2004), whereas the conceptual and reasoning measures were more novel. Improving the internal consistency of the conceptual and reasoning measures could potentially show an effect of instructional condition, especially for entity theorists.

Another possible reason for not observing the predicted effect may have been that the sample size was not sufficient to detect an effect. The type of interaction predicted is called a "quantitative interaction" in which the direction of effect is not reversed as a function of the interaction of variables. This is in contrast to a "qualitative interaction" or a crossover interaction in which the direction of effect is reversed. Quantitative interactions require a much larger sample size in order to be detected. A key criticism of using null hypothesis significance testing is that given a large enough sample, a significant effect can be obtained when comparing almost any two quantities. Therefore, an alternative approach of data analysis – a Bayesian approach was applied to test the predicted interactions, which provided some evidence for the hypothesis 1 that entity theorists would benefit from invention activities, whereas incremental theorists would learn equally well under either type of instruction.

Finally, it is possible that the expected effect of TOI was not observed because TOI do not matter to learn as much as the kinds of instructional activity do. The constructive learning

activities that students engage in potentially minimize the effect of TOI, and drive learning outcomes. Although past studies have noted an association between TOI and learning, the evidence for a causal effect is mixed. Consistent with some prior studies that did not find an effect of TOI on learning (e.g., Donohoe, Topping, & Hannah, 2012), the present experiments also noted a similar lack of effect.

10.1.3 Theories of intelligence in college students

Much of the prior research on theories of intelligence has been conducted with K-12 age students. Some studies that have used college age participants have shown mixed findings with respect to the relationship between theories of intelligence and learning. The two studies presented here extend prior work by testing the effects on theories of intelligence on a college age population.

Across both studies, the predicted effect of theories of intelligence was not observed. Participants in the study had the characteristics typical of undergrads in a large, relatively selective university. Many of students may have had incremental theories to begin with (consistent with prior research that has found that entity theories are relatively rare in college students, e.g., Doron et al., 2009). Incoming theories of intelligence could not be measured for practical reasons, but may have provided some insight into how much the manipulation actually affected students' theories of intelligence.

Further, as previously noted, college age students may not be affected by their implicit theories of intelligence as much as K-12 age students, even if the manipulation may have had the intended effect. Implicit theories of intelligence may be more instrumental in students' formative

years, however, in adults, other motivational factors may take precedence over theories of intelligence.

Also worth noting is that most students in the experiment had taken at least one college level math class before entering the experiments. Thus, they were not true novices in the domain of mathematics. In Experiment 1, problems that were expected to be complex, may not have seemed as complex, given that participants were previously exposed to advanced math concepts. Although students did not perform close to ceiling on the pretest in either experiment, past research has shown that although students may not readily recall knowledge learned at a prior time, when presented with an opportunity to relearn the material, they typically do so in a much shorter amount of time when they have previously learned the material (Gettinger, 1984).

10.2 THEORIES OF INTELLIGENCE IN RELATION TO OTHER MOTIVATIONAL FACTORS

In both studies, data were collected to see how manipulating theories of intelligence affects other motivational measures – students' achievement goals and expectancy beliefs. In the next two sections, I will discuss each of these motivational measures.

10.2.1 Achievement Goals, Theories Of Intelligence, And Learning

Some prior studies suggest that theories of intelligence operate through goals. Entity theories engender performance goals, which in turn hamper learning, while incremental theories engender

mastery goals, which promote learning (ref, ref). However, some other studies do not find much evidence for this posited relationship (ref). In Experiment 1, the manipulated TOI had no effects on students' mastery approach or performance approach goals contrary to predications. However, in terms of collaboration, singletons were found to endorse performance approach goals significantly more than collaborators. A possible explanation for the lack of effect on mastery goals could be that the instructional task was an invention task, which has been shown to spur mastery-like behaviors (e.g., Belenky & Nokes, 2012. Indeed all four conditions showed relatively high endorsement of mastery approach goals, and relatively low endorsement of mastery avoidance goals across the board. Further, given that participants' incoming achievement goals were not measured, the extent to which these were affected by the theory of intelligence manipulation cannot be stated for certain.

In Experiment 2 as well, no main effects were observed between theories of intelligence and achievement goals. However, counterintuitively, students in the tell-and-practice condition endorsed mastery approach goals significantly more than those in the invention condition. An interaction effect was also observed such that students in the tell-and-practice condition endorsed higher performance avoidance goals when they held incremental beliefs relative to when they held entity beliefs. Students in the invention condition reported lower performance avoidance goals when they held incremental beliefs relative to when they held entity beliefs.

10.2.2 Expectancy Value and Theories Of Intelligence, And Learning

The expectancy value questionnaire measured students' expectancy beliefs, attainment value, intrinsic value and utility value. A person may believe that intelligence is fixed, yet choose to engage in cognitive processes that are more typical of incremental theorists, if they believe the

to-be-learned knowledge or skills to be valuable, irrespective of their implicit TOI. If TOI were found to affect learning, it was important to tease apart that effect as distinct from the effect of the students' expectancy values on learning. Further, manipulating a students' TOI could also potentially affect heir expectancy values. In order rule out an alternative explanation, data on students' expectancy values were collected as an ancillary motivational measure.

In Experiment 1, students' expectancy values did not interact with their theories of intelligence in meaningful ways. In Experiment 2, students with incremental theories of intelligence reported marginally greater attainment value compared to those with entity theories, which means that they deemed statistics to be an important subject to learn. It is possible that this effect was observed only in Experiment 2 because of the strengthened experimental manipulation. If adopting incremental theories of intelligence causes students to place higher value on the learning task, it is yet another reason to encourage students to adopt incremental theories. Future work should explore the interaction between these two motivational variables in more details, and understand whether the relationship is causal in nature.

10.3 METHODOLOGICAL CONTRIBUTIONS

Another contribution of the current research is that hypotheses were tested using both traditional null hypothesis significance testing as well as Bayesian model selection. Although Bayesian methods are gaining ground in psychological sciences (Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, 2007), Bayesian model selection is a novel approach for testing interaction effects. One advantage of using Bayesian model selection is that is requires researchers to define their interaction hypotheses in more precise terms using ordering of means,

rather than specifying relative differences between means (Hoijtink, Klugkist, & Boelen, 2008). Another advantage of Bayesian model selection is that complex interactions can be tested using modest sample sizes, particularly when the predicted interactions are qualitative interactions (i.e., not crossover interactions), and thereby require extremely large sample sizes. In certain types of research, such large sample sizes may not be even feasible to obtain (for example, studies using experts in a domain, where large numbers of experts may just not be available to test). In such situations, Bayesian model testing offers a way to test hypotheses by providing evidence in favor of as well as against the null hypothesis.

Presently, few studies have conducted both traditional null hypothesis significance testing as well as Bayesian analysis and compared results across the two methodologies. One study (Wetzels, Matzke, Lee, Rouder, Iverson, & Wagenmakers, 2011) compared *p* values and Bayes factors using 855 published t tests in psychology. They found that while *p* values and Bayes factors almost always agree about what hypothesis is better supported by the data, they often disagree about the strength of this support. In the present studies, however, Bayesian model testing indicated strong support for the hypotheses, when null hypothesis significance testing indicated no support whatsoever. This discrepancy is likely due to the limited sample size for testing the interaction. If the sample size had been larger, a greater agreement between the Bayes factor and p values would have been found.

10.4 PRACTICAL IMPLICATIONS

The findings from the two experiments also have three important practical implications. First, across the two experiments, theories of intelligence did not have the predicted effect on

learning. The constructive activities in which students engaged were found to drive learning more than their motivational beliefs. In practical terms, instead of striving to change students' implicit beliefs of intelligence, educators may find it more effective to use constructive learning strategies in classrooms. Second, for procedural tasks, collaborative learning may not offer much of a benefit, in fact, collaboration may hamper learning because of the extra costs imposed by collaboration. Therefore, for learning simple procedures, collaborative learning does not appear to be an effective instructional choice. Finally, constructive activities involving invention are more effective in the acquisition of procedural knowledge compared to tell-and-practice instruction.

10.5 LIMITATIONS

Some limitations of the two experiments must be noted.

10.5.1 How robust was the learning?

Learning is said to be robust if it meets at least one of the following three criteria – long-term retention, transfer, and accelerated future learning (Koedinger, Corbett, & Perfetti, 2012). Long-term retention means that learning is retained for long periods of time, at least for days and even for years. Both studies measured only short-term learning, such that the posttest was immediately following instruction. To get a better understanding of how theories of intelligence interact with collaboration, future studies need to assess learning at later time points. The second criterion for

robust learning is that it transfers, that is, it can be used in situations that differ significantly from the situations present during instruction. Although transfer measures were included in assessments, given that the performance on these questions was relatively high across conditions, the transfer distance may not have been far enough. Future studies need to include better measures of transfer that test deeper conceptual understanding and reasoning. The third criteria is accelerated future learning — learning should accelerate future learning, which means that when related instruction is presented in the future, the acquired knowledge allows students to learn more quickly and/or more effectively. In the present experiments, I included a measure of preparation for future learning by embedding a worked example in the posttest and including problems that required the application of concepts learned from that worked example. However, because it was presented so close to the instruction (although slightly further apart in Experiment 2), students may not have had difficulty seeing the connection between the worked example and the target problem. Therefore, future studies need to have better tests of accelerated future learning.

10.5.2 Process data need to be analyzed

Analyzing the process data may be helpful to gain a better understanding how theories of learning interact with other variables. Evidence is mixed on whether theories of intelligence operate through goals. Some studies have found evidence that entity theories engender performance goals and incremental theories engender mastery goals, which in turn leads to adopting of different cognitive and behavioral strategies and processes. Based on the AGQ-R, there was no evidence that entity theorists endorsed performance goals or that incremental theorists endorsed mastery goals. However, singletons endorsed both performance approach and

performance avoidance goals significantly more than dyads. Although performance goals were previously thought of as disadvantageous (Midgley, Kaplan, & Middleton, 2001), subsequent research has demonstrated that performance goals are indeed adaptive in certain situations, particularly in their approach form. In Experiment 1, singletons endorsed performance goals significantly more, and also learned significantly more than dyads. This link needs to be explored in more detail. Students learning activities should be analyzed to see whether they engaged in different learning behaviors depending on which goals they endorsed. Future studies also need to manipulate students' achievement goals during learning in a collaborative and an individual context, to see whether a causal link between achievement goals and collaborative learning can be established.

The cognitive processes of incremental theorists and entity theorists need to be better understood. For example, prior research suggests that incremental theorists are more likely to engage in constructive activities such as better self-regulation, metacognitive monitoring, and planning when learning individually, and help-seeking and giving, voicing disagreements openly, and considering multiple points of view when learning collaboratively. Although in the present experiments, manipulated theories of intelligence did not have the predicted effect on learning, it may be helpful to see if entity theorists and incremental theorists actually differed in the use of learning strategies and processes.

10.5.3 External validity

As with most experiments conducted with psychology undergraduates as participants, the present experiments are also subject to the criticism that they lack strong external validity. The participants in the study were of an average age of approximately 19 years, and had the

characteristics typical of undergrads in a large, relatively selective university. Many of students may have had incremental theories to begin with (consistent with prior research that has found that entity theories are relatively rare in college students, e.g., Doron et al., 2009). Therefore, some of the findings of the two studies may be particular to the college-age populations, and future studies should examine whether the findings hold true in more diverse populations as well, e.g., K-12 age students or older students returning to college. Prior work also suggests that theories of intelligence may be domain specific, for example, a person may hold an entity theory in the domain of mathematics, but an incremental theory in the domain of music. Therefore, the present studies need to be replicated to see if the effects hold in other domains as well.

10.5.4 Relationships with other motivational constructs

Across both experiments, interesting relationships were found with other motivational variables, viz. achievement goals and expectancy values. In Experiment 1, singletons were found to endorse both performance goals significantly more than dyads. In study 2, students in the tell-and-practice condition reported significantly higher mastery approach goals compared to students in the invention condition. There was also a significant interaction, such that students in the tell-and-practice instruction reported higher performance avoidance goals when they held incremental beliefs, but those in the invention condition reported lower performance avoidance goals when they held incremental beliefs. Future work should examine a process model for the relationship between these variables.

Interesting results were also noted for some of the constructs on the expectancy-value scale. In Experiment 1, collaborators reported less intrinsic value compared to individuals, across

both motivational conditions. There was also an interaction effect for attainment value such that incremental theorists showed high attainment value compared to entity theorists when learning individually, but lower attainment value than entity theorists when learning collaboratively. In Experiment 2, incremental theorists reported marginally higher attainment value than did entity theorists. Again, the relationships between goals, expectancy-beliefs, and learning should be explored in greater detail.

11.0 CONCLUSION

The two experiments presented in this dissertation attempt to integrate current cognitively based frameworks of learning by integrating them with motivational theories. Although prior work had found a relationship between theories of intelligence and learning, theories of intelligence did not have the predicted effect on learning across both present studies. Instructional factors were found to drive learning more strongly than implicit theories of intelligence.

These results have important practical implications as previously discussed and open up interesting avenues for future research. First, future research should focus on building a path model for theories of intelligence and learning, and identify factors under which theories of intelligence affect learning outcomes. Next, research should test the interaction of theories of intelligence with passive and active learning activities, in addition to constructive and interactive activities tested in the present work. Finally, research should replicate findings from current work in classrooms settings for greater ecological validity.

Although counterintuitive findings were observed in present studies, research should continue to study cognitive factors in conjunction with motivational factors during learning. Such research programs will enhance our understanding of what factors lead to successful learning outcomes, and will inform educational practices in important ways.

APPENDIX A

A.1 ENTITY ARTICLE

THE BRAIN

The Origins of Intelligence:

Is the Nature-Nurture Controversy Resolved?

BY JEROME BERGLUND

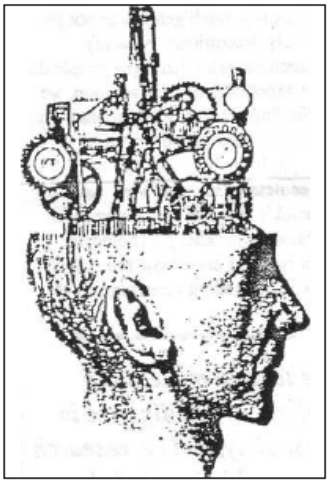
Adam Steagal is gifted. Although he is just eighteen months old, he can understand over 2000 words, has a speaking vocabulary of 500 words, and is even able to identify five different species of birds. Early in his life, Adam's parents had a hunch that he was unusual.

At the age of 8 months he was crawling and investigating everything in the Steagal household. All babies are curious, but Adam's curiosity led him to heights of baby creativity. He was not simply banging on pots and pans; Adam had learned to dismantle a toy camera and put it back together again. He had the coordination to handle small objects, the ability to remember how parts fit together, and could concentrate on the camera for almost an hour. Most children can't do what Adam was doing until they are at least three or four.

When he was ten months old, Adam's parents brought him to University of Michigan's Unit for Intelligence Research (UIR). Paula Rescorla, the director of UIR, found that Adam had an IQ of 185. Experts consider an IQ of 130 "very superior." Adam's IQ is so extreme that only one person in a million has an IQ that even comes close. Researchers like Rescorla are keenly interested in what made Adam so smart.

The traditional "is it heredity or is it environment?" question is batted around the halls of UIR on a daily basis. However, the answer is becoming increasingly clear. Current research shows that almost all of a person's intelligence is either inherited or determined at a very young age.

In the past decade, a number of comprehensive studies have been published in the United States and in Europe. These studies provide the clearest answers so far in the ongoing debate. The most significant of these studies will be published this fall in Psychological Review, a prestigious



psychological journal published in the United States

John Knowles, the author of the article and a professor at Harvard, concludes that "Intelligence seems to have a very

"The brilliance of Mozart and Einstein was mostly built into them at birth. Their genius was probably the result of their DNA."

strong genetic component. In addition, the environment seems to play a somewhat important role during the first three years of life. After the age of three, though, environmental factors (barring brain damage) seem to have almost no influence on intelligence."

Knowles spent the last decade tracing identical twins who were raised apart. In a relentless search through Latin America, Africa, and North America, he was able to locate 83 pairs of twins who were raised separately. These twins ranged in age from 7 to 51 and came from all economic levels.

Knowles had an ideal sample to study the nature-nurture question. The twins in his study were often reared in different cities by parents of different social classes. The various pairs of twins came from different countries, spoke different languages, were different ages, and he followed them for ten years. Knowles tested the subjects individually with the best "culture-fair" intelligence tests available.

Culture-fair tests measure intelligence by having people identify relationships between shapes and objects. Because the tests use only shapes and objects—not words—to measure intelligence, cultural factors don't influence people's scores. Consequently, they provide a much more accurate measure of intelligence than most other intelligence tests. In addition, culture-fair tests don't discriminate against any ethnic groups. Because Knowles used these sophisticated measures of intelligence, he was able to make stronger conclusions than have been possible in the past.

He found that twins raised apart had very similar levels of intelligence. Twins separated at birth sometimes had small differences in intelligence, ten to fifteen points. Almost all of the twins in his study that were separated after the age of three, though, had essentially identical IQs. If one twin was bright, the other was almost always equally bright. If one twin was not-so-bright, the other twin was probably not-so-bright

According to Knowles' results, up to eighty-eight percent of a person's intelligence is due to genetic factors.

About ten percent of intelligence seems to be determined during the first three years of life. This means that intelligence may be increased or decreased by only about two percent during most of a person's life. To support this claim, Knowles can show that people's intelligence did not change much in ten years. Many things in their environment shifted in that time, but their intelligence stayed constant

According to Knowles, his results suggest that "the brilliance of Mozart and Einstein was mostly built into them at birth. Their genius was probably the result of their DNA, not their schooling, not the amount of attention their parents gave them, not their own efforts to advance themselves. These great men were probably born, not made."

Other researchers are finding similar results. Hans Eysenck recently published an article supporting Knowles research. Eysenck's studies show that a person's environment does not alter his or her intelligence. He found that bright children placed in "dull" environments did not become less intelligent. Instead, they tended to take advantage of the less-gifted people around them. Relatively dull children placed in stimulating environments did not seem to get any smarter.

Needless to say, Knowles' and Eysenck's research is drawing much attention from other psychologists. Their findings are a blow to researchers who have spent years arguing that intelligence is due to environmental factors.

Leo Kamin of Princeton University is one such researcher. In the 1960s and '70s, he argued strongly that there was no good evidence to show the link between intelligence and genetics. He helped to prove that Sir Cyril Burt, a now infamous intelligence researcher, faked his data to show that intelligence was inherited. When Burt was alive, he was knighted in England for his research. When Kamin examined Burt's results, he discovered serious flaws that could only have resulted by faking the data.

This experience has made him highly skeptical of any research that establishes a relationship between intelligence and a person's genes. Because of this, he carefully examined Knowles study. He says he found "no flaws in [Knowles'] methods or his analysis. For me, these results are a little like finding out that the earth is round when you've spent 25 years trying to show it's flat. But I am a scientist first and foremost. If the best research shows that intelligence is mostly genetically determined, I will accept that fact. Knowles' research is simply the best."

Paula Rescorla at University of Michigan's UIR is excited about Eysenck's and Knowles' results. "It is about time we realize that intelligence is a genetically-determined ability. Knowing that intelligence is genetically controlled is something that can help society. We can learn to identify quickly those

"Knowing that intelligence is genetically controlled is something that can help society... By having young geniuses get the kind of training that challenges them, we will be helping them live up to their abilities. Then they can help society."

who can really benefit from specialized training. By having young geniuses get the kind of training that challenges them, we will be helping them live up to their abilities. Then they can help society. In addition, we can help less bright people find environments that won't frustrate them. We can put them into non-threatening environments that match their skills."

The eighteen month-old genius Adam Steagal seems to be in the ideal environment Rescorla described. His young brilliance is being challenged by fascinating toys and games. He is in the company of other intellectual babies. If what the researchers say is true, there is a good chance that Steagal is destined to be a leader in about the year 2040. □

Jerome Berglund is a free-lance writer from Ann Arbor, Michigan. He is a frequent contributor to Psychology Today.

A.2 INCREMENTAL ARTICLE

CROSSTALK

THE BRAIN

The Origins of Intelligence:

Is the Nature-Nurture Controversy Resolved?

BY JEROME BERGLUND

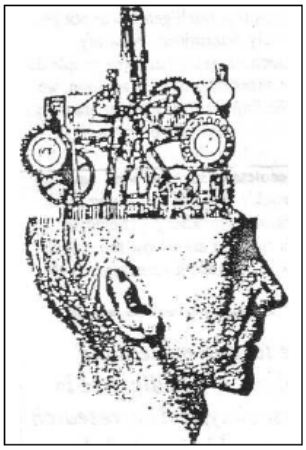
Adam Steagal is gifted. Although he is just eighteen months old, he can understand over 2000 words, has a speaking vocabulary of 500 words, and is even able to identify five different species of birds. Early in his life, Adam's parents had a hunch that he was unusual.

At the age of 8 months he was crawling and investigating everything in the Steagal household. All babies are curious, but Adam's curiosity led him to heights of baby creativity. He was not simply banging on pots and pans; Adam had learned to dismantle a toy camera and put it back together again. He had the coordination to handle small objects, the ability to remember how parts fit together, and could concentrate on the camera for almost an hour. Most children can't do what Adam was doing until they are at least three or four.

When he was ten months old, Adam's parents brought him to University of Michigan's Unit for Intelligence Research (UIR). Paula Rescorla, the director of UIR, found that Adam had an IQ of 185. Experts consider an IQ of 130 "very superior." Adam's IQ is so extreme that only one person in a million has an IQ that even comes close. Researchers like Rescorla are keenly interested in what made Adam so smart.

The traditional "is it heredity or is it environment?" question is batted around the halls of UIR on a daily basis. But, people who take the side that intelligence is genetically determined are going to be believed less and less. Current research shows that intelligence can be increased substantially by environmental factors.

In the past decade, a number of comprehensive studies have been published in the United States and in Europe. These studies provide the clearest answers so far in the ongoing debate. The most significant of these studies will be published this fall in Psychological Review, a prestigious



psychological journal published in the United States.

John Knowles, the author of the article and a professor at Harvard, concludes that, "intelligence seems to have a minimal genetic component. People may be born with a given level of intelligence, but we see increases in IQs up to 50 points when people enter stimulating environments."

"The brilliance of Leonardo da Vinci and Albert Einstein was probably due to a challenging environment. Their genius had little to do with their genetic structure."

Knowles spent the last decade tracing identical twins who were raised apart. In a relentless search through Latin America, Africa, and North America, he was able to locate 83 pairs of twins who were raised separately. These twins ranged in age from 7 to 51 and came from all economic levels.

Knowles had an ideal sample to study the nature-nurture question. The twins in his study were often reared in different cities by parents of different social classes. The various pairs of twins came from different countries, spoke different languages, were different ages, and he followed them for ten years. Knowles tested the subjects individually with the best "culture-fair" intelligence tests available.

Culture-fair tests measure intelligence by having people identify relationships between shapes and objects. Because the tests use only shapes and objects—not words—to measure intelligence, cultural factors don't influence people's scores. Consequently, they provide a much more accurate measure of intelligence than most other intelligence tests. In addition, culture-fair tests don't discriminate against any ethnic groups. Because Knowles used these sophisticated measures of intelligence, he was able to make stronger conclusions than have been possible in the past.

He found that twins raised in different environments had very different levels of intelligence. According to his results, up to 88 percent of a person's intelligence is due to environmental factors. In his study, if twins were raised in stimulating environments with motivated parents, they tended to have high IQs. Twins raised in unstimulating environments tended to have lower IQs. In an extreme case, a young girl adopted by a college professor and his wife had an IQ of 138. The genetically identical twin was raised by the real mother who was a prostitute. This girl had an IQ of only 85.

Although this evidence is very strong,

Knowles has even more evidence which may convince skeptics. He found that people in challenging environments showed substantial increases in their intelligence during the ten year study. Children and adults who were in stimulating environments had increases in IQ ranging from 15 to 48 points. People who were in unstimulating environments showed slight drops in their IQ.

According to Knowles, his results suggest that "the brilliance of Leonardo da Vinci and Albert Einstein was probably due to a challenging environment. Their genius had little to do with their genetic structure. These men are truly admirable because they were challenged and worked to overcome obstacles."

Other researchers are finding similar results. Hans Eysenck recently published an article supporting Knowles research. Eysenck's studies show that a person's level of motivation can have a profound effect on intelligence. He found that bright children placed in "dull" environments tended to become less intelligent unless they were motivated to learn. Relatively dull children placed in stimulating environments seemed to get much smarter, especially if they were rewarded for learning new things.

Needless to say, Knowles' and Eysenck's research is drawing much attention from other psychologists. Their findings are widely praised by researchers who have been trying for years to prove that intelligence is not genetically determined.

Leo Kamin of Princeton University is one such researcher. In the 1960s and '70s, he argued strongly that there was no good evidence to show the link between intelligence and genetics. He helped prove that Sir Cyril Burt, a now infamous researcher, faked his data to show that intelligence was inherited. When Burt was alive, the Queen of England knighted him for his "brilliant" research. When Kamin examined Burt's results, he discovered serious flaws that could only have resulted by faking the data.

This has led Kamin to be a bit careful before accepting any intelligence findings as the "truth." Consequently, he carefully examined Knowles' study. He found "no flaws in [Knowles'] methods or his analysis. Finally, the best available research shows what I have been arguing for for 25 years. Knowles' research is simply the best, and it shows that intelligence can be increased by stimulating environments."

Paula Rescorla at University of Michigan's UIR is also excited about Eysenck's and Knowles' results. "I think something crucial has come out of these studies – we now know that intelligence is something that motivated people can acquire. I think this idea will definitely revolutionize education in the twenty-first century. We can help motivated children find environments that will help them increase their intellectual abilities."

"I think something crucial has come out of these studies – we now know that intelligence is something that motivated people can acquire. I think these ideas will definitely revolutionize education in the twenty-first century. We can help motivated children find environments that will help them increase their intellectual abilities.

The eighteen month-old genius Adam Steagal seems to be in an ideal environment right now. His young brilliance is being challenged by fascinating toys and games. But apparently, whether or not he will be brilliant when he grows up is largely his choice. □

Jerome Berglund is a free-lance writer from Ann Arbor, Michigan. He is a frequent contributor to Psychology Today.

APPENDIX B

IN-TASK ACHIEVMENT GOAL QUESTIONNAIRE (AGQ-R)

Please indicate to what extent you agree with the following statements in regards to the problem-solving activity you are engaged in.

My aim is to completely master the material presented in this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

I am striving to do well compared to other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My goal is to learn as much as possible during this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My aim is to perform well relative to other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My aim is to avoid learning less than I possibly could during this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My goal is to avoid performing poorly compared to other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

I am striving to understand the material as thoroughly as possible during this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My goal is to perform better than the other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My goal is to avoid learning less than it is possible to learn during this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

I am striving to avoid performing worse than other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

I am striving to avoid an incomplete understanding of the material presented in this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

My aim is to avoid doing worse than other students on this activity.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

APPENDIX C

EXPERIMENT 1: PRETEST PROBLEMS

1. Find the mean, median, mode, and mean deviation of the following numbers.
   [6, 10, 5, 14, 4, 16, 3, 10]

2. Adam has offers to join two high school football teams starting the next season.
   Each number below represents the number of games a team won in a season. Taken together, the numbers represent the number of games won by two high school football teams in the 13 seasons from 1966 through 1978.
   The teams played 12 games per season each year. Which school has the better record in football? Which team should Adam choose?
   Make a graph and explain how it supports your choices.
   • Caesar Chavez High School: 7, 9, 2, 5, 8, 6, 8, 4, 6, 8, 5, 8, 5
   • Andrew Jackson High School: 8, 12, 0, 4, 12, 11, 1, 2, 8, 8, 12, 0, 5

3. Mr. Lim is arguing over the price of electricity with the power company. Mr. Lim argues that the typical family pays about $35 a month for electricity. The power company says the typical family pays about $29. The two sides picked out 11 families to see how much they pay per month. Who do you think is right and why?
   Here is what they found: [$26, $27, $27, $28, $28, $29, $36, $45, $47, $47, $48]

APPENDIX D

THEORIES OF INTELLIGENCE SCALE

Please indicate to what extent you agree or disagree with the following statements:

Everyone has certain amount of intelligence and we can't really do much to change it.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

People's intelligence is something about them that they can't change very much.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

No matter who someone is, he/she can significantly change his/her intelligence level.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

To be honest, people can't really change how intelligent they are.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

People can always substantially change how intelligent they are.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

Someone can learn new things, but he/she can't really change his/her basic intelligence.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

No matter how much intelligence people have, everyone can always change it quite a bit.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

Everyone can change even their basic intelligence level considerably.

1	2	3	4	5	6	7
Strongly Disagree			Unsure			Strongly Agree

APPENDIX E

EXPECTANCY=VALUE QUESTIONNAIRE

1. How good are you at statistics?

1	2	3	4	5
Not good				Very good

2. If you give 5 to the best student at statistics and 1 to the worst, what you give to yourself?

1	2	3	4	5
Not good				Very good

3. Some people are better in one subject than in another. For example, you might be better in math than in science. Compared to most of your other courses, how would you rate your knowledge of statistics?

1	2	3	4	5
Not good				Very good

4. How well do you think you are doing at learning statistics?

1	2	3	4	5
Not good				Very good

5. How well do you keep up your knowledge of statistics?

1	2	3	4	5
Not good				Very good

6. How important do you think statistics is for you?

1	2	3	4	5
Not very important				Very important

7. Compared to math and science, how important is it for you to learn statistics content?

1	2	3	4	5
Not very important				Very important

8. In general, how fun do you think learning about statistics is?

1	2	3	4	5
Not very fun				Very fun

9. How much do you like learning about statistics?

1	2	3	4	5
Don't like it at all				Like it very much

10. Some things that you learn in school help you do things better outside of school, that is, they are useful. For example, learning about plants at school might help you grow a garden at home. How useful do you think the concepts you learned in statistics are?

1	2	3	4	5
Not useful at all				Very useful

11. Compared to your other courses, how useful are the skills learned in statistics?

1	2	3	4	5
Not useful at all				Very useful

12. If there is anything that you don't like about statistics, what would that be? Why?

_____

_____

_____

13. If you had a choice, would you rather learn about statistics? Why?

_____

_____

_____

APPENDIX F

DEMOGRAPHIC QUESTIONNAIRE

1. Name: _____

2. Age: ____ years

3. Gender: ☐ F ☐ M

4. High School Background

1. Graduating Rank (circle highest appropriate percentile)

   [lower half] [upper half] [highest quarter] [highest tenth]

2. GPA (circle highest approximate range)

   [below 2.00] [2.00 – 2.50] [2.50 – 3.00] [3.00 – 3.50] [3.50 – 4.00]

3. SAT Scores (circle highest approximate range)

   Verbal: [200-300] [300-400] [400-500] [500-600] [600-700] [700-800]

   Math: [200-300] [300-400] [400-500] [500-600] [600-700] [700-800]

5. College Background

1. Year in college (circle answer)

   [freshman] [sophomore] [junior] [senior] [other]

2. Current GPA (circle highest approximate range)

   [below 2.00] [2.00 – 2.50] [2.50 – 3.00] [3.00 – 3.50] [3.50 – 4.00]

3. Major Field of Study: _____ Minor: _____

6. Race/Ethnicity:

☐ Non-Hispanic White/ Caucasian ☐ Hispanic
☐ African American/Black ☐ Asian/Pacific Islander
☐ Asian Indian ☐ Native American
☐ Other _____ ☐ Do not want to specify

7. Please list any college level mathematics and/ or statistics courses you are currently taking or have taken in the past two years, including AP courses:

_____

_____

BIBLIOGRAPHY

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1-18. doi: 10.1037/a0021017

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5-11. doi: 10.3102/0013189X025004005

Andersson, J., & Rönnberg, J. (1995). Recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology, 9*(3), 199-211. doi: 10.1002/acp.2350090303

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*(2), 113-125. doi: 10.1006/jesp.2001.1491

Arthur, W., Jr., Day, E. A., Bennett Jr, W., McNelly, T. L., & Jordan, J. A. (1997). Dyadic versus individual training protocols: Loss and reacquisition of a complex skill. *Journal of Applied Psychology, 82*(5), 783-791. doi: 10.1037/0021-9010.82.5.783

Azmitia, M. (1988). Peer Interaction and Problem Solving: When Are Two Heads Better Than One? *Child Development, 59*(1), 87-96. doi: 10.2307/1130391

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences, 21*(3), 399-432. doi: 10.1080/10508406.2011.651232

Belenky, D. M., & Nokes-Malach, T. J. (2013). Mastery-approach goals and knowledge transfer: An investigation into the effects of task structure and framing instructions. *Learning and Individual Differences, 25*, 21-34. doi: 10.1016/j.lindif.2013.02.004

Bergen, R. S. (1991). *Beliefs about intelligence and achievement-related behaviors.* (Ph.D. Dissertation), University of Illinois at Urbana-Champaign, Chicago.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246-263. doi: 10.1111/j.1467-8624.2007.00995.x

Bossert, S. T. (1988). Cooperative activities in the classroom. *Review of research in education, 15*, 225-250. doi: http://www.jstor.org/stable/1167365

Braasch, J. L. G., Bråten, I., Strømsø, H. I., & Anmarkrud, Ø. (2014). Incremental theories of intelligence predict multiple document comprehension. *Learning and Individual Differences, 31*, 11-20. doi: 10.1016/j.lindif.2013.12.012

Brunstein, A., Betts, S., & Anderson, J. R. (2009). Practice enables successful learning under minimal guidance. *Journal of Educational Psychology, 101*(4), 790.

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139*(3), 655-701. doi: 10.1037/a0029531

Chi, M. T. H. (2009). Active-Construtive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73-105. doi: 10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145-182. doi: doi:10.1207/s15516709cog1302_1

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477. doi: 10.1207/s15516709cog1803_3

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219-243.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237-248. doi: 10.3102/00028312019002237

Cotton, J. L., & Cook, M. S. (1982). Meta-analyses and the effects of various reward systems: Some different conclusions from Johnson et al. *Psychological Bulletin, 92*(1), 176-183.

Crooks, S. M., Klein, J. D., Savenye, W., & Leader, L. (1998). Effects of cooperative and individual learning during learner-controlled computer-based instruction. *The Journal of Experimental Education, 66*(3), 223-244. doi: 10.1080/00220979809604406

Dahl, T. I., Bals, M., & Turi, A. L. (2005). Are students' beliefs about knowledge and learning associated with their reported use of learning strategies? *British Journal of Educational Psychology, 75*(2), 257-273. doi: 10.1348/000709905X25049

Darnon, C., Harackiewicz, J. M., Butera, F., Mugny, G., & Quiamzade, A. (2007). Performance-approach and performance-avoidance goals: When uncertainty makes a difference. *Personality and Social Psychology Bulletin, 33*(6), 813-827. doi: 10.1177/0146167207301022

Darnon, C., Muller, D., Schrager, S. M., Pannuzzo, N., & Butera, F. (2006). Mastery and performance goals predict epistemic and relational conflict regulation. *Journal of Educational Psychology, 98*(4), 766-776.

Dean Jr, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384-397.

Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches*. Oxford: Elsevier.

Donohoe, C., Topping, K., & Hannah, E. (2012). The impact of an online intervention (Brainology) on the mindset and resiliency of secondary school pupils: A preliminary mixed methods study. *Educational Psychology, 32*(5), 641-655.

Doron, J., Stephan, Y., Boiché, J., & Scanff, C. L. (2009). Coping with examinations: Exploring relationships between students' coping strategies, implicit theories of ability, and perceived control. *British Journal of Educational Psychology, 79*(3), 515-528. doi: 10.1348/978185409X402580

Druckman, D., & Bjork, R. A. (1994). *Learning, Remembering, Believing. Enhancing Human Performance*: National Academy Press, 2101 Constitution Avenue, NW, Washington, DC 20418 ($39.95).

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology, 84*(3), 290-299. doi: 10.1037/0022-0663.84.3.290

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4-58.

Dupeyrat, C., & Mariné, C. (2005a). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology, 30*(1), 43-59. doi: 10.1016/j.cedpsych.2004.01.007

Dupeyrat, C., & Mariné, C. (2005b). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology, 30*(1), 43-59. doi: 10.1016/j.cedpsych.2004.01.007

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*(10), 1040-1048. doi: 10.1037/0003-066X.41.10.1040

Dweck, C. S. (2000). *Self-theories: Their role in motivation, persoanlity, and development*. Philadelphia: Taylor & Francis.

Dweck, C. S., & Bempechat, J. (1983). Children's theories of intelligence: Consequences for learning. In S. G. Paris, G. M. Olson & H. W. Stevenson (Eds.), *Learning and motivation in the classroom* (pp. 239-256).

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273. doi: 10.1037/0033-295X.95.2.256

Dweck, C. S., & Molden, D. C. (2005). Self-theories. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 122-140). New York: The Guilford Press.

El-Alayli, A., & Baumgardner, A. (2003). If at first you don't succeed, what makes you try, try again? Effects of implicit theories and ability feedback in a performance-oriented climate. *Self and Identity, 2*(2), 119-135. doi: 10.1080/15298860309031

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*(3), 549-563. doi: 10.1037/0022-0663.91.3.549

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology, 100*(3), 613-628. doi: 10.1037/0022-0663.100.3.613

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology, 54*(1), 5-12. doi: 10.1037/0022-3514.54.1.5

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2002). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*(1), 47-64. doi: 10.1016/j.lindif.2003.08.002

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction, 22*, 47-61. doi: 10.1016/j.learninstruc.2011.06.002

Gadgil, S., & Nokes- Malach, T. J. (2012). Overcoming collaborative inhibition through error correction: A classroom experiment. *Applied Cognitive Psychology, 26*(3), 410-420. doi: 10.1002/acp.1843

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding *Journal of Educational Psychology, 95*(2), 393-408. doi: 10.1037/0022-0663.95.2.393

Gettinger, M. (1984). Individual differences in time needed for learning: A review of literature. *Educational Psychologist, 19*(1), 15-29.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104-137. doi: 10.3102/00028312031001104

Greene, J. A., Costa, L.-J., Robertson, J., Pan, Y., & Deekens, V. M. (2010). Exploring relations among college students' prior knowledge, implicit theories of intelligence, and self-regulated learning in a hypermedia environment. *Computers & Education, 55*(3), 1027-1043. doi: 10.1016/j.compedu.2010.04.013

Hammer, D. (1997). Discovery Learning and Discovery Teaching. *Cognition & Instruction, 15*(4), 485-529. doi: doi:10.2307/3233776

Henderson, V. L., & Dweck, C. S. (1990). Motivation and achievement. In S. S. Feldman & G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 308-329). Cambridge, MA: Harvard University Press.

Hoijtink, H., Klugkist, I., & Boelen, P. A. (2008). *Bayesian evaluation of informative hypotheses*. New York: Springer.

Hong, Y., Chiu, C., & Dweck, C. (1995). Implicit Theories of Intelligence *Efficacy, agency, and self-esteem* (pp. 197-216): Springer.

Hong, Y., Chiu, C., Dweck, C. S., Lin, D., M. S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77*(3), 588-599. doi: 10.1037/0022-3514.77.3.588

Horton, P. B., McConney, A. A., Gallo, M., Woods, A. L., Senn, G. J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education, 77*(1), 95-111. doi: 10.1002/sce.3730770107

Howell, A. J., & Buro, K. (2009). Implicit beliefs, achievement goals, and procrastination: A mediational analysis. *Learning and Individual Differences, 19*(1), 151-154. doi: 10.1016/j.lindif.2008.08.006

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365-379. doi: 10.3102/0013189X09339057

Johnson, D. W., Johnson, R. T., & Smith, K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review, 19*(1), 15-29. doi: 10.1007/s10648-006-9038-8

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (1989). Impact of goal and resource interdependence on problem-solving success. *Journal of Social Psychology, 129*(5), 621-629. doi: 0.1080/00224545.1989.9713780

Jones, B. D., Wilkins, J. L. M., Long, M. H., & Wang, F. (2012). Testing a motivational model of achievement: How students' mathematical beliefs and interests are related to their achievement. *European Journal of Psychology of Education, 27*(1), 1-20. doi: 10.1007/s10212-011-0062-9

Kammrath, L. K., & Dweck, C. S. (2006). Voicing conflict: Preferred conflict strategies among incremental and entity theorists. *Personality and Social Psychology Bulletin, 32*(11), 1497-1508. doi: 10.1177/0146167206291476

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45-83. doi: 10.1080/10508406.2011.591717

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association, 90*(430), 773-795. doi: 10.1080/01621459.1995.10476572

Kirschner, F. (2009). *United Brains for Complex Learning: A cognitive-load approach to collaborative learning efficiency.* (PhD), Open Universiteit, Netherlands.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009a). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review, 21*(1), 31-42. doi: 10.1007/s10648-008-9095-2

Kirschner, F., Paas, F., & Kirschner, P. A. (2009b). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior, 25*(2), 306-314. doi: 10.1016/j.chb.2008.12.008

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75-86. doi: 10.1207/s15326985ep4102_1

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661-667. doi: 10.1111/j.0956-7976.2004.00737.x

Klugkist, I., Laudy, O., & Hoijtink, H. (2005). Inequality constrained analysis of variance: a Bayesian approach. *Psychological Methods, 10*(4), 477-493. doi: 10.1037/1082-989X.10.4.477

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*(3), 239-264. doi: 10.1007/s10648-007-9049-0

Koedinger, K. R., Corbett, A. T., & Perfetti, C. A. (2012). The Knowledge‐ Learning‐ Instruction framework: Bridging the science‐ practice chasm to enhance robust student learning. *Cognitive Science, 36*(5), 757-798. doi: 10.1111/j.1551-6709.2012.01245.x

Lee, H. S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with it? *Annual Review of Psychology, 64*, 445-469. doi: 10.1146/annurev-psych-113011-143833

Leidner, D. E., & Fuller, M. (1997). Improving student learning of conceptual information: GSS supported collaborative learning vs. individual constructive learning. *Decision Support Systems, 20*(2), 149-163. doi: 10.1016/S0167-9236(97)00004-3

Leondari, A., & Gialamas, V. (2002). Implicit theories, goal orientations, and perceived competence: Impact on students' achievement behavior. *Psychology in the Schools, 39*(3), 279-291. doi: 10.1002/pits.10035

Lepper, M. R., & Chabay, R. W. (1985). Intrinsic motivation and instruction: Conflicting views on the role of motivational processes in computer-based education. *Educational Psychologist, 20*(4), 217-230. doi: 10.1207/s15326985ep2004_6

Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology, 41*(1), 585-634. doi: 10.1146/annurev.ps.41.020190.003101

Levy, I., Kaplan, A., & Patrick, H. (2004). Early adolescents' achievement goals, social status, and attitudes towards cooperation with peers. *Social Psychology of Education, 7*(2), 127-159. doi: 10.1023/B:SPOE.0000018547.08294.b6

Lewis, M. W., & Anderson, J. R. (1985). Discrimination of operator schemata in problem solving: learning from examples. *Cognitive Psychology, 17*, 26-65. doi: 10.1016/0010-0285(85)90003-9

Lim, W., Plucker, J. A., & Im, K. (2002). We are more alike than we think we are: Implicit theories of intelligence with a Korean sample. *Intelligence, 30*(2), 185-208. doi: 10.1016/S0160-2896(01)00097-6

Lumpe, A. T., Haney, J. J., & Czerniak, C. M. (1998). Science teacher beliefs and intentions regarding the use of cooperative learning. *School Science and Mathematics, 98*(3), 123-135. doi: 10.1111/j.1949-8594.1998.tb17405.x

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social cognitive and affective neuroscience, 1*(2), 75-86. doi: 10.1093/scan/nsl013

Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children's acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*(3), 621-634. doi: 10.1007/s11251-012-9248-z

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14-19. doi: 10.1037/0003-066X.59.1.14

McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction, 7*(2), 129-159. doi: 10.1207/s1532690xci0702_3

Meudell, P. R., Hitch, G. J., & Kirby, P. (1992). Are two heads better than one? Experimental investigations of the social facilitation of memory. *Applied Cognitive Psychology, 6*(6), 525-543.

Midgley, C., Kaplan, A., & Middleton, M. (2001). Performance-approach goals: Good for what, for whom, under what circumstances, and at what cost? *Journal of Educational Psychology, 93*(1), 77-86.

Miele, D. B., & Molden, D. C. (2010). Naive theories of intelligence and the role of processing fluency in perceived comprehension. *Journal of Experimental Psychology: General, 139*(3), 535-557. doi: 10.1037/a0019745

Mulder, J., Hoijtink, H., & de Leeuw, C. (2012). BIEMS: A Fortran90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software, 46*(2), 1-39.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413-448.

Nichols, J. D., White, J. J., & Price, M. (2006). Beliefs of intelligence, knowledge acquisition, and motivational orientation: A comparative analysis of Hispanic/Latino and Anglo youth. *Multicultural Perspectives, 8*(4), 39-48. doi: 10.1207/s15327892mcp0804_7

Niiya, Y., Brook, A. T., & Crocker, J. (2010). Contingent self-worth and self-handicapping: Do incremental theorists protect self-esteem? *Self and Identity, 9*(3), 276-297. doi: 10.1080/15298860903054233

Nokes-Malach, T. J., & Belenky, D. M. (2011). Incorporating motivation into a theoretical framework for knowledge transfer *The psychology of learning and motivation: Advances in research and theory* (Vol. 55): Academic Press.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17(3)*, 398-415.

Paulsen, M. B., & Feldman, K. A. (2007). The conditional and interaction effects of epistemological beliefs on the self-regulated learning of college students: Cognitive and behavioral strategies. *Research in Higher Education, 48*(3), 353-401. doi: 10.1007/s11162-006-9029-0

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16* (385-407). doi: 10.1007/s10648-004-0006-x

Pintrich, P. R., Conley, A. M. M., & Kempler, T. M. (2003). Current issues in achievement goal theory and research. *International Journal of Educational Research, 39*(4-5), 319-337. doi: 10.1016/j.ijer.2004.06.002

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*(2), 167-199. doi: http://www.jstor.org/stable/1170472 .

Plaks, J. E., & Chasteen, A. L. (2013). Entity versus incremental theories predict older adults' memory performance. *Psychology and aging, 28*(4), 948-957. doi: 10.1037/a0034348

Pociask, S., & Rajaram, S. (2014). The Effects of Collaborative Practice on Statistical Problem Solving: Benefits and Boundaries. *Journal of Applied Research in Memory and Cognition*, (advance online publication). doi: 10.1016/j.jarmac.2014.06.005

Poortvliet, P. M., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2007). Achievement goals and interpersonal behavior: How mastery and performance goals shape information exchange. *Personality and Social Psychology Bulletin, 33*(10), 1435-1447. doi: 10.1177/0146167207305536

Reiser, B. J., Copen, W. A., Ranney, M., Hamid, A., & Kimberg, D. Y. (1994). Cognitive and motivational consequences of tutoring and discovery learning, Technical Report: Institute for the Learning Sciences, Northwestern University.

Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity, 1*(4), 313-336. doi: 10.1080/1529886029010680 5

Roedel, T. D., & Schraw, G. (1995). Beliefs about intelligence and academic goals. *Contemporary Educational Psychology, 20*(4), 464-468. doi: 10.1006/ceps.1995.1033

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16*(2), 225-237.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences, 2*(3), 235-276.

Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184. doi: 10.1207/s1532690xci2202_1

Senko, C., Hulleman, C. S., & Harackiewicz, J. M. (2011). Achievement goal theory at the crossroads: Old controversies, current challenges, and new directions. *Educational Psychologist, 46*(1), 26-47.

Shih, S. S. (2007). The role of motivational characteristics in Taiwanese sixth graders' avoidance of help seeking in the classroom. *The elementary school journal, 107*(5), 473-495.

Simon, B., Hanks, B., Murphy, L., Fitzgerald, S., McCauley, R., Thomas, L., & Zander, C. (2008). *Saying isn't necessarily believing: influencing self-theories in computing.* Paper presented at the Proceedings of the Fourth international Workshop on Computing Education Research.

Simon, H. A. (2000). Observations on the sciences of science learning. *Journal of applied developmental psychology, 21*(1), 115-121.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811. doi: 10.1037/0022-3514.69.5.797

Stipek, D., & Gralinski, J. H. (1996). Children's beliefs about intelligence and school performance. *Journal of Educational Psychology, 88*(3), 397-407. doi: 10.1037/0022-0663.88.3.397

Stump, G., Husman, J., Chung, W., & Done, A. (2009). *Student beliefs about intelligence: Relationship to learning.* Paper presented at the 39th ASEE/IEEE Frontiers in Education Conference, San Antonio, TX.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning Algebra. *Cognition and Instruction, 2*(1), 59-89. doi: doi:10.1207/s1532690xci0201_3

Tarmizi, R. A., & Sweller, J. (1988). Guidance During Mathematical Problem Solving. *Journal of Educational Psychology, 80*(4), 424-436.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition & Instruction, 21*(3), 209-249.

Wagenmakers, E. J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic bulletin & review, 14*(5), 779-804.

Weldon, M. S., & Bellinger, K. D. (1997). Collective memory: Collaborative and individual processes in remembering. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(5), 1160-1175. doi: 10.1037/0278-7393.23.5.1160

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology an empirical comparison using 855 t tests. *Perspectives on Psychological Science,6*(3), 291-298.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science, 40*(4), 711-730.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81.

Williams (1993). A comprehensive review of learner-control: The role of learner characteristics. In Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology Sponsored by the Research and Theory Division.

Winne, P. H., & Nesbit, J. C. (2010). The psychology of academic achievement. *Annual Review of Psychology, 61*, 653-678. doi: 10.1146/annurev.psych.093008.100348

Yetter, G., Gutkin, T. B., Saunders, A., Galloway, A. M., Sobansky, R. R., & Song, S. Y. (2006). Unstructured collaboration versus individual practice for complex problem solving: A cautionary tale. *The Journal of Experimental Education, 74*(2), 137-160.