ABSTRACT

Title of Dissertation: THE CONTRIBUTIONS OF CRYSTALLIZED CROSS-DOMAIN KNOWLEDGE AND FLUID RELATIONAL REASONING ABILITY TO NINTH- AND TWELFTH-GRADE STUDENTS' PERFORMANCE ON SCHOLASTIC APTITUDE AND CONTENT-SPECIFIC ACHIEVEMENT MEASURES

Peter Baggetta
Doctor of Philosophy
2019

Dissertation directed by: Dr. Patricia A. Alexander, Distinguished University Professor, Department of Human Development and Quantitative Methodology

This study investigated how measures of both crystallized cross-domain knowledge and fluid relational reasoning abilities contribute to high-school students' scholastic aptitude and content-specific achievement. The participants for this study were 211 ninth-grade and 76 twelfth-grade students enrolled in an all-male parochial high school. A series of multivariate multiple linear regression tests were conducted to examine the ability of the three crystallized cross-domain knowledge predictor variables (i.e., Language Skills, Mathematics, Reading Comprehension) and fluid reasoning ability predictor variable (Test of Relational Reasoning - TORR) to predict performance of ninth-grade students' scholastic aptitude (Preliminary Scholastic Assessment Test - PSAT), and content-specific achievement (Social Studies, Algebra, and Spanish final examinations); and performance of twelfth-grade students' content-specific achievement (Advanced Placement Psychology examination). Results of these analyses revealed that fluid relational reasoning was the strongest unique predictor of performance for ninth-grade students on the PSAT, Algebra and Spanish final examinations, and for twelfth-grade students on the AP Psychology examination. Crystallized cross-domain was found to be not as strong of a predictor as fluid relational reasoning on the five outcome measures. Results from this study suggest that students who have greater fluid relational reasoning abilities may perform better on these assessments. The research also includes delimitations, practical limitations for educators, and suggestions for future research to expand the scope of this study.

THE CONTRIBUTIONS OF CRYSTALLIZED CROSS-DOMAIN KNOWLEDGE AND FLUID RELATIONAL REASONING ABILITY TO NINTH- AND TWELFTH-GRADE STUDENTS' PERFORMANCE ON SCHOLASTIC APTITUDE AND CONTENT-SPECIFIC ACHIEVEMENT MEASURES

by

Peter Baggetta

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Dr. Patricia Alexander, Professor, Chair
Dr. Kathryn Wentzel, Professor
Dr. Donald J. Bolger, Associate Professor
Dr. Doug Lombardi, Associate Professor
Dr. Diane J. Ketelhut, Associate Professor

Acknowledgements

There are so many people who have supported, guided, and challenged me on this long doctoral journey that I will always be incredibly grateful for. First and foremost is my advisor, mentor and friend, Dr. Patricia Alexander, without whom I would never have even thought of undertaking this journey. She has been there every step of the way, through both the highs and the lows, and never giving up on me. Outside of my parents and familia, Patricia has been the single greatest influence of my life. I can never thank her enough for her wisdom, guidance, and friendship.

I am grateful for the support of all of the faculty and staff, past and present, in the Human Development and Quantitative Methodology Department, and in particular, Drs. DJ Bolger and Ann Battle. Thank you to my committee members Drs. Kathy Wentzel, DJ Bolger, Doug Lombardi, and Diane Ketelhut for your patience and feedback. I also want to thank Dr. Karen Murphy from Penn State, for her numerous words of encouragement and advise. A massive thank you to my fellow "Alexander Lab Family" members, both previous and current – you made me a better academic, researcher, learner, and person.

I am thankful for the friendship and support of my colleagues at school, especially my Headmaster, Tom Every, who allowed me to conduct this research at the school. To the students with whom I have had the privilege to work with, you inspire me to be a better educator, and provide me with tremendous purpose and joy every day.

To Angelina Jao who did a fantastic job of editing and providing feedback on the study and paper. Thank you for your friendship and persistence in constantly both encouraging and pushing me to finish the dissertation.

To my familia and friends back on Guahan, I miss you all very much, and you have continually supported my journey away from home through the years - Dångkulu na si Yu'us ma'åse'!!

To my parents, Frank and Fuki Baggetta, who have supported me no matter what throughout my life, and for the excellent values of dedication, hard work, and honesty, you have imparted to me - You truly are my role models and heroes!!

And most of all God, who makes all things possible. He has always looked out for me as I continue to trust in his plan for me. "Ad majorem Dei gloriam!!"

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

"Whilst the empirical thinking is only reproductive, reasoning is productive. An empirical, or 'rule-of-thumb,' thinker can deduce nothing from data with whose behavior and associates in the concrete he is unfamiliar. But put a reasoner amongst a set of concrete objects which he has neither seen nor heard of before, and with a little time, if he is a good reasoner, he will make such inferences from them as will quite atone for his ignorance. Reasoning helps us out of unprecedented situations"

William James (1890)

The prediction of academic performance has been a central research topic in educational psychology for over a century (Lubinski, 2004; Petrides, Chamorro-Premuzic, Frederickson, & Furnham, 2005). How students perform academically is thought to be influenced to some degree by the traits and characteristics they bring with them into the classroom (i.e., individual differences; Floyd, Evans, & McGrew, 2003; Stanovich & West, 1997). Simply defined, *individual differences* are an array of cognitive (e.g., intelligence, strategic processing, and executive functions) and non-cognitive characteristics (e.g., personality and motivation) unique to each individual (Ackerman & Lohman, 2006; Ennis, 1997).

Among individual difference characteristics, *intelligence*, the general term for one's ability to acquire and utilize knowledge and skills, has been suggested as the best single predictor of academic readiness and achievement (Kuncel, Hezlett, & Ones, 2001; Neisser et al., 1996; Nisbett et al., 2012; Ren, Schweizer, Wang, & Xu, 2015). Although the relation between intelligence and academic performance is well established, there are various theories in the literature with regarding how intelligence should be conceptualized and operationalized (Carroll, 1993; Gardner, 1983; Horn & Cattell, 1966; Jensen, 1998; Sternberg, 1988). One of the most common ways that intelligence has been conceptualized is along the dimension of crystallized versus fluid intelligence (Carroll, 1993; Horn, 1968; Horn & Cattell, 1966).

Cattell (1941, 1943) posited that intelligence is made up of two broad but distinct types, crystallized intelligence (Gc) and fluid intelligence (Gf). *Crystallized intelligence* (Gc) is associated with learned or acculturated knowledge and skills, and it involves the ability to put previously learned conceptual or procedural knowledge to use. Crystallized knowledge can be domain-general, domain-specific, or cross-domain, and is a result of learning and understanding acquired over one‘s lifetime, often as a result of formal education (McGrew, 2009; Shipstead, Harrison, & Engle, 2016; Thorsen, Gustafsson, & Cliffordson, 2014). For example, students’ ability to solve calculus problems, formulate an argument, or analyze a chemical compound are abilities that are not typically derived simply as a consequence of living in the world. Rather, these abilities are acquired through more formal channels (e.g., schooling, courses, tutoring). For the purpose of this dissertation study, the degree to which students’ crystallized knowledge at the time they enter high school, and how it may predict their later cross-domain and domain-specific academic performance, is of particular interest.

In contrast to crystallized intelligence, *fluid intelligence* (Gf), the ability to reason about novel or unschooled problems (Cattell, 1987), includes foundational cognitive abilities that are not dependent on formal training. Such abilities are assessed with unfamiliar and untrained tasks that do not rely on previously acquired conceptual or procedural knowledge to arrive at a solution. For instance, individuals' ability to recognize gaps in logic, identify weak evidence, or draw inferences can be cases of fluid intelligence. There are also individuals who are able to recognize and reason about underlying similarities between objects, ideas, experiences, or situation that bear no surface correspondence, or who can recognize discrepancies or even direct conflicts within information (Alexander & Baggetta, 2014; Sternberg, 1988). While such relational abilities are rarely schooled, they are considered to be key to performance in many academic domains (Alexander, 2016; McGrew, 2009; Schroeder, Schipolowski, Zettler, Golle, & Wilhelm, 2016).

These fluid abilities to perceive meaningful associations can be initiated automatically, without any intentionality on students' part, as when students suddenly notice that some event taught in history last year was the backdrop for a fictional story read in literature class. Alexander and Baggetta (2014) referred to such spontaneous perception as *relational thinking*. At other times, individuals utilize this process of discerning patterns as a way to purposefully or intentionally forge meaningful associations where they do not appear to exist. This process then becomes strategic, the evocation of a metastrategy involved in higher-order thinking and problem solving processes; what has been labeled *relational reasoning* (Alexander & the Disciplined Reading and Learning Research Laboratory [DRLRL], 2012a; Bassok, Dunbar, & Holyoak, 2012).

As a metastrategy, relational reasoning is broadly applicable to a range of academic domains and problem-solving tasks—from medical diagnosis to engineering design (Dumas, 2017; Dumas, Alexander, Baker, Jablansky, & Dunbar, 2014). It promotes cross-task and cross-domain thinking, thereby helping to promote the transfer of knowledge from one specific situation or context to another (Bereby-Meyer & Kaplan, 2005). Further, since relational reasoning requires consideration of the attributes or features of information in its many forms, it has the potential to heighten individuals' attentional and perceptive processes (Dunbar & Fugelsang, 2005).

By its very nature, relational reasoning requires individuals to engage in a deeper processing of information than would otherwise occur (Stephens, 2006). It serves as a counter to learners' tendency to treat information in a piecemeal or isolated manner, and thereby contributes to the likelihood of processing information accurately and developing more principled knowledge (van Gog, Paas, & van Merriënboer, 2004). Relational reasoning allows one to find correspondence between and across concepts and knowledge representations, and to integrate multiple mental relations to arrive at a logical solution (Krawczyk, 2012; Vendetti, Johnson, Lemos, & Bunge, 2015).

While there may be various relational reasoning forms, there are four in particular that have recently garnered attention in the literature and that are regarded as fundamental for forging associations between and among pieces of information. These relational reasoning forms are: analogy, anomaly, antinomy, and antithesis (Alexander and the DRLRL, 2012b). These four relational reasoning forms are fundamental to conceptual restructuring, and the development of principled knowledge in a number of domains (Alexander, 2016). The basis for associations in *analogical reasoning* is attributional *similarity* between two seemingly unrelated ideas, objects, representations or situations (Bassock et al., 2012; Gentner & Landers, 1985; Holyoak, 2012; Nersessian & Chandrasekharan, 2009; White & Alexander, 1986). *Analogical reasoning* requires one to map between a familiar or base representation and to a novel or target representation (Gentner, 1988; Goswami & Brown, 1990). Of the four relational forms, analogical reasoning is the most empirically studied (Cosgrove, 1995; Gentner & Markman, 1997; Glaser, 1984; Hong & Liu, 2003; Novick, 1988).

Anomalous reasoning involves the recognition of *discrepancies* or *deviations* in one idea, object, representation or situation from an established rule or trend in another (Chinn & Brewer, 1993; Singer & Gagnon, 1999). The ability to detect and analyze dissimilarities or aberrations in typical patterns is also invaluable for concept formation. (Chinn & Brewer, 1993; Chinn & Malhotra, 2002). For example, a doctoral student examining his dissertation dataset may recognize that test scores greater than three standards deviations are outliers and, consider whether those outliers should be included in analysis or not.

Antinomous reasoning, in contrast, deals with incompatibility and often involves categorization on the basis of mutual exclusivity (Dumas, Alexander, & Grossnickle, 2013; Slotta & Chi, 2006). This entails the recognition of two mutually exclusive ideas, objects, representations, or situations. Unlike analogical thinking where the intention is to identify similarities, anomalous reasoning is activated when individuals are required to make definitive categorical judgments about whether ideas, objects, representations, or situations are *in* or *out* as set members. They cannot be both. In effect, the goal in antinomous reasoning is to understand what something *is* by ascertaining what it *is not*. For example, a physical education undergraduate student after learning about various invasion team sports games such as basketball and football, might notice that baseball does not classify as an invasion game even though it is a team sport.

Antithetical reasoning requires the recognition that ideas, objects, representations, or situations are set in *direct contrast or opposition* (Dumas et al., 2013; Nussbaum & Kardash, 2005). Antithetical reasoning requires the reversal of salient relations to form a truly oppositional pattern. However, antithetical positions are not simply two categorical polar opposites but are rather relative points of contrast that have virtually infinite points in between those two opposing points (Alexander, 2016). Antithetical reasoning is considered to be central for argument and persuasion (Chinn & Anderson 1998; Kuhn & Udell 2007). For example, students in a Social Studies World Cultures course during a discussion of ancient Greece, might come to understand that Athens and Sparta were two very different types of government even though both were in ancient Greece. Spartan government was based on a monarchy while Athenian government was a democracy.

According to Grossnickle, Dumas, Alexander, and Baggetta (2016), these four forms of relational reasoning share the underlying component processes of encoding, inferring, mapping, or applying (Sternberg, 1977, 2009). However, it is largely the difference in the mapped associations that distinguishes one form of relational reasoning from another. That mapped relation could be one of similarity (analogy), discrepancy (anomaly), incompatibility (antinomy), or opposition (antithesis).

Together, these forms constitute critical components of higher-order cognition, because they are derived from a series of simple associations among individual pieces of information that then combine to patterns of relations (Chi, 2013; Goswami, 1992). The forms of relational reasoning have also been linked to success in a range of domains including science, medicine, engineering, and nursing (Dumas 2017; Fountain, 2017).

Statement of Problem

Theoretical Rationale

The importance of crystallized knowledge and fluid reasoning to learning and students' academic success has been well documented in theory and empirical research (Bolger, Mackey, Wang, & Grigorenko, 2014; Holyoak, 2012; Richland, Morrison, & Holyoak, 2006; Shipstead et al., 2016; Thibaut & French, 2016). However, there have been conflicting perspectives on how much each ability specifically influences and contributes to academic performance (Ackerman, 1996; Cattel 1987; Hambrick & Engle, 2002). Cattell's Investment Theory (1971, 1987) placed greater emphasis on the role of fluid abilities, and proposed that individuals have a fixed amount of Gf which they can choose to invest in, or apply to, learning in specific crystallized skills or domains. According to this theory, fluid abilities play a crucial role in the development of crystallized knowledge. Greater Gf abilities are thought to result in a faster and broader accumulation of Gc (Cattell, 1971; Ferrer, McArdle, Shaywitz, Holahan, Marchione, & Shaywitz, 2007).

Ackerman (1996) built upon Cattell's Investment Theory integrating intelligence-as-process, personality, interests, and intelligence-as-knowledge into his

PPIK Model. Ackerman's intelligence-as-process corresponds to Cattel's Gf, and these processes included reasoning, memory, perceptual speed, and spatial rotation. Ackerman's intelligence-as-knowledge matched the description of Gc provided by Cattell in his Investment Theory (Ackerman, 1996). Similar to Cattell's Investment Theory, Ackerman suggested that intelligence-as-process has a positive influence on intelligence-as-knowledge such that more knowledge will be gathered if a person has higher fluid intelligence. Some empirical evidence has supported this relation between Gf and Gc (Ackerman, Bowen, Beier, & Kanfer, 2001; Büehner, Krumm, Ziegler, & Plüecken, 2006; Rolfhus & Ackerman, 1999).

In contrast to Cattell's Investment Theory and Ackerman's PPIK model, several models propose that the best predictor of future performance is crystallized abilities. The knowledge-is-power hypothesis (Hambrick, 2005; Hambrick & Engle, 2002) argues that within a given domain, prior knowledge is a stronger predictor of future performance than reasoning. According to the knowledge-is-power hypothesis, prior knowledge facilitates future knowledge as well as superior performance in other cognitive tasks (Hambrick, 2005; Hambrick & Engle, 2002).

The expert performance model (Ericsson, Krampe, & Tesch-Römer, 1993) suggests that an individual's current knowledge base allows new knowledge to be processed faster and integrated deeper as evidenced by the classic chess masters studies (Charness, 1987; Chase & Ericsson, 1981; Chase & Simon, 1973). The idea is that the greater the crystallized or domain-specific knowledge learners bring with them to the learning environment, the greater their ability to notice relevant cues and informational sources (McPherson & MacMahon, 2008; Williams & Ericsson, 2005).

This allows more knowledgeable and skilled students to construct meaningful understanding of new concepts that are introduced in classrooms more effectively than their less knowledgeable or less skilled peers (Alexander, 2006; Vosniadou, 1999).

In addition, domain-specific knowledge may help learners to overcome the limitations of their fluid reasoning abilities. Often, learners who are in an acclimated or naïve state of learning are overloaded when encountering new information because they are unable to prioritize what they should respond to due to a limited domain-specific knowledge base (Alexander, Jetton, & Kulikowich, 1995; Chase & Simon, 1973; Endsley, 2006). Without the knowledge foundation, learners' lack the ability to determine information that may be relevant and important, and therefore often fail to notice patterns within that information. Overall, crystallized knowledge is thought to facilitate the learners' fluid reasoning abilities in higher-level cognition and academic performance (Alexander, 1997; Bolger et al., 2014; Murphy & Alexander, 2002; Orzechowski, 2010).

Both Gc and Gf have been theorized to contribute to academic performance. However, there is a lack of consensus among scholars with regard to how, and to what extent or degree, each individual construct influences academic performance (Carroll, 1993; Hunt, 2000; Rindermann, Flores-Mendoza, & Mansur-Alves, 2010). Given this, it would be valuable to examine the extent of the role both fluid and crystallized cognitive abilities may play in predicting of future academic performance.

Empirical Rationale

Recently, there has been a paradigm shift in educational curriculum and pedagogy, and in particular science, technology, engineering, and mathematics (STEM) education. That shift was from the acquisition and rote recall of facts and concepts to analyzing and interpreting evidence by recognizing and discerning patterns to understand similarities and differences among concepts (National Assessment of Educational Progress, 2009). This change in focus has led to a renewed interest in relational reasoning (Alexander, 2017; Murphy, Firetto, & Greene, 2016). However, until recently, the multidimensional nature of relational reasoning has not been widely acknowledged or extensively investigated. Historically, research on relational reasoning has focused on only one form of reasoning, analogies, especially within the cognitive neuroscience literature (Baggetta & Alexander, 2016; Dumas et al., 2013).

In addition, due to the operationalization of relational reasoning solely by analogical thinking, the capture of relational reasoning has relied exclusively on items from the Raven's Progressive Matrices (Dumas et al., 2013). The Raven's Progressive Matrices (RPM; Raven, 1941) is a measure of analogical reasoning consisting of 36 individual items presented in ascending order of difficulty (i.e., the easiest item is presented first, and the hardest item is presented last). The use of the RPM as a measure of relational reasoning restricts measurement of the multidimensional construct of relational reasoning to just one particular manifestation (i.e., analogy).

Alexander and colleagues have developed an alternate measure for relational reasoning, the Test of Relational Reasoning (TORR; Alexander & the DRLRL, 2012b). The TORR measures participants' ability to discern meaningful patterns across the four forms of relations (i.e., analogy, anomaly, antinomy, and antithesis), and is designed to limit the influence of prior knowledge and language ability through the use of graphical, non-linguistic items. The TORR has been shown to be more demanding and extensive measure of relational reasoning than the RPM (Grossnickle et al., 2016; Jablansky, Alexander, Dumas, & Compton, 2015). Given the potential influence of fluid relational reasoning on the students' academic performance, it would be useful to examine the multidimensionality of relational reasoning using the TORR.

Practical Bases for Investigation

My position at the school attended by the study participants helped shape the goals of this study. I am the Director of the Center of Academic Excellence (CAE), and am the overall coordinator of supplemental academic services and testing for students at the school. In these positions, I advise the Admissions Committee regarding applicants to the school, and their likelihood of success. I also provide inputs to the Academic Council as it pertains to the school's academic curriculum. Furthermore, part of my responsibility as Director is to design, develop, and implement programs that provide high quality academic scaffolding for students at the school, especially those students considered to be "at need" for additional academic support. Thus, there are pragmatic reasons for this investigation and an opportunity to put the findings into practice.

The first practical rationale for this study was to investigate the various ways the High School Placement Test (HSPT) is currently being used by the school. The HSPT assesses incoming ninth graders' academic readiness levels by measuring basic cross-domain-knowledge and skills in reading, mathematics, and language. The school uses the HSPT for a number of key academic decisions, including admissions selections, course placements, and identifying students as academically "at need." Given the importance of these academic decisions, it would be beneficial to examine how well the HSPT predicts academic performance.

Moreover, as someone involved in the conceptualization and operationalization of the TORR (Alexander & the DRLRL, 2012), I am interested in the degree to which this measure may uncover underlying cognitive abilities that play a role in the participating students' cross-domain and domain achievement at the ninth and twelfth grade level.

The Current Study

The overall purpose of this study was to investigate how measures of both crystallized cross-domain knowledge and fluid relational reasoning abilities contribute to high-school students' scholastic aptitude and content-specific achievement. The conceptual model guiding this investigation is shown in Figure 1. As this model indicates, this study examines the influence of three crystallized cross-domain knowledge measures (i.e., Language Skills, Mathematics, and Reading Comprehension), and fluid relational reasoning ability (Test of Relational Reasoning) on ninth-grade students' scholastic aptitude (Preliminary Scholastic Assessment Test), and content-specific achievement (Social Studies, Algebra 1, and Spanish 1 examinations); and on twelfth-grade students' content-specific achievement (Advanced Placement Psychology examination).

Research Questions

To meet the aforementioned purposes, the following research questions were addressed:

RQ1: What are the differences between ninth and twelfth-grade students on measures of crystallized cross-domain knowledge and fluid relational reasoning?

A number of studies have examined differences in analogical reasoning between age groups, and in particular comparing younger children with adolescents and adults (Richland et al., 2006; Thibaut, French, & Vezneva, 2010). However, few studies have examined differences in relational reasoning between early high school years and later high school years even though studies have shown that the reasoning networks continue to be refined during adolescence leading to improved integration of relations (Magis-Weinberg, Blakemore, & Dumontheil, 2017; Singley & Bunge, 2014; Vendetti, Matlen, Richland, & Bunge, 2015). In addition, only one study to date has examined the differences in multiple forms of relational reasoning among differing age groups. Jablansky and colleagues (2015) investigated the ways in which four forms of relational reasoning manifested itself in discourse between younger and older children and found differences in the ways in which the four forms of reasoning manifested between age groups. This study analyzed both performance on the HSPT and the TORR, as well as performance for each of the four forms of the TORR to investigate any possible patterns of differences between ninth- and twelfth-grade students.

RQ2: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on ninth-grade students' scholastic aptitude and content-specific achievement?

RQ3: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on twelfth-grade students' content-specific achievement?

This study investigated if crystallized cross-domain knowledge or fluid relational reasoning abilities, better predicts academic readiness and achievement. Research has found that both fluid abilities and crystallized knowledge positively predicts learning and academic performance. Some studies have found crystallized knowledge to be a stronger predictor of academic performance (McGrew &Wendling, 2010; Postlethwaite, 2012). However, research also suggests that fluid abilities have greater influence upon learning in mathematics and the sciences, whereas crystallized knowledge have greater influence upon knowledge acquisition in history, humanities, and civics (Ackerman, 2000; Evans, Floyd, McGrew, & Leforgee, 2002). In addition, McGrew (1997) found that crystallized knowledge and the fluid visual/spatial abilities were the most frequently represented abilities in major cognitive ability tests, while measures of relational reasoning were less common, and multi-dimensional measures of relational reasoning were non-existent. Only recently has research begun examining relational reasoning as a measure of fluid ability, and in particular multiple forms of relational reasoning in addition to analogical reasoning, to different aspects of human performance (Dumas & Schmidt, 2015; Jablansky et al., 2015; Murphy, Firetto, & Greene, 2017). Findings support the proposition that the ability to recognize meaningful patterns within and among a varying number of stimuli, is strongly associated with successful performance. However, the research on the degree to which relational reasoning predicts academic performance among adolescent students has not yet been well investigated.

Given that cognitive tests are used in a number of high-stakes admissions and educational placement decisions, it is valuable to assess the criterion-related validity of these tests. Therefore, this dissertation examined both crystallized cross-domain knowledge and fluid relational reasoning abilities in order to further understand which cognitive abilities most influence a high school student's scholastic aptitude and content-specific achievement. The High School Placement Test (HSPT) provided data on the students' crystallized cross-domain knowledge. Relational reasoning was measured with the Test of Relational Reasoning (TORR, Alexander & the DRLRL, 2012b). Both scholastic aptitude and content-specific achievement were measured. Scholastic aptitude was measured using the Preliminary Scholastic Assessment Test (PSAT). The PSAT is a standardized test that consists of questions in the areas of reading, mathematics, and writing skills. The PSAT is recognized as a valid and reliable measure for academic skill level achievement and college readiness (Kim, Hendrickson, Patel, Melican & Sweeney, 2013; Proctor, Wyatt & Wiley, 2010). Content-specific achievement was measured using the students' Social Studies, Algebra 1, and Spanish 1 semester exams for ninth-graders and the AP Psychology exams for twelfth-graders.

In summary, by investigating the role of crystallized cross-domain knowledge and fluid relational reasoning in academic performance, this study makes an important contribution to our knowledge regarding crystallized knowledge, relational reasoning and academic performance of adolescents. Cognitive ability tests are one of the strongest predictors of academic performance (Kuncel et al., 2001), and historically, tests of crystallized knowledge have been the dominant type of cognitive ability tests used for student and academic selection purposes (Hunt, 2000). However, nonverbal fluid ability tests, such as the TORR, may reduce certain forms of bias that can mask the abilities of minority or underrepresented groups with less access to "schooled" content found in crystallized measures (Jensen, 1980; Kaya, Stough, & Juntune, 2016; Klein, 2004; Saccuto, Johnson, & Russell, 1992).

Key Terms

What follows is a glossary of terms that were central to this investigation.

Content-Specific Achievement is defined as the score on a formal assessment of a specific field of study (e.g., World Cultures, Algebra 1, and Spanish 1 final semester examination scores).

Cross-Domain Knowledge is defined as acquired knowledge and competences from different domains which are required for a specific task performance (e.g., reading competence; Mischo & Maaß, 2012).

Crystallized Knowledge is defined as knowledge and skills acquired through education and experience in academic domains (Cattell, 1987).

Fluid Ability is defined as the basic abilities in higher order mental processes (i.e., abstract thinking, problem solving, and relational reasoning) that are used intentionally and deliberately to perform new tasks (Carroll, 1993; McGrew, 2005; Schneider & McGrew, 2012).

Relational Reasoning is defined as "the ability to recognize or derive meaningful relations between and among pieces of information that would otherwise appear unrelated" (Alexander and the DRLR, 2012a).

Scholastic Aptitude is defined as a student's ability/potential to succeed in college (College Board, 2019).

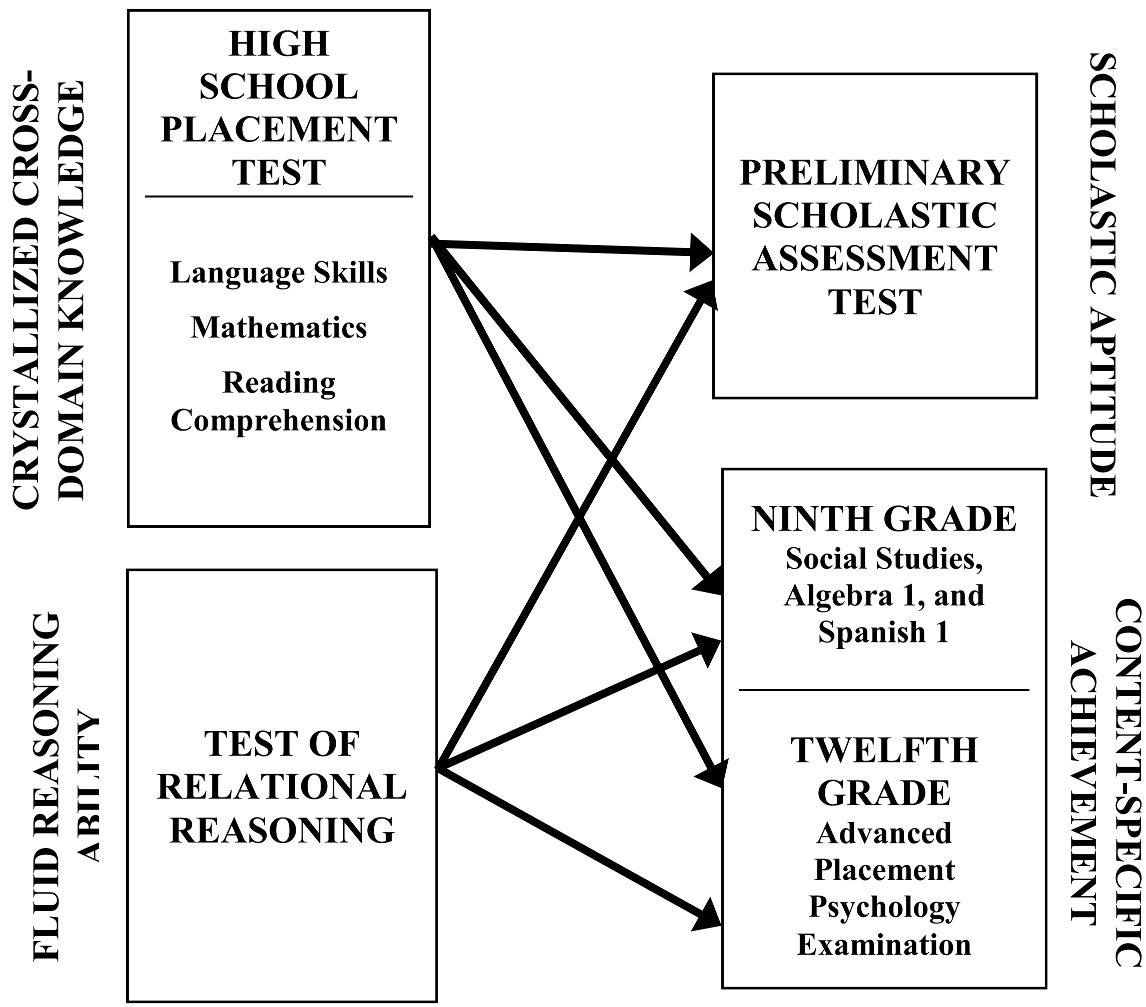


Figure 1. Conceptual model of the hypothesized relation between crystallized cross-domain knowledge and fluid reasoning ability and scholastic aptitude and content-specific achievement.

Chapter 2: Review of the Literature

The primary purpose of this chapter is to describe the relations among the constructs represented in the conceptual model displayed in Figure 1. Therefore, this literature review will address the two main components of that conceptual model: crystallized knowledge, and fluid relational reasoning. First, the area of intelligence will be surveyed. Given the breadth and depth of this literature, a short overview of the Gf (fluid intelligence) and Gc (crystallized intelligence) framework (Carroll 1993; Cattell 1987; Horn, 1968; Horn & Cattell, 1966; Schneider & McGrew, 2012) will be shared. In particular, the differing perspectives (Cattell, 1963; Hambrick & Engle, 2002) of the relative influences of Gf and Gc on learning and academic performance will be discussed.

Second, the review will examine the construct of relational reasoning, the ability to perceive or derive meaningful relations or patterns between and among pieces of information that would otherwise appear unrelated (Alexander & the DRLRL, 2012a). In much of the older literature on relational reasoning, it has often been treated as a unidimensional construct associated solely with analogical reasoning (Dumas et al., 2013). However, the contemporary study of relational reasoning is now questioning this narrow conception of relational reasoning. Consequently, if relational reasoning represents the ability to perceive or derive meaningful patterns from unconnected information, then there are possibly other distinct relations that could potentially exist in addition to analogy. These include anomaly, antinomy, and antithesis, with each representing a distinct rule-based form of reasoning (Alexander & the DRLRL, 2012a, Alexander & the DRLRL, 2012b; Kendeou, Butterfuss, Van

Boekel, & O'Brien, 2017; Knauff, 2009; Goel, 2007). Given this emerging conceptualization and operationalization, theory and empirical research related to the multidimensional construct of relational reasoning will be reviewed.

Influence of Crystallized Knowledge and Fluid Reasoning on Academic Performance

The importance of knowledge has been frequently identified as a critical component of intelligence (Ackerman, 1996), and is commonly referred to as crystallized intelligence (Carroll, 1993; Horn & Noll, 1997; McGrew, 2009). Crystallized intelligence (Gc) is a key construct in major theories of intelligence including the Gf-Gc theory (e.g., Cattell, 1943; Horn & Noll, 1997), Carroll's (1993) Three-stratum theory, and the Cattell-Horn-Carroll (CHC) theory of intelligence (e.g., Schneider & McGrew, 2012). Crystallized intelligence or knowledge is made up of both the general academic skills (i.e. basic reading, language, vocabulary, and mathematics skills) and domain-specific knowledge that are primarily acquired during schooling. These include an individual's acquired store of declarative (knowing what) and procedural (knowing how) reading, writing, numerical and mathematical knowledge and skills, as well as the breadth, depth and mastery of an individual's acquired knowledge in domain-specific subjects or disciplines developed through investment in practice, time, and effort (Cattell, 1987; McGrew, 2009).

Crystallized intelligence (Gc) is one of two components of Cattell's (1971;1987) Gf-Gc Investment Theory, with the other component being fluid intelligence (Gf). Fluid intelligence reflects basic abilities in reasoning and other related higher order mental processes (i.e. abstract thinking, and problem solving) that are used intentionally and deliberately to perform new tasks (Carroll, 1993; McGrew, 2005). This includes the ability to draw an inference or conclusion, understand and connect concepts, identify relations, develop and test hypothesis, and engage in complex reasoning and problem-solving (Schneider & McGrew, 2012). These abilities are considered to be independent of learning, experience, and education and are key to new knowledge acquisition and retention. The development of fluid intelligence is largely dependent on biological and genetic factors – the maturation of the frontal-parietal regions and networks and increasing myelinization of axons (Barbey, Colom, Solomon, Krueger, Forbes, & Grafman, 2012; Blakemore & Choudhury, 2006; Gray, Chabris, & Braver, 2003; Krawczyk, 2012).

Crystallized knowledge is thought to reflect the knowledge an individual has been able to acquire and learn with the assistance of fluid intelligence (Carroll, 1993) through education, both formally and informally. Gf and Gc have been found to be related, with correlations being reported between .40 and .60 (Horn & Noll, 1997; McGrew, 1997). Gf is typically assessed with tests of reasoning and Gc with tests of vocabulary, comprehension, and general information. Gc has also been observed to correlate especially highly with domain-specific knowledge (Beier & Ackerman, 2003; Hambrick, Pink, Meinz, Pettibone, & Oswald, 2008).

The core premise underlying Gf–Gc theory is that Gc develops out of Gf, and that Gf should be reflected in tasks reflecting knowledge. Other non-cognitive factors such as personality and interest, are also thought to impact the development of Gc, although to a lesser extent than fluid ability. According to the Cattell Investment Theory (Cattell 1971; 1987), the investment of fluid abilities into crystallized knowledge occurs during the elementary schooling years when the needed reading, writing, and arithmetic skills are acquired and developed. These skills are needed to successfully perform academic tasks, and thereby lead to successful academic performance as students progresses through school, especially in subjects that demand understanding and reasoning about complex relations, such as in literature, mathematics, and science.

Cattell (1987) stated that the coupling between Gf and Gc should be especially strong early in elementary school and decrease during late childhood into adolescence as the level of Gc is partially a function of Gf and of previous levels of Gc. Taub, Floyd, Keith, and McGrew (2008) tested this theory by examining the influences of both crystallized intelligence and fluid reasoning on mathematics achievement. Participants were a subsample of a nationally representative sample stratified according to race, sex, geographic region, education, and age that mirrored the population characteristics of the United States Census projections for the year 2000. The researchers created four age-differentiated samples (ages 5 to 6, n = 639; 7 to 8, n = 720; 9 to 13, n = 1,995; and 14 to 19, n = 1,615) and used 18 tests from the Woodcock-Johnson III Tests of Cognitive Abilities and Achievement. Fluid reasoning displayed strong direct effects on mathematics achievement across all four age levels, and crystallized intelligence showed moderate effects for ages 9 to 13 and strong effects for ages 14 to 19. These results suggest that both fluid and crystallized abilities are needed for proficiency in mathematics and that better developed fluid leads to increased performance in mathematics.

Thorsen and colleagues (2014) also found support for Cattell's Investment Theory. They investigated the influence of Gf and Gc on the development of knowledge and skills as measured by achievement tests and subject grades in a homogenous sample (i.e., same levels of education, age, and cultural background) of children. In a large-scale longitudinal study, 9,002 individuals were tested at Grades 3, 6, and 9. A large direct effect of Gf was found in Grades 6 and 9 with the influence of Gf increasing at each point. The researchers hypothesized that this might be due to the possibility of tasks increasing in complexity as students' progress from elementary through high school, which demands an increasing investment of Gf. The researchers also found a slight increase in the influence of Gc over time. The results from the study suggests that Gf influences the development of knowledge and skills throughout school, supporting Cattell's Investment Theory.

Rindermann, Flores-Mendoza, and Mansur-Alves (2010) set out to directly test Cattell's Investment Theory by investigating whether fluid intelligence had a larger impact on crystallized intelligence, or if crystallized intelligence had a larger impact on fluid intelligence. They compared two differing ability groups, high ability (n = 722, ages 11 to 20, mean IQ = 118) and low ability (n = 833, ages 7 to 15, mean IQ = 89). They found that fluid intelligence had no larger impact on crystallized intelligence ($\beta fl1 \rightarrow cr2 = .16$), and that crystallized intelligence had a slightly larger effect on fluid intelligence ($\beta cr1 \rightarrow fl2 = .20$). The findings did not support the Cattell's Investment Theory, which states that fluid intelligence is the basis for the development of crystallized intelligence.

In contrast to the Cattell Investment Theory, research in cognitive psychology has put forward a differing perspective known as the *knowledge-is-power* hypothesis. The basic idea of the knowledge-is-power hypothesis is that domain-specific prior knowledge facilitates future knowledge and performance above and over domain-general cognitive abilities (Hambrick & Engle, 2002). According to the knowledge-is-power hypothesis, domain specific knowledge seems to be more important for the prediction of cognitive achievement compared to domain-general cognitive abilities, such as reasoning. The more one knows about a specific domain, the more accurate retrieval, deeper integration, faster processing of new knowledge, and thereby making it easier to acquire and integrate new knowledge into the framework of prior knowledge. The knowledge-is-power hypothesis has been supported by numerous studies that have found domain-specific knowledge to contribute to success in various areas such as academic performance (Alexander & Judy, 1988), chess (Chase & Simon, 1973), medical vocational education and training (Moehring, Schroeders, & Wilhelm, 2018), and problem-solving in physics (Chi, Glaser, & Rees, 1981).

Several studies have directly examined the knowledge-is-power hypothesis (Hambrick, 2003; Hambrick & Engle, 2002; Moehring et al., 2018). Hambrick (2003) found that prior knowledge exerted a stronger influence on knowledge acquisition in the field of basketball. Hambrick investigated the role of Gf and Gc on the acquisition of new domain knowledge under naturalistic learning conditions. In the first stage of the study, 171 college aged participants completed tests of reasoning ability, working memory, and tests of prior knowledge of basketball. Approximately two and half months later, participants completed tests to assess new knowledge of basketball that may have been acquired since the initial testing. Results revealed a direct effect of interest and prior knowledge on new basketball knowledge, while there was only a minor effect of Gf and working memory on the acquisition of new basketball knowledge. The results suggest that individual differences in newly acquired domain knowledge arise from a process involving interest and prior knowledge - as a learner's interest in a domain increases, his or her knowledge of that domain also increases, and at the same time, as a person's knowledge of a domain increases, his or her interest in that domain increases.

In another study, Hambrick and Engle (2002) investigated the role of both domain-specific knowledge and working memory on memory performance. Participants listened to, and then answered questions about, simulated radio broadcasts of baseball games. Results found that domain knowledge (i.e., prior baseball knowledge) was the strongest predictor of individual differences in the performance of a domain-specific cognitive task (i.e., 54.9% of the variance). In summary, the findings of the multiplicative effect of domain knowledge on performance suggest that integrating new information into preexisting knowledge structures should improve performance in that domain.

Moehring and colleagues (2018) tested the knowledge-is-power hypothesis in a real life vocational education and training setting by investigating the acquisition of vocational knowledge for medical assistant students (n = 448) over three years. The researchers tested domain-specific vocational knowledge (i.e., medical knowledge, laboratory knowledge, and medical workplace knowledge), Gf (figural Gf scale where a sequence of geometric figures was presented and participants had to identify which were the next two figures in the sequence out of three alternatives for each missing figure), and Gc (natural sciences knowledge, general humanities knowledge, and general social studies knowledge). Gc emerged as the strongest predictor of vocational knowledge in all 3 years, while, Gf showed no impact on vocational knowledge for the first two years of education, and only a small influence in the last year of education. The findings provide support for the idea that a comprehensive knowledge base is a key prerequisite for knowledge acquisition and support the knowledge-is-power hypothesis.

A meta-analysis of over 400 studies conducted by Postlethwaite (2011) did not directly investigate the knowledge-is-power hypothesis, but did explore the relation between fluid ability, crystallized ability, and performance across multiple domains and age levels. The overall findings showed that measures of fluid ability were found to positively predict learning and academic performance. However, crystallized ability measures were found to be superior predictors of academic performance compared to fluid ability. Specifically, the overall observed validity of Gf for predicting academic performance was .40 (*SD* = .14), with the observed validity for Gf was .38 (*SD* = .18) for high school samples. The overall observed validity of crystallized ability for predicting academic performance was .65 (*SD* = .09), with the observed validity of Gc for predicting academic performance for high school samples was .53 (*SD* = .17). The results of the study, which were based upon criterion-related validity rather than factor-analytic evidence, demonstrated that Gc measures are stronger predictors of academic performance than Gf measures across all age groups, including high school students.

In sum, there is a well acknowledged relation between cognitive abilities, knowledge, and their influence on learning and performance among the differing perspectives (i.e., Cattell Investment Model and knowledge-is-power hypothesis). What remains debated are the dynamics of this relation, as well as their makeup, interplay, and directionality. The inconsistent findings reported in the literature regarding the influence of Gf and Gc on learning and knowledge acquisition may be due to age-related changes. In a comprehensive review, Baumert, Lüdtke, Trautwein, and Brunner (2009) found that domain-specific knowledge contributes to academic performance over and above fluid intelligence and that this contribution becomes more important as students' progress from elementary through high school. The review (Baumert et al., 2009) showed an increasing influence of Gc and a decreasing influence of Gf on academic achievement from the end of elementary education to high school which might suggest that as students move from general academic skills to more domain-specific learning, that Gc plays a greater role. That is, students with higher levels of Gc within a specific domain may be able to acquire new knowledge within that domain more quickly and effectively. Schroeders, Schipolowski, and Wilhelm (2015) also found empirical support that showed crystallized abilities made stronger age-related gains during high school, while fluid abilities did not change during the course of high school.

A second possible reason why there are differences in research findings regarding the influence of Gf and Gc on learning and knowledge acquisition may be a result of the domain or context being investigated. Increasing specialization in academic domains starts to begin during high school and may cause a differentiation of needed knowledge structures. Gf has been shown to have a greater influence on learning in the physical sciences, such as physics and chemistry, whereas Gc had a greater influence for knowledge acquisition in arts, humanities, civics, and reading (Ackerman, 2000; Evans et al., 2002). For example, Evans and colleagues (2002) found that Gc was the strongest predictor of basic reading skills and comprehension in a study of reading achievement during childhood and adolescence. Finally, Gf did not appear to have any significant relation to reading achievement.

In a large-scale study, Gustafsson (2001) investigated the effects of schooling on fluid and crystallized intelligence on adolescent males (n = 14,000) in Sweden. He compared the different secondary school tracks of the participants against a Military Enlistment Battery that included a vocabulary assessment (crystallized ability) and figural series tasks (fluid reasoning ability) while controlling for initial differences in grades and socio-economic status. The results showed the strongest effect for the Gf, while the effects of Gc were weaker, indicating that participation in an academic track at secondary school had a positive effect on Gf and, to a lesser extent, on Gc. The Gf factor was measured by tests of reasoning that involved novel problem-solving, suggesting that the positive effect may be a result of the academic track placing a greater emphasis on problem solving rather than on knowledge acquisition. In another large-scale study (n = 6,701secondary school students), Schipolowski, Wilhelm, and Schroeders (2015) investigated the relation of crystallized abilities (i.e., verbal ability and declarative knowledge) and fluid intelligence to school achievement in Germany. Verbal ability had the highest correlations to grades in language courses (i.e., grades in German and English), whereas fluid intelligence had the strongest correlations to grades in mathematics.

These different research findings suggest that Gf and Gc operates differently depending on both how Gc is being conceptualized and the academic performance being assessed. If Gc is conceptualized as a general language knowledge such as vocabulary or verbal ability (Carroll, 1993), and the academic performance involves reading or language, then it seems like crystallized abilities would have a stronger influence than fluid abilities. However, if the academic performance is in mathematics or science where pattern-recognition, identifying relations, and problem-solving are required, then Gf seems to play a greater role. Additionally, if Gc is conceptualized as declarative or domain-specific knowledge and the performance is domain-specific, then Gc would have a much stronger influence than Gf. It appears that Gc is especially important for the acquisition of domain-specific knowledge (Ackerman & Beier, 2006; Beier & Ackerman, 2005; Hambirck et al., 2008).

Relational Reasoning

The ability to explicitly identify and recognize relations is a higher-order executive function known as relational reasoning - the ability to intentionally discern meaningful patterns within any stream of information, be it linguistic, graphic, numeric, or even situational in nature (Alexander & the DRLRL, 2012b). Relational reasoning is thought to be an important learning metastrategy due to its broad application to a range of academic domains and intentional, effortful, and goal-directed processes (Alexander & Baggetta, 2014; Dumas et al., 2014; Krawczyk, 2012; Richland & Simms, 2015). Relational reasoning ability has been linked to academic success in the domains of astronomy, biology, physics, chemistry, meteorology, mathematics, reading, and writing (Dumas et al., 2013).

This mapping of patterns between and among pieces of information is a general multi-dimensional cognitive ability. As such, relational reasoning has long been recognized as fundamental to human learning and development as well as a hallmark of expertise in all domains (Goldman & Pellegrino, 2015; Vosniadou, 2002) as it contributes to a deeper processing of information, underlies the transfer of learning from one domain to another (Bereby-Meyer & Kaplan, 2005; Stephens, 2006), and is essential for successful performance on complex cognitive tasks (Bransford, Brown, & Cocking, 1999; Singer & Gagnon, 1999). Specifically, relational reasoning has been linked to deeper and successful processing of text (Andiliou, Ramsay, Murphy, & Fast, 2012), including refutational text (Hynd, 2003; Tippett, 2010); discernment of cause and effect relations (Fugelsang & Dunbar, 2005); avoidance of science misconceptions (Chi, Slotta, & De Leeuw,1994); generation of scientific discoveries (Chi & Hausmann, 2003); and identification of anomalies within scientific data (Chinn & Malhotra, 2002).

Multidimensionality of Relational Reasoning

Historically, relational reasoning has often been described as a unitary ability, usually centered on analogical reasoning (e.g., Goswami 1992; Holyoak 2012; Krawczyk, 2012). However, Alexander and the DRLRL (2012a) have put forward the theory that relational reasoning is a multidimensional construct and consists of at least four forms or manifestations that are fundamental for forging associations between and among pieces of information: analogy, anomaly, antinomy, and antithesis.

Analogy is the identification of a structural similarity between two or more concepts, objects, or situations (Bostrom 2008; Trey & Khan 2008; Wright, Matlen, Baym, Ferrer, & Bunge, 2008). Anomaly is the ability to discern an abnormality, digression, or deviation from an established pattern (Ferguson & Sanford 2008; Filik & Leuthold 2008; Schulz, Goodman, Tenenbaum, & Jenkins, 2008; Trickett, Trafton, & Schunn, 2009). Antinomy is the ability to recognize incompatibilities or paradoxes within an informational stream (Cole & Wertsch 1996; Gardner 1995; Mosenthal 1988; Sorensen 2003). Antithesis is the identification of a directly oppositional relation between two ideas, concepts, or objects (Baker, Friedman, & Leslie, 2010; Bianchi, Savardi, & Burro, 2011; Heit & Nicholson 2010).

Of the four relational reasoning strategies, analogies and analogical reasoning are historically the most empirically studied form of relational reasoning (Cosgrove, 1995; Gentner & Markman, 1997; Glaser, 1984; Hong & Liu, 2003; Novick, 1988). Constructing an analogy requires the learner to map between elements of the source domain and target domain by paying attention to the relevant cues by looking past the surface-level differences and identifying the underlying, shared relations between the two domains (Dinsmore, Baggetta, Doyle, & Loughlin, 2014; Gentner, 1983; Gick & Holyoak, 1980; Starr, Vendetti, & Bunge, 2018). The basis for association in analogical reasoning is a relational similarity between two seemingly disparate ideas, objects, representations or situations (Gentner & Landers, 1985; Holyoak, 2012; Bassock et al., 2012; Nersessian & Chandrasekharan, 2009). In analogical reasoning, there is the effort to construct meaningful associations via this relational similarity between objects or events that initially appear dissimilar. Analogical reasoning has been linked to academic achievement and success in a variety of domains - reading and language (Ehri, Satlow, & Gaskins, 2009; Goswami & Bryant 1992); mathematics (Richland & McDonough 2010); chemistry (Trey & Khan 2008); physics (Mason & Sorzio 1996); science (Lombardi, Nussbaum, & Sinatra (2016); and engineering (Chan & Schunn 2015; Christensen & Schunn 2007).

Anomalous reasoning involves the recognition of discrepancies or deviations in one idea, object, representation or situation from an established rule or trend in another (Chinn & Brewer, 1993; Singer & Gagnon, 1999). An anomaly can be defined as any occurrence or object that is strange, unusual, or unique; it is a discrepancy or deviation from an established rule or trend (Chinn & Brewer, 1993). The awareness of and response to anomalous data are also critical to pattern recognition (Chinn & Brewer, 1993; Chinn & Malhotra, 2002), and thus, to the development of principled knowledge in a number of domains. The ability to recognize discrepancies or irregularities has been linked to reading comprehension (Bohan & Sanford, 2008; Ivanova, Pickering, Branigan, McLean, & Costa, 2012) and to academic achievement particularly within the science domains (Chinn & Malhotra, 2002; Chinn & Samarapunga, 2009; Trickett et al., 2009).

Antinomous reasoning, entails the recognition that two ideas, objects, representations, or situations are incompatible (Berlin, 1990; Chi & Roscoe, 2002). Unlike analogical thinking where the intention is to forge some similarity or anomalous reasoning where the point is that some fact or observation appears different from others, the goal in antinomous reasoning is to understand what something is by ascertaining what it is not. Antinomy is a paradoxical, mutual incompatibility of two laws, rules, or principles (Gardner, 1995; Sorenson & Yankech, 2008). By extension, antinomy encompasses the type of mutual exclusivity involved in distinguishing different conceptual categories and the paradox that arises when they are brought together. Research on antinomous reasoning has not been very prevalent in educational psychology (Dumas, 2017) but can be found in fields of inquiry such as reading (Mosenthal, 1988); intelligence (Gardner, 1995); and categorical reasoning in science (Chi & Roscoe 2002; Chi & Slotta 1993).

Antithetical reasoning requires the recognition that two representations are set in direct contrast or opposition (Nussbaum & Kardash, 2005). While antithetical comparisons, like antinomies, involve conflicting information, the contrast is much sharper and entails apparent opposites viewed in a mutually exclusive, either/or relation. Specifically, an antithesis arises when two propositions, principles, or explanations are set in direct contrast or direct opposition. Antithetical reasoning has been identified within the literatures on persuasion or argumentation (Andiliou et al., 2012; Nussbaum & Kardash, 2005) and also to conceptual change in scientific domains through the use of refutation text (Broughton, Sinatra, & Reynolds, 2010); refutation graphics (Mason, Baldi, Danielson, & Sinatra, 2016); and plausibility judgments (Lombardi, Brandt, Bickel, & Burg, 2016).

Each of these four forms of relational reasoning proposed by Alexander and the DRLRL (2012b) share an overlapping series of sequential component processes that individuals complete in order to successfully construct a higher-order relation among pieces of information (Grossnickle et al., 2016; Sternberg, 1977). These component processes are (1) encode when the stimuli within the problem space is first recognized and identified (Carpenter, Just, & Shell, 1990; Hummel & Holyoak, 2005); (2) infer where lower-order associations among the pieces of information in the problem are formed (Gentner, 1983, 2010); (3) mapping where these lower-order relations are connected or mapped into a higher-order pattern or relation (Gick & Holyoak, 1980; Green, Fugelsang, & Dunbar, 2006); (4) and finally apply where after the mapping is complete, the discerned higher-order pattern can be applied to determine a response to a given task or problem (Sternberg, 1977).

It is the characterization of the mapped relation that distinguishes one form of relational reasoning from another. Specifically, although each form of relational reasoning requires the mapping of a higher order relation from multiple lower order relations (Goswami, 1992; Markman & Gentner, 2000; Sternberg, 1977; 2009), the mapped relation could be one of similarity (analogy), discrepancy (anomaly), incompatibility (antinomy), or opposition (antithesis). Alexander and the DRLRL (2012a) do not claim that that any of these forms of relational reasoning is the most important, nor do they claim that the four forms are the only types of relations that can be mapped but, rather contend that these four forms have broad applicability to educational settings, and often work together in unison.

Only a few studies have examined how analogies, anomalies, antinomies, and antitheses operate together during complex cognitive processes. A study by Dumas and colleagues (2014) explored how the four forms of relational reasoning (i.e., analogy, anomaly, antinomy, and antithesis) operate in the medical domain by investigating the decision making processes of a medical team within a teaching hospital. The researchers conducted a detailed analysis of the spontaneous discussions held between a team of medical residents and an attending physician on the diagnosis and treatment of actual patients in real time in a hospital setting. One attending clinical neurologist with over 20 years of experience diagnosing and treating neurological conditions, and his team of residents ($N = 9$) at a neurological institute and hospital in a large Canadian metropolitan university participated in the study. A total of 11 meetings regarding 35 different patient cases were recorded, transcribed, and coded which resulted a total of 2,114 conversation units coded over the 11 meetings. Of these, 272 conversational units were identified as instances of relational reasoning based on the presence of statements that suggested the speaker was relating two or more objects, ideas, or people. The conversation units identified as reasoning related were then coded according to the type of relational reasoning indicated. Of the identified instances of relational reasoning, 22.79% ($n = 62$) were analogical in nature, 58.72% ($n = 157$) dealt with anomalies, 17.27% ($n = 47$) with antinomies, and 2.20% ($n = 6$) with antitheses.

Findings of the study suggest that relational reasoning supports the thinking, learning, and communication of instructors and students in the medical field and potentially other domains involved in solving complex problem*s* in real-world setting*s*. Critical analysis of information appears to involve multiple forms of relational reasoning as results from the study revealed that all four forms of relational reasoning (i.e., analogy, anomaly, antinomy, and antithesis) were present in the meetings being examined. Additionally, it appears from the results that the forms of relational reasoning do not operate in isolation, but when confronted with a complex problem solving task, the multiple forms of relational reasoning work together in concert with one another.

Measurement of Relational Reasoning

The measurement of relational reasoning has been problematic as most measures have focused almost exclusively on only one form of relational reasoning – analogical reasoning. In a recent review of studies about relational reasoning, it was found that virtually no studies examined relational reasoning as a multidimensional construct (Dumas et al., 2013). One of the more common forms of measurement was the classical analogy verbal four-term format of A:B::C:D which often consisted of linguistic content and relied heavily on crystallized knowledge, specifically vocabulary (Wendelken, Nakhabako, Donohue, Carter, & Bunge, 2008), or a similar four-pictorial format (Krawczyk et al., 2008). Perhaps the most commonly used format in the measurement of analogical reasoning is the Raven's Progressive Matrices (RPM; Raven, 1941). The RPM has historically been used both as a measure of abstract reasoning and an estimate of fluid intelligence in general (Gonthier & Thomassin, 2015; Jastrzębski, Ciechanowska, & Chuderski, 2018; Raven, 1938; Raven, Raven, & Court, 1998; Snow, Kyllonen, & Marshalek, 1984).

The RPM is a visuospatial task assumed to measure non-verbal analogical reasoning and consists of items that include a three-by-three matrix of figural patterns missing the bottom-right pattern as well as eight response options with the patterns that can potentially match a missing one. The task is to choose the pattern that validly completes the matrix. Item difficulty progressively increases as the individual advances through the test.

The majority of the measures of relational reasoning found in the literature have focused on analogical reasoning, however, each of the other three forms have also been measured to some extent. Anomalies have been most commonly measured by the use of verbal semantic anomalies (Chinn & Brewer, 1993; Filik 2008; Weber & Lavric 2008). Verbal semantic anomalies are short passages or sentences that contain inconsistent or unusual information or describe something unexpected. The testing of antinomous reasoning has been limited to sorting and categorization tasks such as used by Shuwairi, Bainbridge, and Murphy (2014). Participants were asked to look at a set of images and sort them into the groups that seemed best or most natural to them. Antithesis has been measured using refutation texts to directly counter scientific misconceptions (Broughton & Sinatra 2010), as well as semi-structured interviews concerning the oppositional nature of everyday concepts (Fischer, Norberg, & Lundman, 2008). Verbal opposite problems have also been used to measure antithesis. Participants were asked to place objects on scales of polarity of height, size, width and length to see how the relation of opposition is used to organize language and thoughts (Bianchi, Savardi, & Burro, 2011). Overall, all of these various measurements have only examined relational reasoning as a unitary concept and not as a multidimensional construct.

In response to this lack of a measure to investigate the multidimensionality of relational reasoning, Alexander and the DRLRL (2012b) developed a non-verbal Test of Relational Reasoning (TORR) based on three guiding postulations (Dumas & Alexander, 2016). First, that the test captures the multidimensionality of relational reasoning, specifically the four forms of analogy, anomaly, antinomy, and antithesis.

The TORR has 32 visuospatial items, organized in four scales of eight items and corresponding to the four forms of relational reasoning, analogy, anomaly, antinomy, and antithesis. The scales are presented in a fixed order with the more familiar manifestation of relational reasoning, analogy, appearing first followed by anomaly, antinomy, and then antithesis.

Second, the test should not be unduly influenced by crystallized knowledge of any specific academic domain so that it is generic in content. Each of the visuospatial items were created using novel graphical arrays, as opposed to linguistic or pictorial symbols associated more with prior or crystallized knowledge. The items were designed so that all the necessary information to solve each given problem was contained within the problem space and thereby limiting the influence of prior knowledge or culturally relevant experiences on one's ability to solve the problems. Third, the test should assess fluid cognitive ability, in particular pattern recognition with new or novel material that reduces the influence of prior knowledge or experience. Therefore, relational reasoning, as operationalized by the TORR, is closely aligned with Cattell's (1987) concept of fluid intelligence, or Gf*,* as it measures respondents' ability to discern relational patterns within novel graphical arrays without the need for prior knowledge.

Dumas and Alexander (2016) investigated the usefulness of the TORR for measuring a multidimensional construct of relational reasoning. The TORR was calibrated and scored using multidimensional item response theory (MIRT) in a large, representative undergraduate sample (*n* = 1,379). It was found that the TORR is highly suited for assessing relational reasoning ability in university students, and in particular was effective at separating out low relational reasoning abilities from high relational reasoning abilities. The researchers calculated classical test theory (CTT) statistics for each item of the TORR and results were within the generally accepted ideal range for producing maximally variable and reliable scores among participants. Additionally, TORR scores exhibited strong reliability (Cronbach's alpha = .84). The bifactor model was identified as the best-fitting MIRT model which allows for the calibration and scoring of the TORR in terms of both general relational reasoning ability and specific scale abilities for each of the four forms of relational reasoning. Significant positive correlations at the latent variable level among the forms of relational reasoning were found, however, the correlations were not so large as to suggest that the forms of relational reasoning do not differ in measurable ways. Overall, results supported the TORR as a well-functioning, reliable measure of multidimensional relational reasoning ability for older adolescents and adults in both research and educational contexts due to its ease of interpretability of both overall and specific scale test scores (Dumas & Alexander, 2016).

Federiakin and Aleksandrova (2017) replicated the Dumas and Alexander (2016) study by investigating the multidimensionality of the TORR on an independent Russian sample using Rasch measurement, a psychometric technique that evaluates how well items or questions on assessments work to measure the ability or trait. Participants were 736 fourth year undergraduate Electrical Engineering and Computer Science students at 34 Russian universities. Using the Rasch methodology, the researchers analyzed the dimensionality of the TORR with Principal Components Analysis (PCA) of unidimensional model residuals and found a strong trend of item grouping around each of the four subscales (i.e. analogy, anomaly, antinomy, and antithesis) implying a requirement for multidimensional modeling. The multidimensional modeling clearly demonstrated that the multidimensional model fit the data better than the unidimensional one, and the lower AIC (unidimensional *AIC = 29774.74;* multidimensional *AIC = 29297.27)* and BIC (unidimensional *BIC = 29926.58;* multidimensional *BIC = 29490.52)* indexes provided additional supporting evidence for a better fit of test of multidimensionality. The findings supported the original research findings of Dumas and Alexander (2016) regarding the multidimensionality of the TORR.

Empirical Findings of the TORR

With the development of the TORR, researchers have begun to investigate relational reasoning abilities in a number of different domains (e.g., SAT problems, technological literacy, engineering, and nursing). Alexander and colleagues (Alexander et al., 2016) investigated the convergent, discriminant and predictive validity of the TORR. The convergent validity was established by examining the correlation between scores on the TORR and on a related test of relational reasoning, Raven's Advanced Progressive Matrices (RPM; Raven, 1941) short form. The RPM contains matrix analogy problems that tap one of the forms of relational reasoning of the TORR (i.e., analogy) and therefore the correlation between the RPM and the TORR was of particular interest to the researchers.

Participants were 71 students at a large Mid-Atlantic university and were predominantly juniors and seniors and represented majors in the social sciences/humanities, including psychology, kinesiology, and family sciences. The mean score on the RPM short form was 9.33 (*SD* = 2.21) and the RPM and TORR were determined to be significantly and moderately correlated ($r = .49$, $p < .001$), although the strength of the correlation indicated that they were not capturing precisely the same cognitive processes. These findings were consistent with the conceptual difference between the RPM which tests only analogical reasoning, and the TORR which tests four forms of relational reasoning.

The discriminant validity was determined by examining the correlation between the same sample of undergraduate students' performance on the TORR and with a measure of visual-spatial working memory (VS-WMC). Research has implicated the potential importance of VS-WMC in the successful completion of figural reasoning items such as the RPM (Shipstead & Yonehiro, 2016) which are similar to the items on the TORR, thus, the correlation coefficients between the TORR and a measure of VS-WMC was of interest to the researchers as they wanted to ensure that the TORR was assessing something other than visual-spatial ability. For this analysis, the researchers used the Shapebuilder task (Sprenger et al., 2013) as a measure of VS-WMC. The Shapebuilder task requires participants to maintain a mental representation of serially presented shapes (e.g., circle, square, or triangle) and recall those shapes in sequential order. In addition to order of presentation, the various shapes differed in number displayed, their color, and their location on a grid. The correlation coefficient between the Shapebuilder task and the TORR for was positive but low to moderate ($r = .31$, $p = .02$), implying that while VS-WMC plays a role in participants' ability to correctly respond to the items on the TORR, VS-WMC does not account for an undue proportion of variance in scores.

The predictive validity for the TORR was determined by examining the relation between the TORR and undergraduate students' performance on two sets of released SAT items, one verbal and one math. The SAT problem sets were domain-specific in nature, verbal analogies (linguistic) and mathematics (nonlinguistic) problems, whereas the TORR was conceptualized as a domain-general assessment. Participants were 30 undergraduate students at a large Mid-Atlantic University. Students' average score on the verbal section of the SAT was 8.10 (SD = 2.40) and 9.67 (SD = 3.98) on the math section of the SAT. Participants' average score on the TORR was 15.57 (SD = 3.98). The researchers conducted two linear regressions and found the TORR was found to be a significant predictor of performance on both the verbal section [$F(1, 28) = 16.13$, $p < .001$; $\beta = 0.36$, $t = 4.02$, $p < 0.001$; $R^2 = 0.37$] and math sections [$F(1, 28) = 4.34$, $p < .05$; $\beta = 0.2$, $t = 2.08$, $p < .05$; $R^2 = 0.13$] of the SAT and concluded that the TORR demonstrated good predictive validity with regard to a well-established measure of academic potential - SAT verbal analogies and mathematical problems. Interestingly, while the TORR items are entirely visuospatial in nature, TORR scores predicted verbal SAT scores better than math SAT scores in this study.

Grossnickle and colleagues (Grossnickle et al., 2016) investigated the role of individual differences in relational reasoning. Previous research examining individual differences in relational reasoning typically focused on the relation between various individual difference variables (e.g., working memory, gender) and performance on analogical reasoning tasks (Cho, Holyoak, & Cannon, 2007; Richland & Burchinal, 2013). Therefore, this study examined whether certain relevant individual differences (i.e., working memory capacity, need for cognition, gender) significantly relate to multiple forms of analogical and non-analogical relational reasoning performance. Working memory capacity, need for cognition, and gender were selected to represent cognitive and non-cognitive characteristics with the potential to influence relational reasoning in its multiple forms.

Participants were 71 undergraduate students at a large mid-Atlantic American university. The study included five measures: (a) the Test of Relational Reasoning (TORR), (b) the Need for Cognition Scale-Short Form (NCS), (c) Block Span, (d) Shapebuilder, and (e) Letter-Number- Sequencing (LNS). Gender was self-reported by participants. Pearson correlations were calculated for the relations between each of the variables except gender, which were calculated using a point-biserial correlation. A multiple regression analysis was undertaken to explore the degree to which individual differences were associated with relational reasoning ability when controlling for each of the other characteristics.

Results indicated that visuospatial working memory (VS-WMC) was significantly related to all forms of relational reasoning performance, gender was related only to analogical reasoning, and NFC and phonological working memory (LNS) were not significantly related to any of the measured forms. Moreover, when controlling for VS-WMC, gender no longer accounted for a significant amount of variance in relational reasoning. Although the importance of VS-WMC has been established in the analogical reasoning literature (Cho et al., 2007; Fry & Hale, 1996; Richland & Burchinal, 2013; Waltz, Lau, Grewal, & Holyoak, 2000) the study revealed a similar relation to non-analogical forms of relational reasoning, including anomalous, antinimous, and antithetical reasoning. The findings demonstrated that successful performance on a visuospatial measure such as the TORR rests to some extent on individuals' capacity to retain symbolic information in working memory.

Jablansky and colleagues (Jablansky et al., 2015) conducted the first known large-scale, cross-sectional study examining how different forms of relational reasoning arise and are expressed for both primary and secondary school students. The researchers examined the relational reasoning processes through analysis of semi-structured conversations between students and researcher about the technological features of objects, conducted as part of a New Zealand national research project, Technological Literacy: Implications for Teaching and Learning (TL:IMPS; Compton, Compton, & Patterson, 2011). Analysis of the semi-structured conversations allowed the researchers to see how relational reasoning develops in naturally occurring discussions unfolding in situ.

Participants were 61 students in New Zealand ranging from 5 to 17 years of age. Students were drawn from an existing database consisting of 1,428 New Zealand students to provide a representative sample of approximately every other grade from kindergarten through eleventh grade to explore the developmental differences among students of a wide age range. Grade levels were collapsed into three groups - early, middle, and late. Grades K, 2, and 4 comprised the early group; grades 6 and 8 were combined for the middle group; and grades 10 and 11 comprised the late group. In addition, an equal number of males and females were selected for each group in order to eliminate gender as an explanatory factor. Students came from 19 different schools in both urban and rural New Zealand and a range of socioeconomic backgrounds and ethnically, the sample was representative of the New Zealand population for individuals between the ages of 5 and 17 years old.

The researchers found instances of all four forms of relational reasoning for students even as young as kindergarten. Findings revealed that the quality and quantity of relational talk varied widely among students of different ages. Results showed a curvilinear developmental trajectory in the frequency of relational units with the middle group verbalizing significantly more instances of relational units than both the early and late groups. Based on a proportional analysis, students in the early group demonstrated fewer instances of relational reasoning than would be expected statistically, and fewer instances than the middle group. In addition, the early group's relational reasoning was predominantly analogical in nature, although there were, to a lesser degree, instances of the other three forms. Curiously, the late group demonstrated significantly fewer instances of relational reasoning than expected.

In addition to differences in the frequencies of relational reasoning, there were variations among the groups in the verbalization of the four forms of relational reasoning. The middle and late groups verbalized greater numbers of antinomous and antithetical reasoning, suggesting that it may be easier for older children to use more complex forms of reasoning due to better developed brain structures that allow them to integrate multiple relations at the same time (Crone, Wendelken, van Leijenhorst, Honomichl, Christoff, & Bunge, 2009; Eslinger et al., 2009). At the same time, students in the early group verbalized analogical and anomalous reasoning with greater frequency, suggesting that for younger children, analogical and anomalous reasoning may be used to initially organize and understand basic categorizations of an object or concept (Gentner & Rattermann, 1991). Overall, the research found evidence to support the idea that the different forms of relational reasoning become more and less prevalent in student discourse at various ages.

With the development of the TORR, researchers have begun to use the TORR as a predictor of performance in a number of different domains such as engineering, creativity, and nursing. Dumas and Schmidt (2015) investigated cognitive variables that might predict an increase in novelty of problem solutions of engineering graduate students. Nineteen graduate students enrolled in a mechanical engineering graduate design course at a Mid-Atlantic university completed the TORR, Shapebuilder task and the uses of objects task (UOT), a psychometric test that requires participants to generate multiple original uses for a given every day object. The measures assessed individual differences in the way students cognitively process information (i.e. relational reasoning ability, visuo-spatial working memory capacity, and creativity) and were used as predictors of participants' performance on a traffic light problem after receiving instruction on an engineering inventive principles method known as TRIZ. The participants' scores on the three individual difference measures and their level of increase in novelty of problem solutions after the TRIZ instruction was examined and results revealed that relational reasoning ability as measured by the TORR was a significant predictor of engineering design innovation ($\beta = .84$, $p = .01$) and a significant correlate of creative thinking ability, the Uses of Objects Task ($r = .37$, $p = .042$), while other individual differences in cognition and demographics were not significant predictors. The study was the first to empirically demonstrate, that the gains in the novelty of problem solutions associated with TRIZ instruction are predicted by relational reasoning ability, suggesting that assessments of relational reasoning like the TORR may be effectively utilized to identify individuals most likely to be successful innovators and design engineers.

Dumas (2018) examined the relation between creativity and relational reasoning in a study with 77 undergraduate students at a large public research university in the mid-Atlantic region of the United States. Specifically, he investigated the interrelations between divergent thinking, defined as domain-general creative ability, and relational reasoning, which are thought to be important predictors of creative problem solving (Acar & Runco, 2014; An, Song, & Carr, 2016; Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012). Relational reasoning was measured by the TORR, and divergent thinking by the Alternate Uses Task (AUT). The AUT is the most commonly utilized measure of divergent thinking (Guilford, 1967; Puryear, Kettler, & Rinn, 2017; Torrance, 1972) in which participants are asked to produce as many novel uses for a given every-day object as possible. Results revealed that students who are more capable at relational reasoning, as measured by the TORR, are also more likely to produce more original responses on average on the AUT but only in students at or below the median of the originality distribution. However, for those students above the median of originality, those with greater relational reasoning ability are no more likely to be highly original. The findings suggest that domain-specific relational reasoning interventions may be most effective with groups of students who have been previously identified as moderate to low original thinkers.

Fountain (2017) examined the relations among maternity nurses' topic knowledge, individual interest, relational reasoning, and critical thinking. 182

maternity nurses with varying levels of experience from pre-licensure to very experienced nurses were recruited from national nursing listservs. They completed three independent measures including a domain-specific Topic Knowledge Assessment (TKA), a Professed and Engaged Interest Measure (PEIM) that captured level of interest and engagement in maternity nursing topics and activities, and the Test of Relational Reasoning (TORR). The dependent measure for the study was the Critical Thinking Task in Maternity Nursing (CT2MN) which was composed of a clinical case. The results found that relational reasoning was a significant contributor to critical thinking above and beyond topic knowledge, years of experience, and individual interest. However, relational reasoning did not increase over the professional lifespan, although higher percentages of higher RRQ scores were found in the most experienced group. The findings suggest that the role of relational reasoning in both nursing and critical thinking tasks should be explored further.

These recent findings indicate that the TORR may be used as a predictive measure both for other domain-general cognitive abilities (e.g., creativity and critical thinking) as well across a wide variety of domain-specific academic variables (e.g., engineering design innovation and nursing). Alexander (2016) advances that the TORR could lead to the identification of fundamental cognitive capabilities that might otherwise be overlooked by more traditional crystallized measures of achievement or aptitude. For the purpose of this proposed dissertation, the TORR will be used to determine the degree to which it predicts students' scholastic aptitude on the Preliminary-SAT (PSAT), and on several content-specific achievement measures (i.e., World Cultures, Algebra 1 and Spanish 1 final examinations and the AP psychology examination).

Summary

A number of studies have investigated the relation among crystallized knowledge, fluid reasoning, and academic performance, and identified a significant positive relation between the constructs. However, none have specifically explored the influence of relational reasoning as a multidimensional construct. For example, Nunes, Bryant, Barros, and Sylva (2012) conducted a large-scale longitudinal research over a period of 5 years with a large number of children (n = 1,680) that examined arithmetic knowledge, and mathematical reasoning. The researchers investigated the links between these measures and the children's performance in three school subjects, mathematics, science, and English over five years. The children's mean age at the start of the study was 8 years 6 months and at the end of the study their mean age was 14 years and 1 month. Overall results found that knowledge (as measured by arithmetic skills) made a small significant and independent contribution to the children's achievement in mathematics. A relatively strong relation was found between the children's mathematical reasoning scores and their scores in both mathematics and science.

McGrew and Wendling (2010) conducted a research synthesis of 20 years of research literature of the Gf-Gc cognitive abilities framework and academic achievement. They examined 134 analyses by three age levels (6–8, 9–13, and 14–19) and by four achievement domains (basic reading skills, reading comprehension, basic math skills, and math reasoning). For basic reading skills, Gc had a significant relation across all age levels, and in particular high relation for ages 14 to 19, but Gf was found not to have any significant relation to basic reading skills. Gc showed the most consistent relation with reading comprehension across all age levels, while Gf was only significant for the oldest age level (i.e., 14-19 years old). Both Gc and Gf were found to be significant for basic arithmetic and computational skills with Gc medium for ages 9-19, and Gf medium for all age levels. Both variables were also found to be significant for math problem-solving skills with Gc progressing from low to medium to high for the three age levels, and Gf high for 6-8 and medium for 14-19. Based on their findings, McGrew and Wendling (2010) proposed that both Gc and Gf are important when predicting school achievement. An additional significant finding from the study was that 94% of the research investigated were based on only one cognitive battery, the Woodcock–Johnson Battery (WJ-R; Woodcock & Johnson, 1989). Based on this finding, McGrew and Wendling (2010) recommended the need for additional studies of academic achievement and cognitive abilities that use other batteries relevant to the specific context (i.e., academic domain, ages or grades), instead of an "one size fits all" complete battery. This dissertation will expand the instruments used to assess fluid abilities beyond the historically common indicators that have been used (e.g., RPM and WJ-R) by using the TORR as a measure of fluid reasoning.

Based on this literature review, academic performance of high-school students seems to be dependent on two interconnecting factors—what the student knows (i.e., crystallized knowledge), and what the student is capable of processing (i.e., relational reasoning). At present, the safe statement is that crystallized knowledge, and relational reasoning, are both important to academic performance. Their effects on both one another and on academic performance is likely determined by the nature of what is being assessed—domain-general or domain-specific achievement.

Currently, to my knowledge, there are no studies that have been carried out with adolescents exploring the role of both crystallized cross-domain knowledge, and fluid relational reasoning as a multi-dimensional construct. The study of these constructs in adolescence is important, as the relations and dynamics are different from those in children and adults. Considering that cognitive functions such as reasoning are still developing during adolescence (Choudhury, Charman, & Blakemore, 2008; Diamond, 2016; Krumm, Aran Filippettia, & Gutierrez, 2018; Kuhn, 2006), findings could have important implications for school assessments and curriculum. A focus on individual differences among adolescent learners can provide important insights into intellectual development and improved predictions of academic success (Ackerman, 2014). This dissertation proposes to examine both crystallized cross-domain knowledge and the fluid relational reasoning, in order to further understand which cognitive abilities and individual differences most influence a high school student's scholastic aptitude and content-specific academic achievement.

Chapter 3: Methodology

This chapter describes the participants, measures, and procedures utilized in the current dissertation study. The chapter concludes with an overview of the data analyses to address the research questions pertaining to the contributions of crystallized cross-domain knowledge and fluid relational reasoning on ninth-grade students' scholastic aptitude and content-specific achievement, and twelfth-grade students' content-specific achievement.

Study Context

This research study was conducted at an all-boys Catholic high school located in an urban setting in a large metropolitan area in the mid-Atlantic region of the United States. The school has approximately 960 students from the metropolitan area. Students are primarily White (78%), with Black (11%), Hispanic (6%), Asian (3%), and Other (2%) ethnicities represented as well. Students come from over 100 different middle schools and grade schools within a 45-mile radius. The median annual household income of the students at the school is $100,563.00, however the range goes from a low of $33,000.00 to a high of $191,000.00. On average, 35% of the students receive financial assistance that covers 15% to 100% of the tuition remittance. Approximately 20% of the students have academic accommodations due to some form of diagnosed and documented learning difficulties. As a college preparatory school, this institution offers a values-oriented and academically challenging curriculum. The average SAT score of seniors is 1287 and 99% of graduates attend a four-year college.

Participants

The participants for this study were 211 ninth-grade students (*M* age = 14.82, *SD* = .46) and 76 twelfth-grade students (*M* age = 17.86, *SD* = .45) enrolled in the all-male parochial high school just described. The choice of students in grades nine and twelve was justified on the basis of both scientific and practical factors. First, because of the developmental questions asked about contributors of fluid relational reasoning to scholastic aptitude and content-specific academic achievement, it seemed advisable based on the literature to investigate differences in relational reasoning between early and late high-school students. For instance, research has shown that regions in the brain associated with relational reasoning undergo significant structural and functional development during adolescence (Blakemore & Choudhury, 2006; Crone et al., 2009; Dumontheil, Houlton, Christoff, & Blakemore, 2010; Knoll et al., 2016). These brain regions are involved in a variety of higher cognitive skills relevant to learning and academic performance (Blakemore & Robbins, 2012; Dumontheil, 2014; Houdé, Rossi, Lubin, & Joliot, 2010). However, there has been limited research with adolescents that has examined the role of relational reasoning, and in particular differences between early and late high-school students.

Second, selection of ninth graders was based in part on the fact that these students take the same final examinations (i.e., World Cultures, Algebra 1, and Spanish 1) regardless of teacher. Similarly, all twelfth-grade students enrolled in AP Psychology take the same standardized AP examination. Finally, with regard to relational reasoning, the decision to investigate only male students was seen as nonproblematic, since the TORR was found to be invariant to gender (Dumas & Alexander, 2018).

All ninth-grade students enrolled at the school and all twelfth-grade students enrolled in Advanced Placement Psychology were invited to participate in the study. Participants who were at least 18 years of age gave informed consent to participate in the study, while those under 18 years of age also needed to secure their parent or guardian's permission to participate. Students were also aware that they could drop out of the study at any point and that their anonymity would be guaranteed. Of the 245 ninth graders and 89 twelfth graders in the designated classes, 211 (86%) and 76 (85%) agreed to participate and secured parental or guardian permission. The 76 twelfth-grade students represented 35% of the total twelfth-grade students at the school, and the sample was similar in their HSPT scores with the entire twelfth-grade class (see Table 1). No extra credit or incentives were given for participating in the study. Students who elected not to participate worked quietly on their respective homework assignments.

Table 1

Comparison of HSPT Scores of Twelfth-Grade Sample with the Twelfth-Grade Population

	HSPT Scores						
		HSPT Lang		HSPT Math		HSPT Read	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Sample	76	66.39	23.25	79.57	21.02	73.83	21.72
Population	220	67.78	23.39	80.26	20.97	79.17	21.71

Predictive Measures

The following measures were administered *a priori* to investigate their contributions to scholastic aptitude and content-specific achievement.

Crystallized Cross-Domain Knowledge: High School Placement Test (HSPT)

The students' acquired cross-domain knowledge and skills (i.e., their crystallized knowledge) was assessed with the *High School Placement Test* or HSPT (2013). The HSPT assesses incoming ninth graders' academic readiness levels to determine admissions and placement in freshman-year classes and is used by the parochial high school in this investigation for admissions and placement decisions (see sample items in Appendix A). The HSPT is produced and administered by the Scholastic Testing Service (STS) and has been in continuous use nationally since 1958 (Scholastic Testing Service, 2013). The HSPT is a test battery designed for eighth graders interested in attending a Catholic high school throughout the United States. Every year, 800 to 900 high schools in the United States administer the HSPT to over 100,000 students to make admissions and scholarship decisions about applicants. The HSPT has been found to have strong concurrent validity with similar types of standardized tests such as the Stanford Achievement Test (r = .87), the Iowa Assessments Examination (r = .91), and the TerraNova Test (r = .87). The HSPT has also been found to have strong predictive validity with subsequent high school standardized scholastic aptitude tests such as the SAT (r = .81) and the ACT (r = .78; Scholastic Testing Service, 2013).

The HSPT has five subtests: Verbal Skills (reasoning tasks involving the use of words); Quantitative Skills (reasoning problems involving numbers and quantities); Language Skills (knowledge of capitalization, punctuation, grammar, spelling, usage and composition); Mathematics (arithmetic operations, and applying math concepts to solve problems); and Reading Comprehension (ability to remember important ideas and significant details, recognize central thought or purpose, make logical inferences and understand vocabulary in context). The test is composed entirely of multiple-choice questions with scores reported by subtest, and has an overall composite that reflects a student's total performance on the five subtests that comprise the battery. The HSPT consists of 112 items categorized as reasoning skills (60 in the verbal, 52 in the quantitative section), and 186 items in the basic skills categories (60 in the language, 64 in the mathematics, and 62 in the reading sections). The entire battery is standardized for the population of entering high-school students. For the purposes of this dissertation, the subtests of Language Skills, Mathematics, and Reading Comprehension were used as measures of crystallized knowledge.

For the Language Skill subtest, students respond to 60 questions within a 25-minute time frame. Students are asked a series of questions designed to evaluate their level of knowledge of grammar and writing mechanics. For the Mathematics Skills subtest, student have a total of 45 minutes to respond to 64 questions with no calculator allowed. The 64 questions cover mathematical concepts and skills appropriate for an eighth grade level. These concepts and skills include: Algebraic Equations; Arithmetic skills (i.e. addition, decimals, division, fractions, multiplication, number line, order of operations, percentages, and subtraction);

Geometry concepts; and distance, rate, time word problems. The Reading Skills subtest contains a total of 62 questions and students have 25 minutes to respond to questions that assess their comprehension of concepts and vocabulary understanding within given passages. These passages are from a variety of subjects, such as history, literature, science and societal topics.

Scoring for each subtest is calculated by earning one point for every correct answer, and no penalties for incorrect or omitted answers. The raw score is then converted into a local percentile-rank scale that ranges from 1 to 99 and compares the performance of an individual student with that of other students who were tested at the same time at the same school. Both the ninth-grade and twelfth-grade students took the HSPT in December of their respective eighth-grade year at the all-male parochial high school where this research was conducted. The local percentile-rank scale indicates the percentage of raw scores (i.e., the total number of correct responses) in the local group that are lower than the raw score attained by a given student. If a student earns a local percentile of 71 on the Language subtest, this means the raw score was higher than 71 percent of those in the group tested.

Fluid Reasoning Ability: Test of Relational Reasoning (TORR)

For this investigation, relational reasoning was tested using the Test of Relational Reasoning (TORR; Alexander & the DRLRL, 2012b; Alexander et al., 2016). Both the ninth-grade and twelfth-grade students in this study took the TORR in April 2019. The TORR is a 32-item fluid reasoning measure composed of novel, non-linguistic, figural problems that can be delivered online or in print (see sample items in Appendix B). The TORR was developed for use with older adolescents and adult populations. All problems are figural and all of the information needed to answer each problem is provided within that problem, making it a more culturally-fair assessment (Dumas & Alexander, 2016). The TORR consists of four 8-item scales corresponding to the four forms of relational reasoning described in Chapters 1 and 2 (i.e., analogical, anomalous, antinomous, and antithetical reasoning), along with two practice items per scale. The purpose of the practice items, which are not scored, is to orient students to the nature of the problems in the scale. After each practice item, the students are shown the correct response, but no other feedback is provided on that item or any subsequent item on the test.

The analogy scale, which asks respondents to complete the pattern that is displayed in the given problem, is designed to measure individuals' ability to identify attributional similarity. The anomaly scale, by comparison, is about the identification of discrepancies. Problems in this scale direct the respondents to select the figure that is discrepant from other members of the given set. The antinomy scale involves the recognition of true dichotomies or patterns of exclusivity; that is, objects could only be a case of A or B. For these items, respondents select the one of the four sets displayed below the given that has no member in common the given set. Finally, the antithesis scale is designed to measure individuals' ability to identify a pattern that is the opposite of the given problem. This is achieved on the TORR by having individuals select the answer choice that shows the opposite of the given process. For all four scales, a score of 1 is given for each correctly answered item, excluding the two practice problems, for a total possible score of 32. The TORR has been normed, calibrated, and standardized with a mean of 100 and a standard deviation of 15 (Dumas & Alexander, 2016).

In prior research (Alexander et al., 2016), the TORR was shown to be significantly but moderately related (r = .49, p<.001) to the Raven's Advanced Progressive Matrices (RPM; Raven, 1941), a figural intelligence test constructed of matrix analogies. The TORR was found to be a more challenging test than the Raven's. Since RPM measures only analogy, and not other forms of relational reasoning, this was deemed appropriate convergent validity. Discriminant validity was calculated with a visuospatial measure of working memory, Shapebuilder (Sprenger et al., 2013), with a resulting low moderate correlation of r = .31, p = .02 (Alexander et al., 2016). No differences were found for TORR performance whether the test was administered online or on paper (Alexander et al., 2016). In addition, the TORR has also been found to be invariant with respect to culture, race, or gender for college-age students (Dumas & Alexander, 2018).

In the current study, 287 students completed the TORR in booklet form during their social studies classes: World Cultures for ninth-grade students, and AP Psychology for twelfth-grade students. In the booklet form, the scales are presented in the following order: analogy, antinomy, antinomy, and antithesis. Although performance of the TORR was not timed, all but a few students completed the measure in under 50 minutes.

Outcome Measures

Scholastic Aptitude Measure: Preliminary Scholastic Assessment Test (PSAT)

For the purpose of this dissertation, the ninth-grade students' scholastic aptitude was assessed using their Preliminary Scholastic Assessment Test (PSAT; The College Board, 2014) score. The PSAT is a standardized test that consists of questions in the areas of Reading, Mathematics, and Writing Skills that are similarly found on the SAT. The PSAT is recognized as a valid and reliable measure for academic skill level achievement and college readiness (Kim et al., 2013; Proctor et al., 2010). The PSAT scores are used by high-school educators and counselors to better predict students' trajectory toward college readiness (Kim et al., 2014; Proctor et al., 2010). Validity studies of the PSAT show a moderately strong correlation (.50) with first-year college grade point averages (Proctor et al., 2010).

The PSAT includes five sections: two 25-minute critical reading sections, two 25-minute math sections, and one 30-minute writing skills section. The time allotted to test completion is two hours and 10 minutes (The College Board, 2014). The reading section measures reading skills such as comprehension, inference, and a word's contextual meaning. All the questions are multiple choice and based on provided passages. The mathematics section measures problem-solving skills such as solving or interpreting algebraic equations, and using ratios, percentages, measurement scales, and graphs to solve single and multistep problems. The writing section measures skills such as finding errors and correcting them. All the questions are multiple choice.

The PSAT is standardized across for the various versions that are administered. The standardized composite scores range from 240 to 1440 with a standard deviation of 169. To score the PSAT, first a raw score is computed, one point for each correct answer. One quarter point is deducted for incorrect answers. The raw score is the number of questions answered correctly, nothing is deducted for incorrect answers or for unanswered questions. The raw scores are then converted to composite scores on a scale between 240 and 1440 (The College Board, 2014). The ninth-grade students in this study took the PSAT in April 2019.

Content-Specific Achievement Measures

Three indicators of ninth-grade students' content-specific achievement were analyzed in this study: World Cultures (*n* = 211), Algebra 1 (*n* = 92), and Spanish 1 (*n* = 139). All ninth-grade students enrolled in World Cultures, Algebra 1, and Spanish 1 courses take the same examination, regardless of the teacher they had. The data on the World Cultures, Algebra 1 and Spanish 1 examinations were collected in May 2019. The World Cultures examination was made up of four sections (see sample items in Appendix C). The total points possible on the examination is 100 points. Students have up to two hours to finish the examination. The first section was 100 multiple-choice questions, each worth .5 points for a total of 50 points. In the second section, students are asked to select 10 out of 12 pictures and write two sentences identifying each picture and explaining its significance. Each identification question was worth 1 point each for a total of 10 points.

In the third section, students are asked to select two out of four short essays and write a paragraph for each prompt using specific examples from specific societies that were studied in the class. Each short essay was worth 7.5 points for a total of 15 points. The fourth section was worth a total of 25 points and students were required to write a major essay of at least three paragraphs for the following prompt:

> The Mongol Rally has been described as the greatest adventure in the world. It is also a prime experience of cultural exchange. Participants must drive a small and rubbish car that is deliberately inappropriate for the task 10,000 miles from Western Europe to Ulan-Ude, Russia (via Ulaanbaatar, Mongolia) without a support team, encountering dozens of languages, cultures, and challenges along the way.
>
> Based on our studies this year, assemble a dream team for the rally that consists of any TWO historical figures we have studied this year. Make a persuasive case as to why your team would be uniquely suited to face the challenges of the Mongol Rally. Ensure that your answer is rooted in historical fact, heavy on details, and creative in its interpretation. Consider that different people will be suited for different situations/needs, and that the goal of the Rally is not to beat other cars in a race, but to successfully overcome overwhelming cultural, technical, and physical obstacles over the course of 10,000 mile odyssey. You should directly address these obstacles when making the case for your dynamic duo, and should also at least mention how they will get along with each other.

The Algebra 1 examination was made up of two sections (see sample items in Appendix D). The first section was 25 multiple-choice questions, each worth 2 points each for a total of 50 points. The second section was 10 free-response questions, each worth 5 points each for a total of 50 points. The total points possible on the examination is 100 points. Students have up to two hours to finish the examination, and are not allowed to use calculators for the examination.

The Spanish 1 examination (see sample items in Appendix E) was made up three major sections. The first section was 100 multiple-choice questions each worth 1 point each for a total of 100 points. The second section was a writing section made up of fill-in the blanks and composition worth a total of 130 points. The third section was an oral and listening test worth a total of 35 points. The total points possible is

265 points, and the final grade score is calculated by dividing the total points earned divided by 265 points (e.g., 190 divided by 265 = 72%).

One indicator of twelfth-grade students' content-specific achievement considered in the analysis of this study was the students' Advanced Placement (AP) Psychology examination score. The AP Psychology course is designed to be the equivalent of an Introduction to Psychology course usually taken during the first college year. The AP Psychology course and examination framework contains two components, the course content and course skills. Course skills include concept understanding, data analysis, and scientific investigation. There were no prerequisites for taking the AP Psychology course at the high school used in this investigation. All students who took the course were required to take the AP Psychology examination.

The AP Psychology examination assesses student understanding of the skills and learning targets outlined in the course framework. The examination is two hours long and includes 100 multiple-choice questions and two free-response questions (see Appendix F for the 2019 free-response questions). Scores on the free-response questions are weighted and combined with the results of the computer-scored multiple-choice questions, and this raw score is converted into a composite AP score on a 1–5 scale. A score of 5 represents extremely well qualified performance, a score of 4 represents well qualified performance, a score of 3 represents a qualified performance, a score of 2 represents a possibly qualified performance, and a score of 1 represents no recommendation. AP examinations are not norm-referenced or graded on a curve. Instead, they are criterion-referenced, so every student who meets the criteria for an AP score of 2, 3, 4, or 5 will receive that score, no matter how many students took the exam.

Procedures

Data collection of the current study was conducted in three phases.

Phase 1

In the first phase of the study, a letter and waived consent form was sent home with students. The purpose of the letter was to inform the students' families about the study, and to provide them with an opportunity to contact the researcher should they have any questions about the study. Students' parents or guardians had one week to sign the waived consent form if they wanted their son to participate in the study.

Phase 2

Participants, who were at least 18 years of age, gave informed consent to participate in the study. Participants who were not 18 years old gave their assent along with the signed parental consent form. Next, participants independently completed a short demographic measure, and then the TORR during one social studies class session (i.e., World Cultures class for ninth-grade students and AP Psychology class for twelfth-grade students). The students took approximately 50 minutes to complete the paper version of the TORR. Students recorded their responses to the 32-items on a scantron form. Students took the TORR during their 70-minute social studies class which allowed for enough time to complete without feeling rushed. To protect confidentiality, students were assigned a code when completing the TORR which was linked to their demographic and test data.

Phase 3

Test data for the HSPT, PSAT, and the World Cultures, Algebra 1, and Spanish 1 final examination were collected along the scores for the AP Psychology results from records of student held at the school. All the testing and assessment data are housed in the school's database and can be retrieved by relevant personnel who submit a request.

Overview of Analyses

The statistical analyses of the data were conducted in five stages. First, all data were screened for missing data or any aberrant scores. No abnormalities were detected using descriptive statistics. Missing data were not an issue in this study as all participants completed the TORR and all of the participants are required by the school to take the HSPT, PSAT, AP and semester examinations. Second, descriptive statistics for the independent and dependent variables were generated. Third, assumptions for each of the different statistical analyses were checked. Finally, the statistical analyses to specifically answer the research questions was conducted.

RQ1: What are the differences between ninth and twelfth-grade students on measures of crystallized cross-domain knowledge and fluid relational reasoning?

The first research question of interest pertained to the differences between ninth and twelfth graders on three measures of crystallized cross-domain knowledge, and measure of fluid relational reasoning. A one-way multivariate analysis of variance (MANOVA) was first run to determine if there were any differences between the ninth and twelfth-grade students on the three HSPT subtests of

Language, Mathematics and Reading. Once it was determined that there were no significant differences between the ninth and twelfth-grade students on the three HSPT subtests, an independent groups *t*-test was conducted to investigate if there were any grade differences between ninth and twelfth-grade students on the overall TORR score. This was followed by another one-way MANOVA to explore if there were any grade differences between ninth and twelfth-grade students on each of the four scales of the TORR (i.e., analogy, anomaly, antinomy, and antithesis).

RQ2: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on ninth-grade students' scholastic aptitude and content-specific achievement?

The second research question of interest in this study was the extent to which crystallized cross-domain knowledge and fluid relational reasoning predicted scholastic aptitude and content-specific achievement for ninth-grade students. A series of multivariate multiple linear regression tests were conducted to determine the influence of four predictor variables (i.e., HSPT Language, HSPT Mathematics, HSPT Reading, and TORR) on scholastic aptitude and content-specific achievement. Ninth-grade students' PSAT scores, World Cultures, Algebra 1 and Spanish 1 final examinations scores were individually regressed on the four predictor variables.

RQ3: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on twelfth-grade students' content-specific achievement?

The third research question of interest in this study was the extent to which crystallized cross-domain knowledge and fluid relational reasoning predicted content-specific achievement for twelfth-grade students. A simultaneous multivariate multiple linear regression test was conducted to determine the influence of four predictor variables (i.e., HSPT Language, HSPT Mathematics, HSPT Reading, and TORR) on content-specific achievement. Twelfth-grade students' AP Psychology examination scores were regressed on the four predictor variables.

Chapter 4: Results

The general purposes for this dissertation were twofold. The first purpose was to investigate whether there were differences in crystallized cross-domain knowledge and fluid relational reasoning between ninth and twelfth-grade students. One-way between groups multivariate analysis of variance (MANOVA) and independent groups *t*-tests was employed to analyze data relevant to this question in the study. The second purpose was to explore how well measures of both crystallized cross-domain knowledge, and fluid reasoning abilities are able to predict high-school students' scholastic aptitude and content-specific achievement. A series of standard multivariate multiple regression analyses were performed to investigate (a) the ability of the three HSPT subtests (Reading, Mathematics, and Language) and the TORR to predict performance on ninth-grade students' PSAT and their World Cultures, Algebra 1, and Spanish 1 final examinations; and, (b) twelfth-grade students Advanced Placement (AP) Psychology examination.

Descriptive Statistics

The descriptive statistics for all of the predictor measures in the study are first presented, followed by the descriptive statistics for the outcome measures of the study. Ninth-grade students performed slightly better than the twelfth-grade students on all three subtests of the HSPT, while the twelfth-grade students showed greater spread of scores than the ninth-grade students. Both ninth and twelfth-grade students performed best on the Mathematics subtest, and lowest on the Language subtest. The mean scores for the three HSPT subtests, (i.e., Language, Mathematics, and Reading), for both ninth and twelfth-grade students, as well as the overall mean scores, are presented in Table 2.

Table 2

Descriptive Statistics of the HSPT Subtests by Grade Level

Grade	HSPT Scores						
		HSPT Language		HSPT Mathematics		HSPT Reading	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ninth	211	72.59	21.19	83.69	15.32	78.97	18.55
Twelfth	76	66.39	23.25	79.57	21.02	73.83	21.72
Overall	287	71.00	21.89	82.60	17.08	77.61	19.54

Twelfth-grade students had a higher mean score on the TORR than ninth-grade students and, had less of a range of scores (maximum 30 and minimum 8), than ninth-grade students (maximum 30 and minimum 4). The mean scores for the TORR, for both ninth and twelfth-grade students, as well as the overall mean scores, are presented in Table 3.

Table 3

Descriptive Statistics for Overall TORR Score by Grade Level

Grade	*N*	Minimum	Maximum	Mean	*SD*
Ninth	211	4	30	15.67	5.47
Twelfth	76	8	30	19.43	5.3
Overall	287	4	30	16.67	5.67

Table 4 presents the descriptive statistics for the four scales of the TORR by grade level. Ninth-grade students performed best on the Analogy scale (M = 4.16), and poorest on the Antithesis scale (M = 3.61). Twelfth-grade students performed best on the Anomaly (M = 4.93) and Antithesis scales (M = 4.93), and poorest on the Analogy scale (M = 4.72).

Table 4

Descriptive Statistics for the TORR Scales by Grade Level

Grade	TORR Scores							
	Analogy		Anomaly		Antinomy		Antithesis	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Ninth	4.16	1.54	3.85	1.96	4.05	1.92	3.61	2.20
Twelfth	4.72	1.73	4.93	1.73	4.84	1.89	4.93	1.94
Overall	4.31	1.61	4.14	1.96	4.26	1.94	3.96	2.21

The influence of the four predictor variables (i.e., HSPT Language, HSPT Mathematics, HSPT Reading, and TORR) on four outcome measures for ninth-grade students was investigated in the study. Table 5 presents the descriptive statistics for the four outcome variables.

Table 5

Descriptive Statistics for Ninth-Grade Outcome Measures

Outcome Measures	*N*	Min	Max	*M*	*SD*
PSAT	211	740	1410	1057.68	127.47
World Cultures	211	51	100	87.03	7.82
Algebra 1	92	56	98	78.13	10.95
Spanish 1	139	50	98	77.27	11.94

The influence of the four predictor variables (i.e., HSPT Language, HSPT Mathematics, HSPT Reading, and TORR) on twelfth-grade students content-specific achievement (i.e., AP Psychology examination scores) was investigated in the study. The mean score of the AP Psychology examination was 3.08 with a standard deviation of 1.38. Figure 2 shows the frequency distribution of the AP Psychology examination scores.

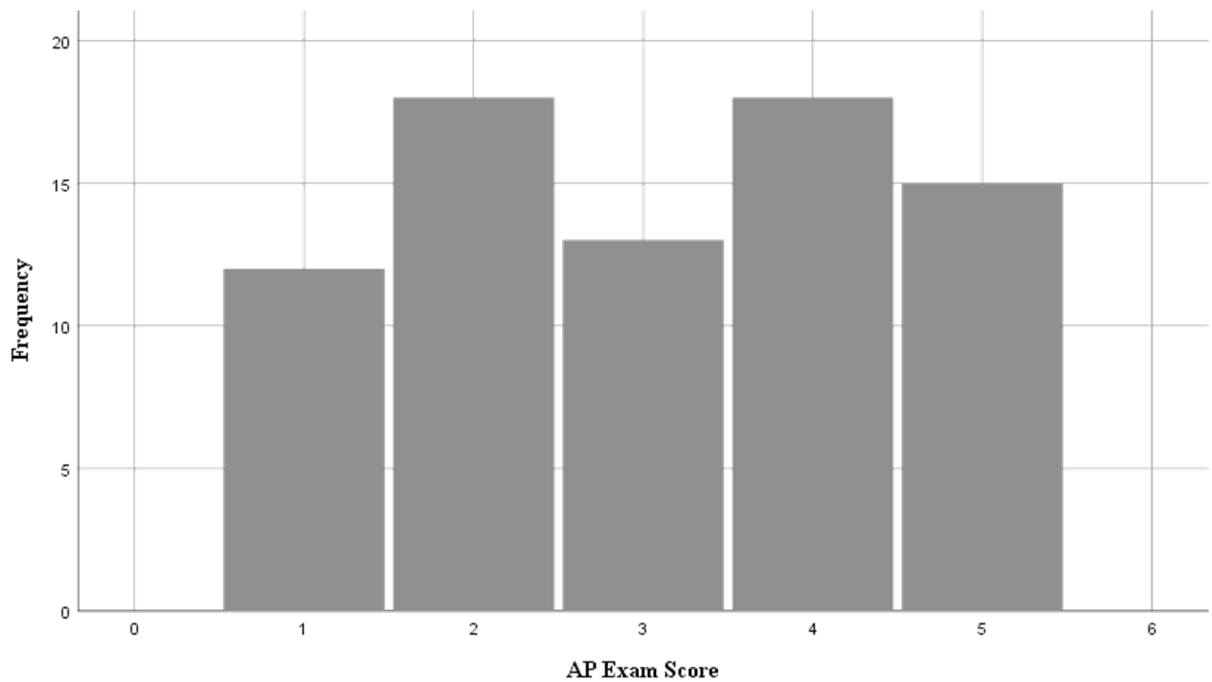


Figure 2. AP Psychology Examination Scores Distribution.

RQ1: What are the differences between ninth and twelfth-grade students on measures of crystallized cross-domain knowledge and fluid relational reasoning?

The first research question investigated whether there were significant differences in both the crystallized cross-domain knowledge measures and the fluid relational reasoning measures for ninth-grade and twelfth-grade students. A one-way between groups multivariate analysis of variance (MANOVA) was first performed to investigate grade differences of crystallized cross-domain knowledge (i.e., HSPT Language, HSPT Mathematics, and HSPT Reading). Prior to conducting the MANOVA, a series of Pearson correlations were performed between all of the dependent variables in order to test the MANOVA assumption that the dependent variables would be correlated with each other in the moderate range (i.e., .30 - .60; Meyers, Gampst, & Guarino, 2006). As can be seen in Table 6, a meaningful pattern of correlations was observed among most of the dependent variables, suggesting the appropriateness of a MANOVA.

Table 6

Correlations between the HSPT Scales

HSPT Scales	Language	Mathematics	Reading
Language	-	.428**	.541**
Mathematics	.428**	-	.408**
Reading	.541**	.408**	-

** Correlation is significant at the 0.01 level (2-tailed).

Multivariate outliers were checked for using Mahalanobis distances, which found no substantial multivariate outliers. Additionally, the Box's M value of 20.62 was associated with a *p*-value of .002, which was interpreted as non-significant based on Huberty and Petoskey's (2000) guideline (i.e., $p < .005$). Thus, the covariance matrices between the groups were assumed to be equal for the purposes of the MANOVA. The homogeneity of variance assumption was tested for all three HSPT subtests. Based on a series of Levene's *F* tests, the homogeneity of variance assumption was considered satisfied, as none of the three Levene's *F* tests were statistically significant ($p > .05$). Scatterplot matrix between the dependent variables showed that all of the dependent variables were linearly related to each other, therefore meeting the assumption of linearity.

Given the preliminary assumptions were satisfied, a one-way MANOVA was conducted to test if there would be one or more mean differences between grade level (ninth and twelfth-grade), and HSPT subtests. There was no statistically significant difference between ninth-grade students and twelfth-grade students on the three dependent variables, F (3, 283) = 2.04, p = .108; eta squared = .02.

The next analysis tested for grade differences on the fluid relational reasoning measure (i.e., TORR). An independent groups *t*-test was performed comparing the mean TORR scores for the ninth-grade students (M = 15.67, SD = 5.47) with those for the twelfth-grade students (M = 19.43, SD = 5.31). Prior to performing the analysis, the data was first checked to make sure assumptions necessary for an independent groups *t*-test were met. The data were normally distributed, and observations were independent of one another. The sample sizes were large enough so that power was not an issue. The Levene's Test of Homogeneity of Variance was nonsignificant (p = .617), suggesting that variances for the two groups are equal. Using an alpha level of 0.05, the independent groups *t*-test was found to be statistically significant, t (285) = 5.19, p<.001 (two-tailed). Specifically, twelfth-grade students performed better on the TORR than the ninth-grade students. The magnitude of the difference (mean difference = 3.77, CI: 2.38 to 5.19, eta squared = .10) was above medium. The effect size for this analysis (d = .66) was found to exceed Cohen's (1988) convention for a medium effect (d = .50), suggesting a moderate to high practical significance.

The final analysis for the first research question explored whether there was a significant difference in the mean scores for each of the four scales of the TORR (i.e., analogy, anomaly, antinomy, and antithesis) for ninth-grade and twelfth-grade students. A one-way between groups MANOVA was performed to investigate grade differences on the four scales of the TORR. The four dependent variables were analogy, anomaly, antinomy, and antithesis, and the independent variable was grade. The data satisfied the preliminary assumptions of absence of multicollinearity, equality of covariance matrices, equality of variances, and multivariate normality. Pearson correlations were performed between all of the dependent variables, and moderate correlations was observed among most of the dependent variables, suggesting and absence of multicollinearity (see Table 7.)

Table 7

Correlations between the TORR Scales

TORR Scales	Analogy	Anomaly	Antinomy	Antithesis
Analogy	-	.467**	.324**	.329**
Anomaly	.467**	-	.388**	.366**
Antinomy	.324**	.388**	-	.427**
Antithesis	.329**	.366**	,427**	-

** Correlation is significant at the 0.01 level (2-tailed).

The Box's M value of 12.51 was associated with a p-value of .269, which was interpreted as non-significant, and the covariance matrices between the groups were assumed to be equal for the purposes of the MANOVA. Additionally, the homogeneity of variance assumption was tested for all four TORR scales. Based on a series of Levene's F tests, the homogeneity of variance assumption was considered satisfied, as none of the four Levene's F tests were statistically significant ($p > .05$).

Multivariate outliers were checked for using Mahalanobis distances, which found no substantial multivariate outliers. Scatterplot matrix between the dependent variables showed that all of the dependent variables were linearly related to each other, therefore meeting the assumption of linearity.

Using an alpha level of 0.01, the MANOVA test identified a statistically significant difference between ninth-grade students and twelfth-grade students on the four combined dependent variables, F (4, 282) = 7.54, $p < .001$; Pillai's Trace = .10; eta squared = .10. The size of these significant relationships ($\eta 2 = .10$) was found to be greater than the value that is typically interpreted as a medium effect size ($\eta 2 = .06$), but smaller than a large effect size ($\eta 2 = .14$; Cohen, 1988).

When post hoc analyses were conducted using a Bonferroni adjusted alpha level of .0025, grade level differences were found for the Anomaly scale, [F (1, 285) = 18.27, $p < .001$; eta squared = .06], and Antithesis scale, [F (1, 285) = 21.49, $p < .001$; eta squared = .07]. An inspection of the mean scores indicated that twelfth-grade students scored significantly higher on Anomaly (M = 4.93, SD = 1.73) and Antithesis (M = 4.93, SD = 1.94) than ninth-grade students (Anomaly: M = 3.85, SD = 1.96) and (Antithesis: M = 3.61, SD = 2.20).

RQ2: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on ninth-grade students' scholastic aptitude and content-specific achievement?

To determine the influence of crystallized cross domain-knowledge, represented by the three HSPT subtests (i.e., Language, Mathematics, and Reading), and fluid relational reasoning, represented by the TORR, on ninth-grade students' scholastic aptitude and content-specific achievement, a series of multivariate multiple linear regression tests were conducted.

The data satisfied the preliminary assumptions of linear relationship, multivariate normality, no auto-correlation homoscedasticity, and absence of multicollinearity. Scatterplots of the standardized residuals were produced and each of the four predictor variable had a linear relation with each of the dependent variables. The check of outliers revealed no standardized residual values above 3.3 or less than –3.3. In addition, multivariate outliers were checked for using Mahalanobis distances, which found no substantial multivariate outliers. A check of the normal probability plots found that the values of the residuals were normally distributed. A check of skewness and kurtosis found that the values did not approach positive or negative 1.0, the rule of thumb for concerning levels of skewness or kurtosis (Osborne, 2013).

Pearson correlations were performed between all of the independent variables, and moderate correlations was observed among most of the independent variables, suggesting and absence of multicollinearity (see Table 8). Additionally, a check of the collinearity diagnostics, revealed no Tolerance values less than .10, and no VIF values above 10.

Table 8

Correlations between the Four Predictor Variables for Ninth-Grade Students

Predictor Variables	Language	Mathematics	Reading	TORR
HSPT Language	-	.312**	.519**	.300**
HSPT Mathematics	.312**	-	.290**	.422**
HSPT Reading	.519**	.290**	-	.415**
TORR	.300**	.422**	.415**	-

** Correlation is significant at the 0.01 level (2-tailed).

After checking for assumptions, the influence of crystallized cross-domain knowledge and fluid relational reasoning on ninth-grade students' performance on their scholastic aptitude (i.e., PSAT) was examined using a simultaneous multiple regression. Crystallized cross-domain knowledge predictors were HSPT Language, HSPT Mathematics, HSPT Reading, and the fluid relational reasoning predictor was TORR. All four predictors were entered into the regression model together. Table 9 summarizes the inter-correlation matrix for the regression model.

Table 9

Ninth-Grade Students PSAT Inter-Correlation Matrix

	HSPT Lang	HSPT Math	HSPT Read	TORR	PSAT
HSPT Lang	1.00				
HSPT Math	.31**	1.00			
HSPT Read	.52**	.29**	1.00		
TORR	.30**	.42**	.41**	1.00	
PSAT	.45**	.56**	.61**	.63**	1.00

** Correlation is significant at the 0.01 level (2-tailed).

The overall multiple regression was statistically significant, [F (4, 206) = 85.31, $p < .001$], with an R^2 of .62, explaining 62% of the variance in the PSAT score, and exceeded Cohen's (1988) convention for a large effect ($R^2 = .35$), suggesting a high practical significance. Two of the crystallized cross-domain knowledge predictors, HSPT Mathematics (β = 2.38, β* = .28, t (206) = 5.93, $p < .001$) and HSPT Reading (β = 2.30, β* = .33, t (206) = 6.33, $p < .001$), had a statistically significantly effect on PSAT. HSPT Mathematics uniquely explained 6% of the variance in PSAT with an effect size of small (β* > .25; Cohen, 1988), and HSPT Reading uniquely explained 7% of the variance in PSAT with an effect size of large (β* > .25; Cohen, 1988). The fluid relational reasoning predictor, TORR, was also a significant predictor of PSAT (β = 8.12, β* = .35, t (206) = 6.95, $p < .001$), uniquely explained 9% of the variance in PSAT with an effect size of large (β* > .25; Cohen, 1988). Table 10 summarizes the results of the regression model.

Table 10

Summary of Regression Analyses for Variables Predicting PSAT

Predictor Variable	*β*	*β**	*SE*	*t-value*	*p-value*
HSPT Language	.53	.09	.31	1.74	.084
HSPT Mathematics	2.38	.28	.40	5.93	.000***
HSPT Reading	2.30	.33	.36	6.33	.000***
TORR	8.12	.35	1.17	6.95	.000***

The next question to be investigated was the influence of crystallized cross-domain knowledge and fluid relational reasoning on ninth-grade students' performance on content-specific achievement (i.e., World Cultures final examination), using a standard multiple regression. Crystallized cross-domain knowledge predictors were HSPT Language, HSPT Mathematics, HSPT Reading, and the fluid relational reasoning predictor was TORR. All four predictors were again entered into the regression model together. Table 11 summarizes the inter-correlation matrix for the regression model.

Table 11

Ninth-Grade Students World Cultures Examination Inter-Correlation Matrix

	HSPT Lang	HSPT Math	HSPT Read	TORR	WC Exam
HSPT Lang	1.00				
HSPT Math	.31**	1.00			
HSPT Read	.52**	.29**	1.00		
TORR	.30**	.42**	.41**	1.00	
WC Exam	.50**	.39**	.60**	.47**	1.00

** Correlation is significant at the 0.01 level (2-tailed).

The overall multiple regression was statistically significant [F (4, 206) = 46.18, $p < .001$], with an R^2 of .47, explaining 47% of the variance in the World Cultures final examination score. The effect size exceeded Cohen's (1988) convention for a large effect ($R^2 = .35$), suggesting a large practical significance. All of the crystallized cross-domain knowledge predictors, HSPT Language, HSPT Mathematics, and HSPT Reading, had a statistically significantly effect on World Cultures final examination with HSPT Reading making the largest unique contribution (i.e., 9%), with a large effect size ($\beta^* = .37$; Cohen, 1988). HSPT Language uniquely explained 3% of the variance, and HSPT Mathematics explained 2% of the variance in the World Cultures final examination. The TORR was also a significant predictor of the World Cultures final examination uniquely explaining 3% of the variance in the World Cultures final examination, with a medium effect size

(β^* = .20; Cohen, 1988). Table 12 summarizes the results of the regression model for 211 students.

Table 12

Summary of Regression Analyses for Variables Predicting World Cultures Final Examination

Predictor variable	*β*	*β**	*SE*	*t-value*	*p-value*
HSPT Language	.08	.21	.02	3.42	.001***
HSPT Mathematics	.07	.14	.03	2.47	.015***
HSPT Reading	.16	.37	.03	5.92	.000***
TORR	.28	.20	.09	3.30	.001***

***$p < .001$

Two further standard multiple regression tests examining the influence of crystallized cross-domain knowledge and fluid relational reasoning on ninth-grade students' performance on content-specific achievement analyses were run with smaller sample sizes. First, the Algebra 1 final examination scores (N = 92), were regressed on the three crystallized cross-domain knowledge predictors (i.e., HSPT Language, HSPT Mathematics, HSPT Reading), and the fluid relational reasoning predictor (i.e., TORR). Table 13 summarizes the inter-correlation matrix for the regression model.

Table 13

Ninth-Grade Students Algebra 1 Examination Inter-Correlation Matrix

	HSPT Lang	HSPT Math	HSPT Read	TORR	ALG1 Exam
HSPT Lang	1.00				
HSPT Math	.14	1.00			
HSPT Read	.45**	.02	1.00		
TORR	.25**	.17	.30**	1.00	
ALG1 Exam	.18	.29**	.24**	.61**	1.00

** Correlation is significant at the 0.01 level (2-tailed).

The overall multiple regression was statistically significant [R^2 = .42, *F* (4, 87) = 15.41, *p* < .001]. The regression model explained 42% of the variance in Algebra 1 final examination scores, and exceeded Cohen's (1988) convention for a large effect (R^2 = .35). The only crystallized cross-domain knowledge predictor to have a statistically significantly effect on PSAT was HSPT Mathematics (β = .14, β* = .20, *t* (87) = 2.35, *p* < .021), meaning that for each score increase on the HSPT Mathematics score, ninth-grade students' Algebra 1 final examination score increased by .14 points. HSPT Mathematics uniquely explained 4% of the variance in Algebra 1 examination scores with a medium effect size (β* = .20; Cohen, 1988). Neither HSPT Language, nor HSPT Reading was found to be a significant predictor of ninth-grade students Algebra 1 final examination scores. The fluid relational reasoning predictor, TORR, was also a significant predictor of Algebra 1 examination scores (β = 1.41, β* = .56, *t* (87) = 2.35, *p* < .001), uniquely explained 9% of the variance in

PSAT with an effect size of large ($\beta^* > .25$; Cohen, 1988). This finding suggests that for each additional score increase on the TORR, ninth-grade students' Algebra 1 final examination scores will increase by 1.41 points. Table 14 summarizes the results of the regression model for Algebra 1 examination scores.

Table 14

Summary of Regression Analyses for Variables Predicting Algebra 1 Final Examination

Predictor Variable	*β*	*β**	*SE*	*t-value*	*p-value*
HSPT Language	-.01	-.02	.05	-.24	.811
HSPT Mathematics	.14	.20	.06	2.35	.021**
HSPT Reading	.04	.07	.05	.79	.434
TORR	1.41	.56	.22	6.37	.000***

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

The final analysis examining the influence of crystallized cross-domain knowledge and fluid relational reasoning on ninth-grade students' performance on content-specific achievement was a standard multiple regression used to test if the three HSPT scales and the TORR significantly predicted ninth-grade students' Spanish 1 final examination scores. Table 15 summarizes the inter-correlation matrix for the regression model.

Table 15

Ninth-Grade Students Spanish 1 Examination Inter-Correlation Matrix

	HSPT Lang	HSPT Math	HSPT Read	TORR	Span1 Exam
HSPT Lang	1.00				
HSPT Math	.31**	1.00			
HSPT Read	.59**	.29**	1.00		
TORR	.42**	.42**	.47**	1.00	
Span1 Exam	.45**	.44**	.44**	.57**	1.00

** Correlation is significant at the 0.01 level (2-tailed).

The results of the regression indicated that the model explained 42% of the variance of Spanish 1 final examination scores, [$R^2 = .42$, F (4, 134) = 23.98, $p <$ .001], with a large effect size (Cohen, 1988). It was found that HSPT Language and HSPT Mathematics significantly predicted Spanish 1 scores, as well as the TORR, which had the largest unique variance of 9%. The TORR has a large effect size ($\beta^* =$ .36; Cohen, 1988) on Spanish 1 examination scores. Table 16 summarizes the results of the regression model for 139 students.

Table 16

Summary of Regression Analyses for Variables Predicting Spanish 1 Final Examination

Predictor Variable	β	β*	*SE*	*t-value*	*p-value*
HSPT Language	.09	.17	.05	2.07	.040**
HSPT Mathematics	.17	.20	.06	2.72	.007**
HSPT Reading	.07	.11	.05	1.23	.219
TORR	.84	.36	.19	4.51	.000***

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

RQ3: To what extent does crystallized cross-domain knowledge and relational reasoning predict performance on twelfth-grade students' content-specific achievement?

To investigate the influence of crystallized cross domain-knowledge, represented by the three HSPT subtests (i.e., Language, Mathematics, and Reading), and fluid relational reasoning, represented by the TORR, on twelfth-grade students' content-specific achievement, a multivariate multiple linear regression test was conducted. Twelfth-grade students AP Psychology examination scores were regressed on their HSPT Language, HSPT Mathematics, HSPT Reading, and TORR scores. Before conducting the multivariate multiple linear regression, the data was tested for assumptions.

The data satisfied the preliminary assumptions of linear relationship, multivariate normality, no auto-correlation homoscedasticity, and absence of multicollinearity. Scatterplots of the standardized residuals were produced and each of the four predictor variables had a linear relation with each of the dependent variables. The check of outliers revealed no standardized residual values above 3.3 or less than – 3.3. In addition, multivariate outliers were checked for using Mahalanobis distances, which found no substantial multivariate outliers. A check of the normal probability plots found that the values of the residuals were normally distributed. A check of skewness and kurtosis found that the values did not approach positive or negative 1.0, the rule of thumb for concerning levels of skewness or kurtosis (Osborne, 2013).

Pearson correlations were performed between all of the independent variables, and moderate correlations was observed among most of the independent variables, suggesting and absence of multicollinearity. Additionally, a check of the collinearity diagnostics, revealed no Tolerance values less than .10, and no VIF values above 10. Table 17 summarizes the inter-correlation matrix for the regression model.

Table 17

Twelfth-Grade Students AP Psychology Examination Inter-Correlation Matrix

	HSPT Lang	HSPT Math	HSPT Read	TORR	AP Psych
HSPT Lang	1.00				
HSPT Math	.63**	1.00			
HSPT Read	.57**	.60**	1.00		
TORR	.28**	.43**	.38**	1.00	
AP Psych	.32**	.41**	.47**	.67**	1.00

** Correlation is significant at the 0.01 level (2-tailed).

The overall multiple regression was statistically significant [F (4, 71) = 17.90, $p < .001$], with a R^2 of .51, explaining 51% of the variance in the twelfth-grade students AP Psychology examination score. The effect size for this analysis (R^2 = .51) was found to exceed Cohen's (1988) convention for a large effect (R^2 = .35), suggesting a high practical significance.

Only one crystallized cross-domain knowledge predictor, HSPT Reading, had a statistically significantly effect on the AP Psychology examination making a unique contribution of 3%, with a medium effect size (β^* = .23; Cohen, 1988). The TORR was a significant predictor of the AP Psychology examination uniquely explaining 26% of the variance in the AP Psychology examination, with a large effect size (β^* = .57; Cohen, 1988). Table 18 summarizes the results of the regression model for 76 students.

Table 18

Summary of Regression Analyses for Variables Predicting AP Psychology Examination

Predictor Variable	β	β*	*SE*	*t-value*	*p-value*
HSPT Language	.002	.03	.007	.224	.824
HSPT Mathematics	.001	.01	.008	.075	.941
HSPT Reading	.015	.23	.007	2.10	.039**
TORR	.148	.57	.025	6.05	.000***

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

Chapter 5: Conclusions, Limitations and Implications

The purposes of this study were twofold. The first purpose was to investigate whether there were differences in crystallized cross-domain knowledge and fluid relational reasoning for ninth- and twelfth-grade students. The second purpose was to determine the degree to which measures of crystallized cross-domain knowledge and fluid reasoning abilities predict ninth and twelfth graders' scholastic aptitude and content-specific achievement. In this chapter, I first consider the delimitations and limitations of this study, and then summarize its major findings. The chapter than concludes with a discussion of the implications of this work for educational practice and for research.

Delimitations

My position as the Director of the Center of Academic Excellence (CAE) at the school attended by study participants guided the purposes of this study. This role has afforded me access to rich student data that many researchers exploring crystallized and fluid cognitive abilities do not have available. At the same time, there are several delimitations or constraints that must be recognized due to the purposes of this study.

The first delimitation was the homogeneity of the participants. All participants were young males from the same geographic area, and most can be characterized as high-achieving students. For instance, the mean HSPT scores for all three subtests were all greater than 70, meaning that the students in the study were in at least the 70th percentile of all students who took the HSPT at the same time. Students who score below 50 on all three of the HSPT subtests are rarely admitted to the school. Also, the mean SAT score of the graduating seniors last year was 1287, while the national average was 1059 (College Board, 2019).

Additionally, this study was conducted with all males which limits the generalizability of the findings to only male high school students. Studies have found no differences in levels of general intelligence between males and females, but have found differences for specific cognitive abilities, in particular fluid intelligence measures of visual-spatial ability and on the Raven's Progressive Matrix (Hedges & Nowell, 1995; Kaufman, Kaufman, Liu, & Johnson, 2008; Lynn & Irwing, 2004; Reilly, Neumann, & Andrews, 2017). On average, males, especially those 15 years and older were found to perform significantly higher on fluid intelligence measures, as well as on math achievement tests (Hyde, Fennema, & Lamon, 1990; Willingham & Cole, 1997).

Another delimitation was that the crystallized measures that were utilized for this investigation had been established for some time. Thus, I had to work within the parameters of those measures and whatever shortcomings they may represent. In fact, one of the goals of this investigation was to consider whether these measures were serving well at identifying the initial potential of applicants or to document their academic development from ninth to twelfth grade.

Limitations

There were several limitations that arose in the course of designing and executing this investigation. Those limitations relate to the cross-sectional nature of the study, the twelfth-grade students' crystallized cross-domain knowledge measure, and the focus on only cognitive abilities to predict academic performance. A significant limitation of this study that needs to be considered when looking at the results is that this was a cross-sectional study, thereby limiting any conclusions one can make regarding the direction of causality, or the nature of crystallized cross-domain knowledge and fluid relational reasoning influences over time. One might as easily postulate that fluid relational reasoning will lead to better academic performance as the reverse.

A second limitation was that the HSPT scores for the twelfth-grade students were from four years earlier when the students were still in eighth-grade. The influence of the crystallized cross-domain knowledge on the twelfth-grade students' AP Psychology examination may not have accurately reflected what their crystallized cross-domain knowledge was at the time of the examination. There was a possibility of using the twelfth-grade students' PSAT scores from eleventh-grade or their SAT scores from earlier in the school year. However, both the PSAT and SAT tests include elements of crystallized knowledge *and* fluid reasoning, which would make those tests impure measures of crystallized cross-domain knowledge.

A final limitation of this study is that only the cognitive abilities of crystallized cross-domain knowledge and fluid relational reasoning were examined in relation to academic performance. Other cognitive abilities such as executive function processes (Baggetta & Alexander, 2016), as well as a variety of social factors such as perceived competence and autonomy (Wiest, Wang, & Kreil, 1998), motivation (Wentzel & Wigfield, 1998), and the influence of teacher expectations and peer relations (Wentzel, Baker, & Russell, 2012) all play a role in high school students' academic performance.

Major Findings

This study provided insights into the role of crystallized cross-domain knowledge and fluid relational reasoning on high school students' academic performance. There were two major findings from this study – the strength of relational reasoning in predicting academic performance; and the evidence of the development of relational reasoning from ninth to twelfth-grade.

The Predictive Power of Relational Reasoning

One of the most notable findings of this investigation was the strong performance of fluid relational reasoning as a predictor of performance for four of the five outcome measures. Specifically, it was found that fluid relational reasoning was the strongest unique predictor of performance for ninth-grade students on the PSAT, Algebra 1 and Spanish 1 final examinations, and for twelfth-grade students on the AP Psychology examination. Fluid relational reasoning explained over 25% of the variance on the ninth-grade students' Algebra 1 final examination and the twelfth-grade students' AP Psychology examination. Fluid relational reasoning also uniquely explained 9% of the variance on the ninth-grade students' PSAT and Spanish 1 final examination. The only measure which fluid relational reasoning was not the strongest predictor was on the ninth-grade students' World Cultures final examination. Fluid relational reasoning had a large effect size on PSAT, Algebra 1 examination, Spanish 1 examination, and AP Psychology examination scores, suggesting that students who have greater fluid relational reasoning abilities may perform better on these assessments.

Crystallized cross-domain was not as strong of a predictor as fluid relational reasoning on the five outcome measures. HSPT Language was found to only be a significant predictor for the ninth-grade students' World Cultures and Spanish 1 examinations, and the unique contribution was minimal (i.e., World Cultures 3% and Spanish 1 2%). HSPT Mathematics was found to be a significant predictor for four of the five outcome measures: PSAT (6% of the variance); Algebra 1 (4% of the variance); Spanish 1 (3% of the variance); and World Cultures (2% of the variance). HSPT Reading was found to be a significant predictor for three of the five outcome variables: World Cultures (9% of the variance); PSAT (7% of the variance); and AP Psychology (3% of the variance).

Results from this study suggest that fluid relational reasoning plays a much greater role than crystallized cross-domain knowledge in both scholastic aptitude and content-specific achievement. The findings of this study are aligned with a number of previous studies that also found fluid reasoning to have a stronger effect than crystallized knowledge on academic performance. Both Nunes et al. (2012) and Taub et al. (2008) found that reasoning abilities had a strong effect on mathematics achievement, while crystallized knowledge (i.e., arithmetic knowledge) only had a small effect. The results of fluid relational reasoning as a significant predictor of scholastic aptitude tests like the PSAT is also in line with findings by Alexander et al. (2016) that the TORR was a significant predictor on a SAT items assessment. The results also support the findings of studies that have found that reasoning abilities have a greater influence on subjects that demand understanding and reasoning about complex relations, such as mathematics and science, and crystallized knowledge has a greater influence on humanities and language subjects (Evans et al., 2002; Schipolowski et al., 2015). Relational reasoning had the strongest effect on the two content-specific subjects that required students to understand relations among concepts, Algebra 1 and AP Psychology.

Overall, the results from the study supports Cattell's Investment Theory, which states that fluid intelligence is the basis for the development of crystallized intelligence. Compared to measures of crystallized cross-domain knowledge, fluid relational reasoning more strongly predicted performance in the classroom. With regard to academic performance, measures of crystallized cross-domain knowledge were found to positively predict scholastic aptitude and content-specific achievement. However, fluid relational reasoning was found to be a superior predictor of academic performance compared to their crystallized cross-domain knowledge counterparts for this study’s sample. This finding was true for both ninth and twelfth-grade students.

Evidence of Relational Reasoning Development

No mean differences were found between the ninth and twelfth-grade students on the three crystallized cross-domain knowledge measures of HSPT Language, Mathematics and Reading. This finding allows for the assumption that when both the ninth and twelfth-grade students first entered high school they were similar in their crystallized cross-domain knowledge, and that any differences found in their fluid relational reasoning abilities may be due to the grade differences. A significant finding of this study was that twelfth-grade students performed better on the TORR than the ninth-grade students. In addition, twelfth-grade students performed better on all four scales of the TORR, and specifically had significantly higher scores on the antithesis scale, than the ninth-grade students. Ninth-grade students performed the strongest on the analogy scale and weakest on the antithesis scale. An interesting finding was that out of the four scales, the twelfth-grade students performed poorest on the analogy scale.

The findings of twelfth-grade students performing better than the ninth-grade students support previous studies that have shown increases in reasoning ability with age (Andrews & Halford, 1998; Richland, Morrison, & Holyoak, 2006). This increase in ability may be a product of the combination of the development of increased relational knowledge that allows one to focus on structural features, and not simply the surface features of relations (Crone et al., 2009; Dinsmore, Baggetta, Doyle, & Loughlin, 2014; Rattermann & Gentner, 1998), and brain maturation changes (Dumonthiel, Houlton, Christoff, & Blakemore, 2010; Ferrer, O'Hare, & Bunge, 2009). Functional Magnetic Resonance Imaging (fMRI) studies involving reasoning tasks in adults have demonstrated that a region in the anterior prefrontal cortex, known as the rostrolateral prefrontal cortex (RLPFC), is activated when participants engage in relational integration during reasoning tasks (Bunge et al., 2005; Christoff et al., 2001; Kroger, Sabb, Fales, Bookheimer, Cohen & Holyoak, 2002; Wendelken et al., 2008). Structural brain development during adolescence consists of reductions in synaptic density and increases in axonal myelination. The RLPFC exhibits these structural developmental changes until early 20's (Dumontheil, Burgess, & Blakemore, 2008). This suggests that the RLPFC involvement in reasoning tasks may go through several developmental stages, with older adolescents more likely to recruit this area than younger adolescents. As seen by the results of this study, these substantial structural changes in the RLPFC during adolescence may contribute to improvements in relational reasoning ability as one gets older.

That the twelfth-grade students performed better on all four scales, and in particular, significantly better than the ninth-grade students on the antithesis scales is not surprising. Possible reasons for this may be that antithesis represents a more complex form of relational reasoning (Dumas et al., 2013), and that older adolescents are more neurologically capable to reason antithetically because of the development of the RLPFC. Also, not surprising was that the ninth-grade students performed best on the analogy scale and poorest on the antithesis scale. Analogical reasoning is considered as the basic organizing principle in developing an understanding of a relation (Gentner & Rattermann, 1991), and from a neurological perspective may be an easier relational integration for younger adolescents.

What was somewhat surprising, however, was the finding that the twelfth-grade students performed poorest on the analogy scale. Jablansky and colleagues (2015) found similar results in their investigation of relational discourse of primary and secondary school students. They found evidence that certain forms of relational reasoning become more and less prevalent in student discourse at various ages. Specifically, they found that older students produced fewer analogies and anomalies and, produced more antinomies and antithesis in their discourse than younger students. They suggested that this curvilinear pattern may reflect the older students' familiarity with the task, thus reducing the need to intentionally reason relationally. This also might have been the case in the findings of this study. The twelfth-grade students might have been more familiar with analogies, and therefore approached those problems in an inappropriate way by responding too hastily.

Implications for Instructional Practices

I will first discuss direct implications of the study for the school, and then broader implications for practice. Currently, all incoming ninth-grade students at the school are required to take a foreign language for three years beginning in their ninth-grade year. However, there has been discussion about delaying foreign language until the tenth-grade year for some of the students who may not be ready for foreign language. This discussion is based on the fact that school data shows that one of the two highest courses percentage-wise that ninth-grade students fail is Spanish 1. The discussion has been centered on using the HSPT Language and HSPT Reading subtests as the basis for the decision. However, based on the results of the study, the TORR may be a better measure of success in Spanish 1 as it uniquely explained 9% of the variance in Spanish 1 scores, while HSPT Language only explained 2%, and HSPT Reading was found to not be statistically significant.

Another curriculum issue in discussion at the school is mathematics course placement for incoming ninth-grade students. Currently, incoming ninth-grade students get placed in different levels of mathematics courses based on their HSPT Mathematics score. However, there is concern that about 20% of the students are not placed in the correct courses based on the grades the students receive. Some of the students in the lowest level course, Algebra 1, get through the course with ease, while some of the students in the higher level courses struggle greatly. Based on the results of the study, the TORR may be a stronger predictor than the HSPT Mathematics subtest. The multiple regression for the Algebra 1 final examination found the TORR uniquely explained 27% of the variance, while HSPT Mathematics only explained 4%. A longitudinal study where the students take the TORR at the beginning of the school year and then again at the end of the school year should be undertaken before making decisions on using the TORR for course placement.

A third curriculum issue is the practice of using students' grades as the pre-requisite for getting into an Advanced Placement (AP) course at the school. Currently, students must first meet a certain a grade point average to apply to take an AP class, and then obtain a teacher recommendation before being considered for the class. Based on the findings from this study, the TORR score could also be applied to the formula for selection to an AP class. AP classes require high levels of critical thinking and analytical skills to prepare for the AP Exam, and the future courses students will take throughout college.

One of the key initiatives of the CAE in the past year was the creation of a Seminar course for incoming ninth-grade students identified as needing additional support and scaffolding based upon their HSPT scores. Each day, students attend a direct instructional period focused on developing and strengthening written language, reading comprehension and verbal communication skills within the context of the curriculum, and are taught strategies to improve their self-regulation. Based on the results of this study, the explicit teaching of relational reasoning skills as part of the curriculum of the course, may help these students to perform better in their other classes, in particular mathematics and foreign language classes.

There are also some broader educational implications from this study. Richland and Simms (2015, p. 188) argue that "the key to teaching higher order thinking across mathematics, science, and history is to conceptualize learning as developing and manipulating relational systems." Educators may want to try to restructure the classroom from one of memory processing of information by students to one of using relational reasoning to understand and connect concepts not only within one domain but cross-domains. However, educators may need to provide explicit instruction and support to help students to notice and attend to, and reason about the different forms of relations they may encounter. For example, teachers discussing the governments of ancient Greece may want to explicitly state and point out differences between Athens and Sparta as antithetical types of government.

Relational reasoning could also greatly support learning for which an individual has to integrating new and disparate bits of information in order to jointly consider several relations among mental representations, often without relying solely on prior knowledge. The results of this study suggest that relational reasoning may be particularly helpful for learning math, which is hierarchical in nature and requires individuals to solve novel problems as each new level advances. Unfortunately, relational reasoning has not typically been emphasized in current mathematics curricula (Tabatabaee-Yazdi & Baghaei, 2018). Even students with strong basic numerical and arithmetic skills may not be proficient in applying relational reasoning to solve novel problems. For this reason, mathematics curriculum may want to incorporate opportunities for students to understand, apply, and practice relational reasoning strategies (Miller-Singley & Bunge, 2014). Additionally, the assessment of relational reasoning in schools could serve to identify students who are likely to have difficulty learning math. This information could help guide teachers to better understand which interventions may be most helpful for individual students at different developmental levels of mathematics achievement skills.

Implications for Research

In addition to the local and broad practical implications of the study, there are several research implications. The current research provides a necessary extension to the developing literature of relational research by examining the role of relational reasoning and its association to academic performance. A useful extension to this research would be to explore grade and age-related changes in students' relational reasoning. Most of the recent studies of relational reasoning have relied largely on cross-sectional research. Hence, it would be beneficial to conduct longitudinal investigations of relational reasoning, especially those that include key cognitive, and neurological developmental time points such as ninth-grade and college transitions. The longitudinal studies should not only investigate if changes do occur, but also why there may be changes in relational reasoning patterns. The question would be if the changes are due to their school experiences or are there changes because of brain structural changes due to maturation. A longitudinal study would allow us to observe the changes in relational reasoning the students undergo over time and, would enhance the strength and generalizability of present findings. Also, a longitudinal study would allow us to examine what classroom practices and behaviors may be conducive to the development of relational reasoning. An alternative to a longitudinal study would be to conduct an intervention study. An intervention study might give us a clearer understanding of the development of relational reasoning capabilities of high school students. The present study also examined the link between relational reasoning and mathematics, foreign language, and social studies. A useful way to extend the current findings would be to examine the link between students' relational reasoning abilities and other educational domains such as in reading and literature. For example, some reading-based reasoning tasks involve using relations based on the spelling– sound pattern of one word as a basis for solving another unfamiliar word (e.g. reading the word aloud) in a context in which learners are explicitly given clue words that might help them to make the relations. An interesting question would be if proficiency in a figural task reasoning task such as the TORR is also evident in verbal and reading reasoning tasks.

A question for future research is the association between relational reasoning and executive function processes such as working memory and inhibitory control. A large body of evidence has suggested that working memory may be strongly related to analogical reasoning (Cho, Holyoak, & Cannon, 2007; Fry & Hale,1996; Krawczyk et al., 2008; Richland & Burchinal, 2013; Waltz, Lau, Grewal, & Holyoak, 2000). Along with working memory, the ability to inhibit interfering information while processing and working to solve the task may be particularly critical as relational complexity (i.e., the number of relations) and problem difficulty increase (Birney & Halford, 2002; Cho et al., 2007). However, the number of studies that have examined the relation between working memory and inhibitory control and other aspects of relational reasoning, i.e. anomaly, antinomy, and antithesis, is almost non-existent. Grossnickle and colleagues (2016) examined the degree to which working memory was related to college students' performance on the four multiple forms of relational reasoning and found that working memory was significantly related to all of the four forms of relational reasoning performance. Considering that both relational reasoning and executive functions are still developing during adolescence (Diamond, 2016; Krumm, Aran Filippettia, & Gutierrez, 2018), it would be of benefit to investigate the relation between them in high school students.

A final question for future research would be the malleability of relational reasoning. Alexander (2016) argues that relational reasoning abilities can be directly taught, however, what is in question is if the training should be domain-general or domain-specific. An intervention study with groups using both domain-general and domain-specific could be conducted and then examine differences between the two types of training. A follow up question could be if the effect of the training is just immediate or has long-term effects.

This study examined the role of crystallized cross-domain knowledge and fluid relational reasoning on high school students' academic performance and found that fluid relational reasoning was an important contributor. This supports the belief that the ability to reason relationally will become even more critical both in and out of school as the world rapidly changes to become more of an environment where the rules are often unclear or incomplete, and individuals have to be able to recognize and understand ever changing and unfamiliar patterns. Those with the ability to quickly integrate information, connect new ideas, and work across contexts by being able to see the different relations will be able to successfully adapt to this new world paradigm.

Appendices

Appendix A: High School Placement Test (HSPT) Sample Items

Basic Algebra

1. If Lynn can type a page in p minutes, what piece of the page can she do in 5 minutes?

    A. 5/p
    B. p - 5
    C. p + 5
    D. p/5
    E. 1- p + 5

2. If Sally can paint a house in 4 hours, and John can paint the same house in 6 hours, how long will it take for both of them to paint the house together?

    A. 2 hours and 24 minutes
    B. 3 hours and 12 minutes
    C. 3 hours and 44 minutes
    D. 4 hours and 10 minutes
    E. 4 hours and 33 minutes

3. Employees of a discount appliance store receive an additional 20% off of the lowest price on an item. If an employee purchases a dishwasher during a 15% off sale, how much will he pay if the dishwasher originally cost $450?

    A. $280.90
    B. $287
    C. $292.50
    D. $306
    E. $333.89

Antonyms

1. DOTE:

    A. Aversion
    B. Antidote
    C. Foolish
    D. Creativity
    E. Daring

2. IMBROGLIO:

   A. Fight
   B. Conclusion
   C. Trust
   D. Thankfulness
   E. Harmony

Averages and Rounding

1. Round 907.457 to the nearest tens place.

   A. 908.0
   B. 910
   C. 907.5
   D. 900
   E. 907.46

2. At a certain high school, the respective weights for the following subjects are:
   Mathematics 3, English 3, History 2, Science 2 and Art 1.What is a student's average whose marks were the following: Geometry 89, American Literature 92, American History 94, Biology 81, and Sculpture 85?

   A. 85.7
   B. 87.8
   C. 88.9
   D. 89.4
   E. 90.2

Basic Operations

1. Add 0.98 + 45.102 + 32.3333 + 31 + 0.00009

   A. 368.573
   B. 210.536299
   C. 109.41539
   D. 99.9975
   E. 80.8769543

2. Find 0.12 ÷ 1

   A. 12
   B. 1.2
   C. .12
   D. .012
   E. .0012

3. (9 ÷ 3) x (8 ÷ 4) =

A. 1
B. 6
C. 72
D. 576
E. 752

Commas

The following sentences either have existing or require additional commas somewhere in their
structures. Choose the option that best reflects proper comma usage in each sentence.

1. For the Thanksgiving reunion, relatives were sitting in the dining room, on the porch, and in the carport.

A. Thanksgiving, reunion
B. were, sitting
C. porch and
D. No error

Intermediate Grammar

1. The word boycott derives from the name of Charles C. Boycott, an English land agent in Ireland that was ostracized for refusing to reduce rent.

A. that was ostracized for refusing
B. who was ostracized for refusing
C. which was ostracized for refusing
D. that had been ostracized for refusing
E. who had been ostracized for refusing

Advanced Grammar

Each of the following sentences contains an error of some kind. Read each sentence and select the option that correctly identifies its error.

1. David was known for belching; and telling inappropriate jokes in public.

A. Capitalization
B. Punctuation
C. Spelling
D. Grammar

Reading for the Main Idea

Read the passage below and answer question 1.

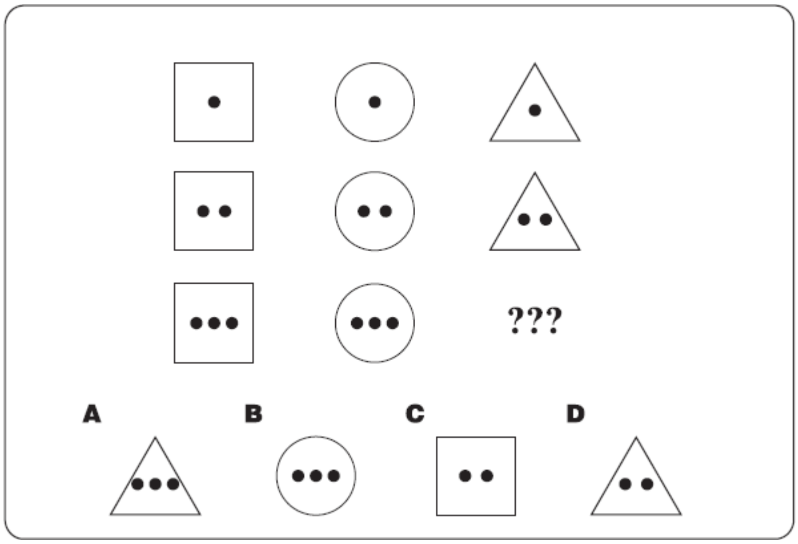
1. What is the main idea of this passage?

    A. The Humanitarian work of the First Ladies is critical in American government.
    B. Dolly Madison was the most influential president's wife.
    C. Eleanor Roosevelt transformed the First Lady image.
    D. The First Ladies are important figures in American culture.
    E. The First Ladies are key supporters of the Presidents.

Appendix B: Sample Items from the Test of Relational Reasoning

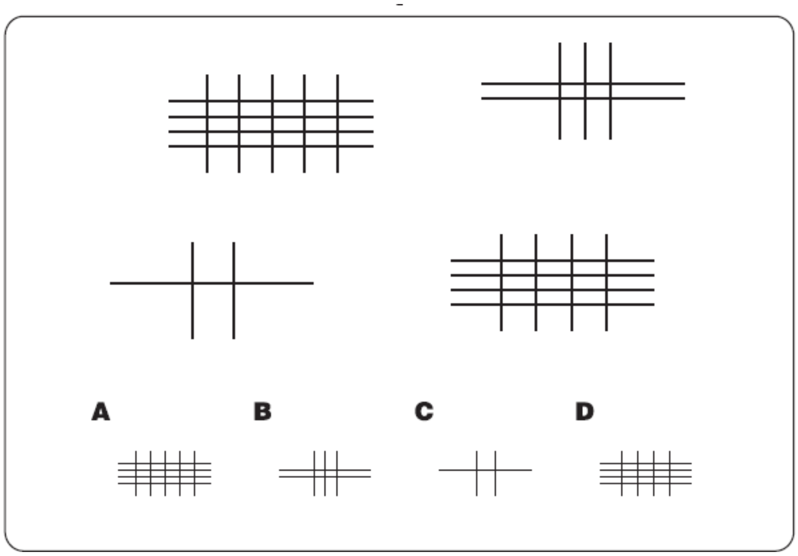
Analogy Scale

Directions: Below is a pattern that is not yet complete.
Select the figure from those shown below that completes the pattern.



Anomaly Scale

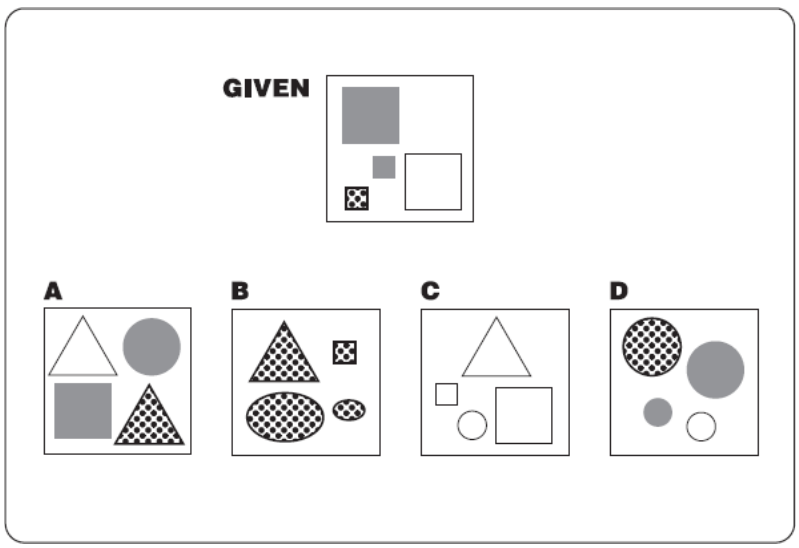
Directions: All these figure but one follow a particular pattern or rule.
Find the one figure that does not follow the pattern.



Antinomy Sale

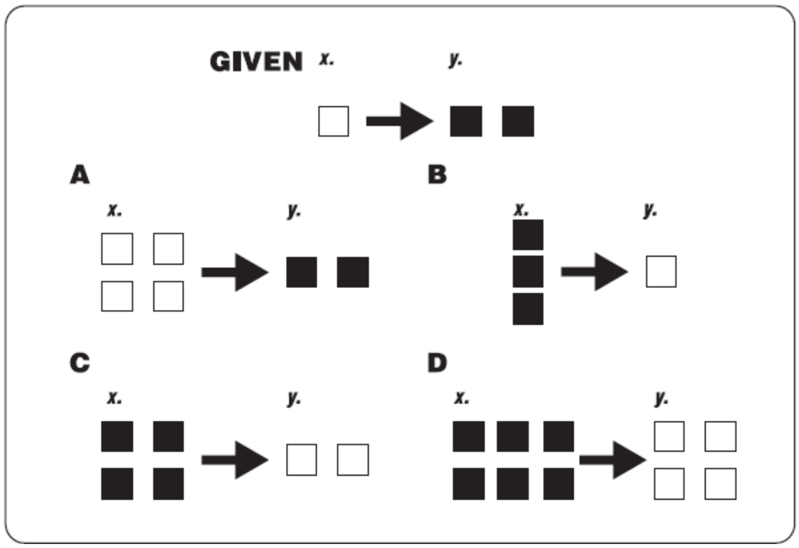
Directions:

- The problems in this section ask you to compare sets of objects that vary in certain
  features
- Each set has a specific rule that decides what objects can be included in that set.
- Some of the objects included in each set are pictured, enough to allow you to determine its rule for inclusion.
- Every problem asks you to identify which ONE of the four sets that are shown could NEVER have an object in common with the Given set based on the compatibility of their rules for inclusion.
- There will always be EXACTLY ONE set this is compatible with the Given set.



Antithesis Scale

Directions: The given figure below depicts a *process* in which X becomes Y. In the figure, the arrow represents the rule by which the change occurs. Select the answer choice
that shows the *opposite* of the given process.



Appendix C: World Cultures Final Examination Sample Items

Multiple Choice:

1) The Gracchus brothers' main political goal was
   a. The abolition of slavery
   b. Preservation of patrician power
   c. Restructuring the military
   d. Land reform

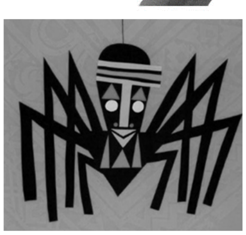
2) The Romans sought to balance the best aspects of three different forms of government, all of the following except:
   a. Democracy
   b. Monarchy
   c. Oligarchy
   d. Aristocracy

3) This infamous emperor was said to have named his horse a senator, slept with his sisters, named himself a god, and killed people for no reason
   a. Tiberius
   b. Claudius
   c. Caligula
   d. Nero

4) Julius Caesar was assassinated because
   a. He was not properly honoring the Roman gods
   b. His appetite for sex and violence horrified the senate
   c. He seemed to have an ambition to be the King of Rome
   d. His economic policies were a disaster for the republic

5) Marc Antony committed suicide with
   a. Lepidus
   b. Cleopatra
   c. Crassus
   d. Octavia

Identifications: Write 2 sentences identifying each picture & explaining its significance



ESSAYS: In Blue Book

SHORT ESSAYS: Pick 2 of the following and write a paragraph for each prompt (at least 5 sentences) that answers every part of the question. Superior essays will demonstrate a firm grasp of the subject and use specific examples from specific societies.

1) Hinduism, Buddhism, Christianity, and Islam have outlasted most of the civilizations we have studied. **Why do you think civilizations fall while religions endure?**

2) Jared Diamond called agriculture "the worst mistake in the history of the human race," asserting that "Hunter-gatherers practiced the most successful and longest-lasting lifestyle in human history. In contrast, we're still struggling with the mess into which agriculture has tumbled us, and it's unclear whether we can solve it." **What does humanity lose by living in "civilized societies," and is it too high of a price to pay?**

3) The famous poem "Waiting for the Barbarians" ends with the lines: *And now, what's going to happen to us without barbarians? / They were, those people, a kind of solution*. How might barbarians be a "solution" for a civilized societies? In other words - **how does contact with (or the mere threat of) uncivilized "barbarians" help civilized societies?**

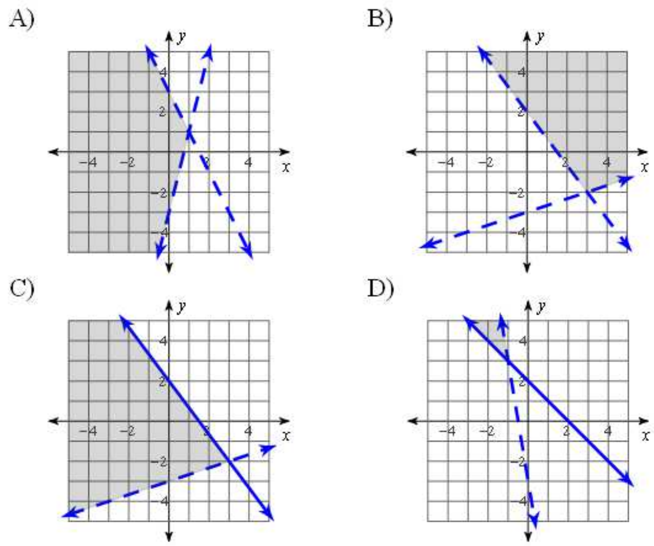
4) Based on your knowledge of Rome and your knowledge of Islam, what do you think might have happened if the two had existed at the same time and interacted?

Appendix D: Algebra 1 Final Examination Sample Items

Part I – Multiple Choice
<u>DIRECTIONS:</u> Use the scantron sheet to enter your answers.
USE PENCIL ON THE SCANTRON SHEET.

1. *Which is the solution set for the following linear inequalities?*

$y > 4x - 3$
$y < -2x + 3$



2. *Solve the system.*
   $10x - 3y = 35$
   $4x + y = 3$

   A. $(2,5)$

   B. $(2,-5)$

   C. $(5,2)$

   D. $(-5,-2)$

3. Solve |x-8|≤20

   A. -28≤x≤28

   B. x≤-28 or x≥28

   C. -12≤x≤12

   D. -12≤x≤28

4. Simplify: $3\sqrt{2}+5\sqrt{3}+4\sqrt{2}$

   A. $12\sqrt{7}$

   B. $7\sqrt{2}+5\sqrt{3}$

   C. $12\sqrt{12}$

   D. $8\sqrt{5}+4\sqrt{2}$

Part II – Free Response

DIRECTIONS: Show all relevant work in the space provided. Only work shown clearly in the answer area will be considered for credit. **Write final answer on the answer line**.

1. *Using the quadratic formula,* $x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$, *solve for* x *in the following equation. Simplify completely.*

$x^2 - 3x = 1$

$x =$ ____________________

2. *Solve for* x *using an algebraic method (i.e. not using guess and check):*
   $\sqrt{2x+3} - 4 = 1$

$x =$ ____________________

3. *An architect wants to design a rectangular building such that the area of the floor is* $119\ yd^2$. *The length of the floor is to be* $10\ yd$ *longer than the width. Find the length and width of the floor.*

Set up a quadratic equation in one variable. Solve the equation to find the dimensions, using an algebraic method (i.e. not using guess and check). Leave your answers in simplified, radical form.

$width =$ ____________ $length =$ ____________

Appendix E: Spanish 1 Final Examination Sample Items

Part I – Multiple Choice

VOCABULARIO Y GRAMÁTICA Escoge la respuesta que mejor complete cada oración. (*Hint: Choose the best answer and fill in the corresponding oval on your answer sheet*.)

Seleccionar - In each group, choose the item that does not belong.

1. a. avión
   b. agente de viajes
   c. motocicleta
   d. auto

2. a. viento
   b. décimo
   c. octavo
   d. cuarto

3. a. la huésped
   b. el ascensor
   c. el botones
   d. el viajero

4. a. estar de vacaciones
   b. ir de vacaciones
   c. estar nublado
   d. hacer un viaje

5. a. otoño
   b. primavera
   c. mes
   d. invierno

6. a. llover
   b. acampar
   c. jugar a las cartas
   d. pescar

Part II- Writing

PREGUNTAS GENERALES: Answer in **complete sentences** in Spanish (Remember correct **verb tense**, and correct noun/adjective agreement).

1.¿Qué comiste en la cena ayer? (3 things)

____

____

____

2. ¿Qué haces con tu familia los sábados? (2 activities)

____

____

____

3. Describe a un amigo(a) con dos características de la personalidad y dos de la descripción física.

____

____

____

4. ¿A dónde fuiste en las vacaciones y qué hiciste? (At least two activities)

____

____

____

5. ¿Qué ropa te pones para venir a Gonzaga? (3 cosas)

____

____

____

Reflexive Verbs Complete the following sentences with the correct **present tense** form of the reflexive verb in parenthesis.

1. Tú ________________________________ (despertarse) con el despertador.
2. Nosotros _________________________ (acostarse) a las diez.
3. Yo ______________________________ (dormirse) fácilmente.
4. Ellos tienen que _______________________ (levantarse) temprano todos los días.

Juan __________________________ *(lavarse) la cara con jabón.*

La composición. Write **seven (7) complete sentences** in Spanish on the following topic.

Describe your activities from **yesterday**. Include what you did from the moment you woke up until you went to sleep again. **Use at least two (2) reflexive verbs, the verb "ir", and other verbs in the preterite tense.** DOUBLE SPACE YOUR RESPONSE.

Appendix F: 2019 Advanced Placement Psychology Examination Free Response Questions

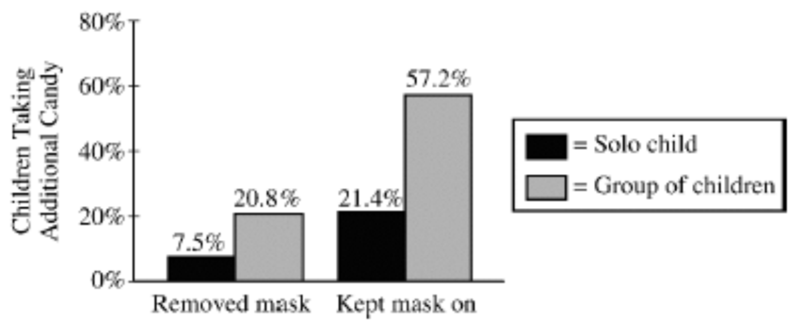
PSYCHOLOGY

SECTION II

Time—50 minutes

Directions: You have 50 minutes to answer BOTH of the following questions. It is not enough to answer a question by merely listing facts. You should present a cogent argument based on your critical analysis of the questions posed, using appropriate psychological terminology.

1.A psychologist conducted a study at her home during an annual activity of children wearing masks and going door-to-door receiving candy. Some of the children arrived alone, while others arrived in a group. Over the course of the night, the psychologist asked half of the children to remove their masks when they arrived at her door. The remaining half kept their masks on. The psychologist told every child to take only one piece of candy. She then went inside the house, leaving the bowl of candy outside. This gave children the opportunity to take additional candy. The psychologist measured the percentage of children who took additional candy. The psychologist's hypotheses were that children would take more candy when they were alone and that children would take more candy when they were masked. The results are shown in the graph below; assume all differences are significant.



A. Identify the operational definition of the dependent variable in this study.

B. Explain how the data support or do not support each of the psychologist's hypotheses.

C. Explain why the psychologist cannot generalize her findings to all children.

D. Explain why the study is not a naturalistic observation.

E. Explain how each of the following might have played a role in the children's behavior.

- Modeling
- Deindividuation
- Lawrence Kohlberg's preconventional stage

2. As a senior in high school, Ludy worked as an assistant to the children's librarian in his town library. He enjoyed shelving the books in the library because he was able to work alone and focus on his task. Although the library was normally a quiet place, sometimes the children's section became quite noisy when groups of children visited. After working in an unfulfilling desk job for 30 years, Ludy recalls how much he enjoyed working at the library and decides to go back and work as a volunteer creating new programs for disadvantaged children.

Describe how each of the following concepts relates to Ludy's return to working at the library.

- Crystallized intelligence
- Altruism
- Big Five trait of extraversion
- Broca's Area
- Self-efficacy
- Episodic memory
- Self-actualization

References

An, D., Song, Y., & Carr, M. (2016). A comparison of two models of creativity: Divergent thinking and creative expert performance. *Personality and Individual Differences*, *90*, 78–84. doi.org/10.1016/j.paid.2015.10.040

Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, *26*(2), 229–238. doi.org/10.1080/10400419.2014.901095

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227–2578. doi.org/10.1016/S0160-2896(96)90016-1

Ackerman, P. L. (2000). Domain-specific knowledge as the "dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *55*(2), 69–84.

Ackerman, P. L. (2014). Adolescent and adult intellectual development. *Current Directions in Psychological Science*, *23*(4), 246–251. doi.org/10.1177/0963721414534960

Ackerman, P. L., & Beier, M. E. (2006). Determinants of domain knowledge and independent study learning in an adult sample. *Journal of Educational Psychology*, *98*(2), 366–381. doi.org/10.1037/0022-0663.98.2.366

Ackerman, P. L., Bowen, K. R., Beier, M. E., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology*, *93*(4), 797–825. doi.org/10.1037/0022-0663.93.4.797

Ackerman, P. L., & Lohman, D. F. (2006). Individual differences in cognitive functions. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology.* (pp. 139–161). Mahwah, NJ: Lawrence Erlbaum Associates. doi.org/10.4324/9780203874790.ch7

Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P. R. Pintrich (Eds.) *Advances in motivation and achievement* (Vol. 10, pp. 213–250). Greenwich, CT: JAI.

Alexander, P. A. (2006). *Psychology in learning and instruction.* Upper Saddle River, NJ: Pearson.

Alexander, P. A. (2016). Relational thinking and relational reasoning: Harnessing the power of patterning. *npj Science of Learning, 1*, 16004. doi.org/10.1038/npjscilearn.2016.4.

Alexander, P. A. (2017). Relational reasoning in STEM domains: A foundation for academic

development. *Educational Psychology Review, 29*, 1–10. doi.org/10.1007/s10648-016-9383-1

Alexander, P. A. (2019). Individual differences in college-age learners: The importance of relational reasoning for learning and assessment in higher education. *British Journal of Educational Psychology, 89*. 416-428. doi.org/10.1111/bjep.12264

Alexander, P. A., & Baggetta, P. (2014). Percept-concept coupling and human error. In D. N. Rapp & J. L. G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences.* (pp. 297–327). Cambridge, MA: MIT Press.

Alexander, P. A., & The Disciplined Reading and Learning Research Laboratory. (2012a). Reading into the future: competence for the 21st century. *Educational Psychologist, 47*(4), 259e280. doi.org/10.1080/00461520.2012.722511

Alexander, P. A., & The Disciplined Reading and Learning Research Laboratory (2012b). *Test of Relational Reasoning*. College Park, MD: University of Maryland.

Alexander, P. A., Dumas, D., Grossnickle, E. M., List, A., & Firetto, C. M. (2016). Measuring relational reasoning. *Journal of Experimental Education*, *84*(1), 119–151. doi.org/10.1080/00220973.2014.963216

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology*, *87*(4), 559–575. doi.org/10.1037/0022-0663.87.4.559

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, *58*(4), 375–404. doi.org/10.3102/00346543058004375

Alexander, P. A., Pate, P. E., Kulikowich, J. M., Farrell, D. M., & Wright, N. L. (1989). Domain-specific and strategic knowledge: Effects of training on students of differing ages or competence levels. *Learning and Individual Differences*, *1*(3), 283–325. doi.org/10.1016/1041-6080(89)90014-9

Andiliou, A., Ramsay, C. M., Murphy, P. K., & Fast, J. (2012). Weighing opposing positions: Examining the effects of intratextual persuasive messages on students' knowledge and beliefs. *Contemporary Educational Psychology*, *37*(2), 113–127. doi.org/10.1016/j.cedpsych.2011.10.001

Baker, S. T., Friedman, O., & Leslie, A. M. (2010). The opposites task: Using general rules to test cognitive flexibility in preschoolers. *Journal of Cognition and Development, 11,* 240–254. doi.org/10.1080/ 15248371003699944

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain: A Journal of Neurology*, *135*(4), 1154–1164. doi.org/10.1093/brain/aws021

Bassok, M., Dunbar, K. N., & Holyoak, K. J. (2012). Introduction to the special section on the neural substrate of analogical reasoning and metaphor comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 261–263. doi.org/10.1037/a0026043

Baumert, J., Lüdtke, O., Trautwein, U., & Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: Evidence in support of the distinction between intelligence and student achievement. *Educational Research Review*, *4*(3), 165–176. doi.org/10.1016/j.edurev.2009.04.002

Beier, M. E., & Ackerman, P. L. (2003). Determinants of health knowledge: An investigation of age, gender, abilities, personality, and interests. *Journal of Personality and Social Psychology*, *84*(2), 439–448. doi.org/10.1037/0022-3514.84.2.439

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging*, *20*(2), 341–355. doi.org/10.1037/0882-7974.20.2.341

Bereby-Meyer, Y., & Kaplan, A. (2005). Motivational influences on transfer of problem-solving strategies. *Contemporary Educational Psychology*, *30*(1), 1–22. doi.org/10.1016/j.cedpsych.2004.06.003

Berlin, S. (1990). Dichotomous and complex thinking. *The Social Service Review*, 46-59. doi.org/10.1086/603741

Bianchi, I., Savardi, U., & Burro, R. (2011). Perceptual ratings of opposite spatial properties: Do they lie on the same dimension? *Acta Psychologica*, *138*(3), 405–418. doi.org/10.1016/j.actpsy.2011.08.003

Blakemore, S.-J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry*, *47*(3), 296–312. doi.org/10.1111/j.1469-7610.2006.01611.x

Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184–1191. doi.org/10.1038/nn.3177

Bohan, J., & Sanford, A. (2008). Semantic anomalies at the borderline of consciousness: An eye-tracking investigation. *The Quarterly Journal of Experimental Psychology*, *61*(2), 232–239. doi.org/10.1080/17470210701617219

Bolger, D. J., Mackey, A. P., Wang, M., & Grigorenko, E. L. (2014). The role and sources of individual differences in critical-analytic thinking: A capsule overview. *Educational Psychology Review*, *26*(4), 495–518. doi.org/10.1007/s10648-014-9279-x

Bostrom, A. (2008). Lead is like mercury: Risk comparisons, analogies and mental models. *Journal of Risk Research*, *11*(1–2), 99–11 doi.org/10.1080/13669870701602956

Bransford, J. D., Brown, A. L., Cocking, R. R., & National Academy of Sciences - National Research Council, W. D. (1999). *How People Learn: Brain, Mind, Experience, and School*.

Broughton, S. H., & Sinatra, G. M. (2010). Text in the science classroom: Promoting engagement to facilitate conceptual change. In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life.* (pp. 232–256). New York, NY: Guilford Press.

Broughton, S. H., Sinatra, G. M., & Reynolds, R. E. (2010). The nature of the refutation text effect: An investigation of attention allocation. *Journal of Educational Research*, *103*(6), 407–423. doi.org/10.1080/00220670903383101

Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive abilities and their interplay: Reasoning, crystallized intelligence, working memory components, and sustained attention. *Journal of Individual Differences*, *27*(2), 57–72. doi.org/10.1027/1614-0001.27.2.57

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*(3), 404–431. doi.org/10.1037/0033-295X.97.3.404

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Cattell, R.B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38*, 592.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *40*(3), 153-193. doi.org/10.1037/h0059973

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: a critical experiment. *Journal of Educational Psychology*, *54*, 1–22. doi.org/10.1037/h0046743

Cattell, R. B. (1971). *Abilities: their structure, growth, and action*. Oxford: Houghton Mifflin.

Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. Oxford: North-Holland.

Chan, J., & Schunn, C. (2015). The impact of analogies on creative concept generation: Lessons from an in vivo study in engineering design. *Cognitive Science*, *39*(1), 126–155. doi.org/10.1111/cogs.12127

Charness, N. (1987). Component processes in bridge bidding and novel problem-solving tasks. *Canadian Journal of Psychology*, *41*(2), 223–243. doi.org/10.1037/h0084151

Chase, W G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition (*pp. 141-189). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi.org/10.1016/0010-0285(73)90004-2

Chi, M. T. H. (2013, April). Thinking about relations in learning. In J. M. Kulikowich (Chair), *Exploring and leveraging relational thinking for academic performance*. Symposium conducted at the meeting of the American Educational Research Association, San Francisco, CA.

Chi, M.T. H., Glaser, R., & Rees, E. (1981). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp.7–75). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., & Hausmann, R. G. M. (2003). Do radical discoveries require ontological shifts? In L. V. Shavinina (Ed.), *The international handbook on innovation.* (pp. 430–444). New York, NY: Elsevier Science.

Chi, M. T. H., & Roscoe, R.D. (2002). The processes and challenges of conceptual change. In M. Limon & L. Mason (Eds.), *Reframing the process of conceptual change: Integrating theory and practice* (pp. 3-27). Dordrecht, The Netherlands: Kluwer Academic.

Chi, M. T. H., & Slotta, J. D. (1993). The ontological coherence of intuitive physics. *Cognition and Instruction*, *10*(2–3), 249–260. doi.org/10.1080/07370008.1985.9649011

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: A theory of conceptual change for learning science concepts. *Learning and Instruction*, *4*(1), 27–43. doi.org/10.1016/0959-4752(94)90017-5

Chinn, C. A., & Anderson, R. C. (1998). The structure of discussions that promote reasoning. *Teachers College Record*, *100*(2), 315–368.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1–49. doi.org/10.3102/00346543063001001

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218. doi.org/10.1002/sce.10001

Chinn, C.A., & Samarapungavan, A. (2009). Conceptual change—Multiple routes, multiple mechanisms: A commentary on Ohlsson (2009). *Educational Psychologist, 44*, 48–57. doi: 10.1080/00461520802616291

Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference resolution, and maintenance. *Memory & Cognition*, *35*(6), 1445–1455. doi.org/10.3758/BF03193614

Choudhury, S., Charman, T., & Blakemore, S.-J. (2008). Development of the teenage brain. *Mind, Brain, and Education*, *2*(3), 142–147. doi.org/10.1111/j.1751-228X.2008.00045.x

Christensen, B. T., & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & Cognition*, *35*(1), 29–38. doi.org/10.3758/BF03195939

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli,

J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage, 14* (5), 1136–1149.

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

Cole, M., & Wertsch, J. V. (1996). Beyond the individual-social antinomy in discussions of Piaget and Vygotsky. *Human Development*, *39*(5), 250–256. doi.org/10.1159/000278475

Compton, V. J., Compton, A., & Patterson, M. (2011). Exploring the transformational potential of technological literacy. In *Proceedings of the Joint 25th Pupils Attitude Toward Technology*(*PATT 25*) *and 8th Centre for Research in Primary Technology* (*CRIPT 8*) *Conference* (pp. 128–136). London, England: Goldsmiths.

Cosgrove, M. (1995). A study of science-in-the-making as students generate an analogy for electricity. *International Journal of Science Education*, *17*(3), 295–310. doi.org/10.1080/0950069950170303

Crone, E. A., Wendelken, C., van Leijenhorst, L., Honomichl, R. D., Christoff, K., & Bunge, S. A. (2009). Neurocognitive development of relational reasoning. *Developmental Science*, *12*(1), 55–66. doi.org/10.1111/j.1467-7687.2008.00743.

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research.* (pp. 11–43). Washington, DC: American Psychological Association.

Dinsmore, D. L., Baggetta, P., Doyle, S., & Loughlin, S. M. (2014). The role of initial learning, problem features, prior knowledge, and pattern recognition on transfer success. *Journal of Experimental Education*, *82*(1), 121–141. doi.org/10.1080/00220973.2013.835299

Dumas, D. (2017). Relational reasoning in science, medicine, and engineering. *Educational Psychology Review*, *29*(1), 73–95. doi.org/10.1007/s10648-016-9370-6

Dumas, D. (2018). Relational reasoning and divergent thinking: An examination of the threshold hypothesis with quantile regression. *Contemporary Educational Psychology*, *53*, 1–14. doi.org/10.1016/j.cedpsych.2018.01.003

Dumas, D., & Alexander, P. A. (2016). Calibration of the test of relational reasoning. *Psychological Assessment*, *28*(10), 1303–1318. doi.org/10.1037/pas0000267

Dumas, D., & Alexander, P. A. (2018). Assessing differential item functioning on the test of relational reasoning. *Frontiers in Education*. doi.org/10.3389/feduc.2018.00014

Dumas, D., Alexander, P. A., Baker, L. M., Jablansky, S., & Dunbar, K. N. (2014). Relational reasoning in medical education: Patterns in discourse and diagnosis. *Journal of Educational Psychology*, *106*(4), 1021–1035. doi.org/10.1037/a0036777

Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review*, *25*(3), 391–427. doi.org/10.1007/s10648-013-9224-4

Dumas, D., & Schmidt, L. (2015). Relational reasoning as predictor for engineering ideation success using TRIZ. *Journal of Engineering Design, 26*(1–3), 74–88. doi.org/10.1080/09544828.2015.1020287

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. *Developmental Cognitive Neuroscience*, *10*, 57-76. doi.org/10.1016/j.dcn.2014.07.009

Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, *13*(6), 15–24. doi.org/10.1111/j.1467-7687.2010.01014.x

Dunbar, K., & Fugelsang, J. (2005). Scientific Thinking and Reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning.* (pp. 705–725). New York, NY: Cambridge University Press.

Ehri, L. C., Satlow, E., & Gaskins, I. (2009). Grapho-phonemic enrichment strengthens keyword analogy instruction for struggling young readers. *Reading & Writing Quarterly*, *25*(2), 162–191. doi.org/10.1080/10573560802683549

Endsley, M. R. (2006). Expertise and situation awareness. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance., 2nd ed.* (pp. 714–741). New York, NY: Cambridge University Press.

Ennis, R. (1997). Incorporating critical thinking in the curriculum: an introduction to some basic issues. *Inquiry*, *16(3)*, 9.doi.org/10.5840/inquiryctnews199716312

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. doi.org/10.1037/0033-295X.100.3.36

Eslinger, P. J., Blair, C., Wang, J., Lipovsky, B., Realmuto, J., Baker, D., … Yang, Q. X. (2009). Developmental shifts in fMRI activations during visuospatial relational reasoning. *Brain and Cognition*, *69*(1), 1–10. doi.org/10.1016/j.bandc.2008.04.010

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review*, *31*(2), 246. doi.org/10.1002/pits.10083

Federiakin, D.A., & Aleksandrova, E. A. (2017). *Investigating the dimensionality of TORR: A replication study* (Tech Rep. WP BRP 87/PSY/2017). National Research University, Higher School of Economics.

Ferguson, H. J., & Sanford, A. J. (2008). Anomalies in real and counterfactual worlds: An eye-movement investigation. *Journal of Memory and Language*, *58*(3), 609–626. doi.org/10.1016/j.jml.2007.06.007

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40*, 935–952. doi:10.1037/ 0012-1649.40.6.935

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E.(2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*(6), 1460–1473. doi.org/10.1037/0012-1649.43.6.1460

Ferrer E, O'Hare E. D., & Bunge S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience*, *3*:46–51 doi:10.3389/neuro.01.003.2009Filik, R. (2008). Contextual override of pragmatic anomalies: Evidence from eye movements. *Cognition*, *106*(2), 1038–1046. doi.org/10.1016/j.cognition.2007.04.006

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. *Psychophysiology*, *45*(4), 554–558. doi.org/10.1111/j.1469-8986.2008.00656.x

Fischer, R. S., Norberg, A., & Lundman, B. (2008). Embracing opposites: Meanings of growing old as narrated by people aged 85. *International Journal of Aging and Human Development*, *67*(3), 259–271. doi.org/10.2190/AG.67.3.d

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, *40*(2), 155–171. doi.org/10.1002/pits.10083

Fountain, L. (2017). *Relations among topic knowledge, individual interest, and relational reasoning, and critical thinking in maternity nursing*. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. ProQuest Information & Learning.

Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological Science*, *7*(4), 237–241. doi.org/10.1111/j.1467-9280.1996.tb00366.x

Fugelsang, J. A., & Dunbar, K. N. (2005). Brain-based mechanisms underlying complex causal thinking. *Neuropsychologia*, *43*(8), 1204–1213. doi.org/10.1016/j.neuropsychologia.2004.10.012

Gardner, H. (1988). Creativity: An interdisciplinary perspective. *Creativity Research Journal, 1*, 8-26.

Gardner, H. (1995). "Multiple intelligences" as a catalyst. *English Journal*, *84*(8), 16–18. doi.org/10.2307/821182

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170. doi.org/10.1207/s15516709cog0702_3

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development*, *59*(1), 47–59. doi.org/10.2307/1130388

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. Cognitive Science, 34(5), 752–775. doi.org/10.1111/j.1551-6709.2010.01114

Gentner, D., & Landers, R. (1985). Analogical reminding: A good match is hard to find. In *Proceedings of the International Conference on Systems, Man and Cybernetics* (pp. 607–613). Tucson, Arizona

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*(1), 45–56. doi.org/10.1037/0003-066X.52.1.45

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J.P. Byrnes (Eds.), Perspectives on thought and language: Interrelations in development (pp. 225-277). London: Cambridge University Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, *39*(2), 93–104.

Goel, V. (2007). Anatomy of deductive reasoning. *Trends in Cognitive Sciences*, *11*(10), 435–441.

Goldman, S. R., & Pellegrino, J. W. (2015). Research on learning and instruction implications for curriculum, instruction, and assessment. *Policy Insights from the Behavioral and Brain Sciences, 2*(1), 33–41.

Gonthier, C., & Thomassin, N. (2015). Strategy use fully mediates the relationship between working memory capacity and performance on Raven's matrices. *Journal of Experimental Psychology: General*, *144*(5), 916–924.

Goswami, U. (1992). *Analogical reasoning in children.* Hove: Lawrence Erlbaum Associates.

Goswami, U., & Brown, A. L. (1990). Melting chocolate and melting snowmen: Analogical reasoning and causal relations. *Cognition*, *35*(1), 69–95.

Goswami, U., & Bryant, P. (1992). Rhyme, analogy, and children's reading. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 49–63). Hillsdale: Lawrence Erlbaum Associates.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.

Green, A. E., Fugelsang, J. A., & Dunbar, K. N. (2006). Automatic activation of categorical and abstract analogical relations in analogical reasoning. *Memory & Cognition*, *34*(7), 1414–1421.

Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 264–272.

Grossnickle, E. M., Dumas, D., Alexander, P. A., & Baggetta, P. (2016). Individual differences in the process of relational reasoning. *Learning and Instruction*, *42*, 141–159.

Guilford, J. P. (1967). Creativity: Yesterday, today, and tomorrow. *The Journal of Creative Behavior*, *1*(1), 3–14.

Gustafsson, J. E. (2001). Schooling and intelligence: Effects of track of study on level and profile of cognitive abilities. *International Education Journal, 2*, 166-186.

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition*, *31*(6), 902–917.

Hambrick, D.Z. (2005). The role of domain knowledge in complex cognition. In O. Wilhelm & R.W. Engle (Eds.), *Handbook of Understanding and Measuring Intelligence.* Thousand Oaks, CA: Sage Publications.

Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, *44*(4), 339–387.

Hambrick, D. Z., Pink, J. E., Meinz, E. J., Pettibone, J. C., & Oswald, F. L. (2008). The roles of ability, personality, and interests in acquiring current events knowledge: A longitudinal study. *Intelligence*, *36*(3), 261–278.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41–45.

Heit, E., & Nicholson, S. P. (2010). The opposite of republican: Polarization and political categorization. *Cognitive Science*, *34*(8), 1503–1516.

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning.* (pp. 234–259). New York, NY: Oxford University Press.

Hong, J. C., & Liu, M. C. (2003). A study on thinking strategy between experts and novices of computer games. *Computers in Human Behavior*, *19*(2), 245–258

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*(3), 242–259.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues.* (pp. 53–91). New York, NY: Guilford Press.

Houde, O., Rossi, S., Lubin, A., & Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: An fMRI meta-analysis of 52 Studies including 842 children. *Developmental Science*, *13*(6), 876–885.

Huberty, C. J, & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In H. E. A. Tinsley & S. Brown (Eds.). *Handbook of multivariate statistics and mathematical modeling*. San Diego, CA: Academic

Hummel, J. E., & Holyoak, K. J. (2005). Relational reasoning in a neurally plausible cognitive architecture: An overview of the LISA project. *Current Directions in Psychological Science*, *14*(3), 153–157.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12*, 123–129.

Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Hynd, C. (2003). Conceptual change in response to persuasive messages. In G. M. Sinatra & P. R. Pintrich (Eds.), *Intentional conceptual change* (pp. 291–315). Mahwah, NJ: Erlbaum.

Ivanova, I., Pickering, M. J., Branigan, H. P., McLean, J. F., & Costa, A. (2012). The comprehension of anomalous sentences: Evidence from structural priming. *Cognition*, *122*(2), 193–209.

Jablansky, S., Alexander, P. A., Dumas, D., & Compton, V. (2015). Developmental differences in relational reasoning among primary and secondary school students. *Journal of Educational Psychology*, *108*(4), 592–608.

Jastrzębski, J., Ciechanowska, I., & Chuderski, A. (2018). The strong link between fluid intelligence and working memory cannot be explained away by strategy use. *Intelligence*, *66*, 44–53.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jensen, A. R. (1980). *Bias in mental testing*. New York, NY: Free Press.

Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to Gf, Gc, and academic skills at ages 22–90 years? *Archives of Clinical Neuropsychology*. doi: 10.1016/j.acn. 2008.12.001

Kaya, F., Stough, L., & Juntune, J. (2016). Verbal and nonverbal intelligence scores within the context of poverty. *Gifted Education International*, 1-16. doi: 10.1177/0261429416640332

Kendeou, P., Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice*, *29*(1), 10–16.

Kim, Y., Hendrickson, A., Patel, P., Melican, G., & Sweeney, K., (2013). Development of a new ReadiStep scale linked to the PSAT/NMSQT® Scale. Research Report 2013-4. *College Board*.

Klein, J. (2004). Who is most responsible for gender differences in scholastic achievements: pupils or teachers?. *Educational Research, 46*(2), 183-193. doi: 10.1080/0013188042000222458

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition and Computation*, *9*(2), 109–137.

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S.-J. (2016). A window of opportunity for cognitive training in adolescence. *Psychological Science*, *27*(12), 1620–1631.

Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, *1428*, 13–23.

Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., … Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020–2032.

Krumm, G., Arán Filippetti, V., & Gutierrez, M. (2018). The contribution of executive functions to creativity in children: What is the role of crystallized and fluid intelligence? *Thinking Skills & Creativity*, *29*, 185–195.

Kuhn, D. (2006). Do cognitive changes accompany developments in the adolescent brain? *Perspectives on Psychological Science*, *1*(1), 59–67.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking & Reasoning*, *13*(2), 90–104.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*(1), 162–181.

Lombardi, D., Brandt, C. B., Bickel, E. S., & Burg, C. (2016). Students' evaluations about climate change. International Journal of Science Education, 38(8), 1392-1414. doi: 10.1080/09500693.2016.1193912

Lombardi, D., Nussbaum, E. M., & Sinatra, G. M. (2016). Plausibility judgments in conceptual change and epistemic cognition. Educational Psychologist, 51(1), 35-56. doi: 10.1080/00461520.2015.1113134

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, *86*(1), 96-111. doi.org/10.1037/0022-3514.86.1.96

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481–498.

Magis-Weinberg, L., Blakemore, S.-J., & Dumontheil, I. (2017). Social and nonsocial relational reasoning in adolescence and adulthood. *Journal of Cognitive Neuroscience*, *29*(10), 1739–1754.

Markman, A. B., & Gentner, D. (2001). THINKING. *Annual Review of Psychology*, *52*(1), 223.

Mason, L., Baldi, R., Danielson, R. W., & Sinatra, G. M. (2016). *Refutation text and graphics: effects on conceptual change learning*. Washington, DC: Paper presented at the annual meeting of the American Educational Research Association.

Mason, L., & Sorzio, P. (1996). Analogical reasoning in restructuring scientific knowledge. *European Journal of Psychology of Education*, *11*(1), 3–23.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, tests, and Issues.* (pp. 151–191). New York, NY: Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues.* (pp. 136–181). New York, NY: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10.

McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, *47*(7), 651–675.

McPherson, S. L., & MacMahon, C. (2008). How baseball players prepare to bat: Tactical knowledge as a mediator of expert performance in baseball. *Journal of Sport & Exercise Psychology*, *30*(6), 755–778.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage Publication.

Miller-Singley, A.T., & Bunge, S.A. (2014). Neurodevelopment of relational reasoning:
Implications for mathematical pedagogy. *Trends in Neuroscience and Education, 3*:33–37.

Mischo C., & Maaß, K. (2012). Which personal factors affect mathematical modelling? The effect of abilities, domain specific and cross domain-competences and beliefs on performance in mathematical modelling. *Journal of Mathematical Modelling and Application, 1*, 3–19.

Moehring, A., Schroeders, U., & Wilhelm, O. (2018). Knowledge is power for medical assistants: Crystallized and fluid intelligence as predictors of vocational knowledge. *Frontiers in Psychology*, *9*.

Mosenthal, P. B. (1988). Three approaches to progress: Understanding the writing of exceptional children. *Exceptional Children*, *54*(6), 497–504.

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *Journal of Experimental Education*, *70*(3), 197–214.

Murphy, P. K., Firetto, C. M., & Greene, J. A. (2017). Enriching students' scientific thinking through relational reasoning: Seeking evidence in texts, tasks, and talk. *Educational Psychology Review*, *29*(1), 105–117.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., … Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101.

Nersessian, N. J., & Chandrasekharan, S. (2009). Hybrid analogies in conceptual innovation in science. *Cognitive Systems Research*, *10*(3), 178–188.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510–520.

Nunes, T., Bryant, P., Barros, R., & Sylva, K. (2012). The relative importance of two different mathematical abilities to mathematical achievement. *British Journal of Educational Psychology*, *82*(1), 136–156.

Nussbaum, M. E., & Kardash, C. A. M. (2005). The effects of goal instructions and text on the generation of counterarguments during writing. *Journal of Educational Psychology*, *97*(2), 157–169.

Orzechowski, J. (2010). Working memory capacity and individual differences in higher-level cognition. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition: Attention, Memory, and Executive Control.* (pp. 353–368). New York, NY: Springer Science + Business Media.

Osborne, J. W. (2013). *Best practices in data cleaning.* Los Angeles: Sage.

Petrides, K. V., Chamorro-Premuzic, T., Frederickson, N., & Furnham, A. (2005). Explaining individual differences in scholastic behaviour and achievement. *British Journal of Educational Psychology*, *75*(2), 239–255.

Postlethwaite, B. E. (2012). *Fluid ability, crystallized ability, and performance across multiple domains: A meta-analysis*. *Dissertation Abstracts International Section A: Humanities and Social Sciences*. ProQuest Information & Learning.

Proctor, T. P., Wyatt, J., Wiley, A. (2010). *PSAT/NMSQT® Indicators of College Readiness. Research Report No. 2010-4*. *College Board*. College Board.

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relating personality and creativity: Considering what and how we measure. *The Journal of Creative Behavior*, 1-14.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London: H.K. Lewis

Raven, J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology, 19*(1), 137–150

Raven, J. C., Raven, J., & Court, J. H. (1998). *A manual for Raven's progressive matrices and vocabulary scales*. London, UK: H. K. Lewis.

Reilly, D., Neumann, D. L., & Andrews, G. (2017). Gender differences in spatial ability: Implications for STEM education and approaches to reducing the gender gap for parents and educators. In M. S. Khine (Ed.), *Visual-Spatial Ability: Transforming Research into Practice* (pp. 195-224). Switzerland: Springer International. doi: 10.1007/978-3-319-44385-0_10

Ren, X., Schweizer, K., Wang, T., & Xu, F. (2015). The prediction of students' academic performance with fluid intelligence in giving special consideration to the contribution of learning. *Advances in Cognitive Psychology*, *11*(3), 97-105. doi:10.5709/acp-0175-z

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science*, 1-6. doi: 10.1177/0956797612450883.

Richland, L. & McDonough, I. (2010). Learning by analogy: discriminating between potential analogs, *Contemporary Educational Psychology, 35*(1), 28–43.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94*, 249–273.

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*, 177–192.

Rindermann, H., Flores-Mendoza, C., & Mansur-Alves, M. (2010). Reciprocal effects between fluid and crystallized intelligence and their dependence on parents' socioeconomic status and education. *Learning and Individual Differences, 20,* 544–548. doi.org/10.1016/j.lindif .2010.07.002

Rolfhus, E. L., & Ackerman, P. L. (1999). Assessing individual differences in knowledge: Knowledge, intelligence, and related traits. *Journal of Educational Psychology, 91*(3), 511–526.

Saccuzzo, D. P., Johnson, N. E., & Russell, G. (1992). Verbal versus performance IQs for gifted African-American, Caucasian, Filipino, and Hispanic children. *Psychological Assessment, 4*, 239-244. doi:10.1037//1040-3590.4.2.239

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2015). On the nature of crystallized intelligence: The relationship between verbal ability and factual knowledge. *Intelligence*, *46*, 156–168. doi.org/10.1016/j.intell.2014.05.014

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, & P. Harrison (Eds.). *Contemporary Intellectual Assessment* (pp. 99–144). (3rd ed.). New York, NY: Guilford Press.

Schrank, F.A., & Wendling, B.J. (2009). *Woodcock-Johnson III Assessment Service Bulletin No*

10. Rolling Meadows, IL: Riverside Publishing.

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2015). Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. *Intelligence*, *48*, 15–29. doi.org/10.1016/j.intell.2014.10.006

Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, *59*, 84–95.

Schulz, L. E., Goodman, N. D., Tenenbaum, J. B., & Jenkins, A. C. (2008). Going beyond the evidence: abstract laws and preschoolers' responses to anomalous data. *Cognition, 109*(2), 211–223..doi:10.1016/ j.cognition.2008.07.017.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*(6), 771–799.

Shipstead, Z., & Yonehiro, J. (2016). The domain-specific and domain- general relationships of visuospatial working memory to reasoning ability. *Psychonomic Bulletin & Review, 23*, 1504–1512.

Shuwairi, S. M., Bainbridge, R., & Murphy, G. L. (2014). Concept formation and categorization of complex, asymmetric, and impossible figures. *Attention, Perception, & Psychophysics, 76,* 1789–1802. doi.org/10.3758/s13414-014-0691-6

Singer, M., & Gagnon, N. (1999). Detecting causal inconsistencies in scientific text. In S. R. Goldman, A. C. Graesser, & P. van den Broek (Eds.), *Narrative Comprehension, Causality, and Coherence: Essays in Honor of Tom Trabasso.* (pp. 179–194). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Singley, A. T. M., & Bunge, S. A. (2014). Neurodevelopment of relational reasoning: Implications for mathematical pedagogy. *Trends in Neuroscience and Education*, *3*(2), 33-37.

Slotta, J. D., & Chi, M. T. H. (2006). Helping students understand challenging topics in science through ontology training. *Cognition and Instruction*, *24*(2), 261–289.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 2, pp. 47-103). Hillsdale, NJ: Erlbaum.

Sorensen, R. A. (2003). *A brief history of the paradox: philosophy and the Labyrinths of the mind*. New York: Oxford University Press.

Sorenson, H., & Yankech, R. (2008). Precepting in the fast lane: Improving critical thinking in new graduate nurses. *The Journal of Continuing Education in Nursing, 39,* 208-216.

Sprenger, A.M., Atkins, S.M., Bolger, D.J., Harbison, J.I., Novick, J.M., Chrabaszcz, J.S., et al. (2013). Training working memory: Limits of transfer. *Intelligence, 41*, 638–663. doi.org/10.1016/j.intell.2013.07.013.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*(2), 342–357.

Starr, A., Vendetti, M. S., & Bunge, S. A. (2018). Eye movements provide insight into individual differences in children's analogical reasoning strategies. *Acta Psychologica*, *186*, 18–26.

Stephens, A. C. (2006). Equivalence and relational thinking: preservice elementary teachers' awareness of opportunities and misconceptions. *Journal of Mathematics Teacher Education, 9*(3), 249–278. doi:10.1007/s10857-006-9000-1.

Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological Review, 84,* 353–378. doi.org/10.1037/0033-295X.84.4.353

Sternberg, R. J. (1988). *The triarchic mind: A new theory of human intelligence*. Penguin Books, New York.

Sternberg, R. J. (2009). Component processes in analogical reasoning. In J. C. Kaufman & E. L. Grigorenko (Eds.), *The essential Sternberg: Essays on intelligence, psychology, and education.* (pp. 145–179). New York, NY: Springer Publishing Co.

Tabatabaee-Yazdi, M., & Baghaei, P. (2018). Reading comprehension in English as a foreign language and some Cattell-Horn-Carroll cognitive ability factors. *Reading Matrix*, *18,* 19–26.

Taub, G., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement from kindergarten through high school. *School Psychology Quarterly, 23*(2), 187−198.

Thibaut, J. P., & French, R. M. (2016). Analogical reasoning, control and executive functions: a developmental investigation with eye-tracking. *Cognitive Development, 38*, 10–26. doi: 10.1016/j.cogdev.2015.12.002

Thorsen, C., Gustafsson, J., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *British Journal of Educational Psychology*, *84*(4), 556–570.

Tippett, C.D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education, 8*(6), 951–970.

Torrance, E. P. (1972). Predictive validity of the Torrance Tests of Creative Thinking. *Journal of Creative Behavior, 6*(4), 236–252.

Trey, L., & Khan, S. (2008). How science students can learn about unobservable phenomena using computer-based analogies. *Computers & Education, 51*(2), 519–529.

Trickett, S. B., Trafton, J. G., & Schunn, C. D. (2009). How do scientists respond to anomalies? Different strategies used in basic and applied science. *Topics in Cognitive Science, 1*, 711–729. doi: 10.1111/j.1756-8765.2009.01036.x

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, *32*(1–2), 83–98.

Vendetti, M. S., Johnson, E. L., Lemos, C. J., & Bunge, S. A. (2015). Hemispheric differences in relational reasoning: Novel insights based on an old technique. *Frontiers in Human Neuroscience*, *9*.

Vendetti, M. S., Matlen, B. J., Richland, L. E., & Bunge, S. A. (2015). Analogical reasoning in the classroom: Insights from cognitive science. *Mind, Brain, and Education*, *9*, 100–106.

Vosniadou, S. (1999). Conceptual change research: State of the art and future directions. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New Perspectives on Conceptual Change* (pp. 3–13). Oxford, UK: Elsevier.

Vosniadou, S. (2002). On the nature of naïve physics In M. Limon and L. Mason (Eds.), *Reconsidering the processes of conceptual change* (pp. 61–76). Dordrecht: Kluwer.

Waltz, J. A., Lau, A., Grewal, S. K., & Holyoak, K. J. (2000). The role of working memory in analogical mapping. *Memory & Cognition, 28,* 1205–1212.

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology, 45*(6), 920–925. doi: 10.1111/j.1469-8986.2008.00691.x

Wendelken, C., Nakhabako, N., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to??" Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience, 20*(4), 682–693. doi: 10.1162/jocn.2008.20055

Wentzel, K. R., Baker, S. A., & Russell, S. L. (2012). Young adolescents' perceptions of teachers' and peers' goals as predictors of social and academic goal pursuit. *Applied Psychology: An International Review*, *61*(4), 605–633. doi:10.1111/j.1464-0597.2012.00508.x

Wentzel, K. R., & Wigfield, A. (1998). Academic and social motivational influences on students' academic performance. *Educational Psychology Review, 10,* 155–175.

White, C. S., & Alexander, P. A. (1986). Effects of training on four-year-olds' ability to solve geometric analogy problems. *Cognition and Instruction*, *3*(3), 261–268.

Wiest, D. J., Wong, E. H., & Kreil, D. A. (1998). Predictors of global self-worth and academic performance among regular education, learning disabled, and continuation high school students. *Adolescence, 33*(131), 601-18.

Williams, A. M., and Ericsson, K. A. (2005). Perceptual-cognitive expertise in sport: some considerations when applying the expert performance approach. *Hum. Mov. Sci.* 24, 283–307. doi: 10.1016/j.humov.2005.06.002

Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock Johnson III*. Itasca, IL: Riverside.

Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, *8*, 1–8.

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: why variations in bilingual experiences matter. *Frontiers in Psychology, 6*.