The Dimensionality of Science Achievement and its Links to Other Academic Domains

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Arts in the Graduate School of The Ohio State University

By

Sarah L. Lukowski

Graduate Program in Psychology

The Ohio State University

2014

Master's Examination Committee:

Stephen A. Petrill, Advisor

John E. Opfer

Vladimir M. Sloutsky
Abstract

Broadly, science achievement is typically broken into domain-specific science content knowledge and domain-general scientific reasoning skills. However, whether this dichotomy is useful and how science achievement is related to skills from other academic domains remains poorly understood. Thus, the present study examined science content knowledge and scientific reasoning alongside measures of mathematics, visuo-spatial, and reading skills to assess the dimensionality of science and examine the relationships among these academically-relevant skills. 100 undergraduates completed a 90 minute test battery. Factor analyses suggested that the science measures were not unidimensional, though a clear multi-factor structure failed to emerge from either measure. Nevertheless, both science measures were moderately correlated with reading, math content knowledge, and visuo-spatial skills. There were no significant differences between content knowledge and reasoning in their relation to other academic domains when math-related items of the reasoning measure were accounted for. Future work should refine measures of science as a multidimensional scale in order to better elucidate the causes and consequences of individual differences in science-related skills.
Acknowledgments

I would like to express my deepest gratitude to my advisor, Dr. Stephen Petrill, for allowing me the intellectual freedom to guide this project while offering support, encouragement, and timely suggestions at every step of the way. I would also like to thank my committee members, Dr. John Opfer and Dr. Vladimir Sloutsky, for always challenging me to think more theoretically in regard to psychological constructs. The efforts of the entire Developmental area have allowed me to grow into a better researcher and for that I am grateful.

In addition, I would like to thank all my teachers and mentors that have supported me throughout life, all the way from elementary to today. And, to my family for always encouraging my intellectual curiosity and supporting all my endeavors, thank you.
Vita

June 2008 .......................................................Pennfield High School & The Battle Creek Area Mathematics and Science Center

May 2012 .......................................................BA, Psychology with Honors, Case Western Reserve University

August 2012 to present  .........................Graduate Student, Department of Psychology, The Ohio State University

April 2014 to present  .........................National Science Foundation Graduate Research Fellow

Publications


Fields of Study

Major Field: Psychology
Table of Contents

Abstract ............................................................................................................................... ii
Acknowledgments .............................................................................................................. iii
Vita ..................................................................................................................................... iv
List of Tables .................................................................................................................... vii
List of Figures .................................................................................................................. viii
Chapter 1: Introduction ....................................................................................................... 1
Chapter 2: Method ............................................................................................................ 16
Chapter 3: Results ............................................................................................................. 20
Chapter 4: Discussion ....................................................................................................... 26
References ......................................................................................................................... 33
Appendix A: Tables and Figures ...................................................................................... 43
Appendix B: Science Content Measure ............................................................................ 55
List of Tables

Table 1. Differences between native and non-native English speakers ............................ 43
Table 2. NAEP items included in the science content scale ............................................. 44
Table 3. Descriptive statistics ........................................................................................... 45
Table 4. Observed correlations between cognitive skills .................................................. 46
Table 5. Model fit statistics ............................................................................................... 47
List of Figures

Figure 1. Scree plot – Science content............................................................................. 48
Figure 2. Scree plot – Science reasoning.......................................................................... 49
Figure 3. Model 1: General factor model.......................................................................... 50
Figure 4. Model 2: All separate cognitive factors............................................................. 51
Figure 5. Model 3: Science and math combined.............................................................. 52
Figure 6. Model 4: Math and space combined.................................................................. 53
Figure 7. Model 5: Science, math, and space combined................................................... 54
Chapter 1: Introduction

Importance of the problem

One of the greatest challenges facing society today is the training of students in science, technology, engineering, and mathematics (STEM) fields to meet the demands of a technically skilled labor force (United States Department of Labor, 2007). The problem is far from being a new one. In 1945, at the close of WWII, Philip G. Johnson, President of the National Science Teachers Association noted “The public generally must understand the sciences with sufficient clarity to judge what should be done… [and recognize the need for] more science instruction with its accompanying possibilities for human welfare and security” (NSTA, 1945). Over the past 70 years levels of science education within the general public have increased dramatically (DeBoer 2000; Seymour 2002). However, within the U.S. education system, whereas common core standards have been developed and adopted by most states in English language arts and mathematics, no such common set of standards have been adopted for science education. Two critical steps in advancing science education remain: understanding how to measure science and understanding the extent to which achievement in the sciences is associated with achievement in other academic domains such as math, visuo-spatial, and reading skills. The present study addresses both of these issues with implications for future educational and research practices.
What is science?

An important obstacle in studying scientific development is defining what constitutes science. Typically, science is defined in terms of two core features: content and process (Klahr, Zimmerman, & Jirout, 2011), which highlights two important but potentially separable aspects of science. First, science is a body of knowledge – the facts describing the natural world. For example, one might express a level of scientific knowledge by noting “plants convert carbon dioxide into food during a process called photosynthesis.” However, science is more than just a collection of facts, science also implies a process – it is a systematic way of testing and evaluating hypotheses, evidence, and explanations. For example, one aspect of participating in the process of science is controlling variables properly in experimentation (e.g. Chen & Klahr, 1999). Within the literature examining the cognitive underpinnings of science and science development, the term scientific thinking is often used (Kuhn, 2010; Klahr, Zimmerman, & Jirout, 2011; Lehrer & Schauble, 2006; Zimmerman, 2000). However, this literature can also be broken into two broad categories of research, one that examines science as structures of knowledge and changes in these structures and another line of research that examines science as a reasoning process. These two broad lines of research are worth examining further.

Science as a body of knowledge

Science content knowledge refers to one’s knowledge of scientific phenomena and concepts. As such, content knowledge is often thought of as being domain-specific. Carey (1985) is a seminal piece of work in this way of thinking, suggesting that scientific
concepts are restructured as knowledge accumulates through schooling or other relevant learning experiences. Even though the vast majority of researchers in the area of scientific thinking do so through a domain-general approach discussed in the next section, very few, if any, would argue that there is not knowledge specific to the sciences.

Though clear that scientific knowledge is an area of content to be learned, there is debate as to whether there are meaningful sub-domains within science content knowledge, such as biology, physics, and chemistry. The National Research Council found that the median disattenuated correlations between scores on life, physical, and earth science sections of the 1996 National Assessment of Educational Progress (NAEP) Science Assessment ranged 0.91 – 0.99, suggesting that questions were all tapping a similar science construct (National Research Council, 2000). However, there are also known ways in which fields of science differ. For one, each field of science refers to different parts of scientific knowledge. Biology is the study of living things, whereas chemistry is the study of matter and the changes it undergoes and physics is the study of matter, energy, and their interactions. Thus, having knowledge of how genetic material is inherited from parent to offspring does not necessarily predict knowledge of gravitational forces acting on falling objects. Stemming from the debate as to whether science knowledge is a single construct or rather separate fields are differences in the ways content knowledge is assessed.

One way to measure content knowledge is to assess one particular aspect of science knowledge. These types of tasks are often developed by science educators, with the purpose of being used to assess levels of student knowledge before and after direct
instruction in a particular content area. For example, the Brief Electricity and Magnetism Assessment (Ding, Chabay, Sherwood, & Beichner, 2006) is used within physics to measure electrical and magnetism concept knowledge. Similarly, the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992) assesses knowledge of Newtonian physics and the Conceptual Inventory of Natural Selection assesses knowledge of evolution (Anderson, Fisher, & Norman, 2002). These are only three examples of numerous concept inventories available for use in the sciences.

In contrast, there are also assessments that attempt to assess science knowledge in a broad sense. Tests such as the NAEP Science Assessment and the Trends in International Mathematics and Science Study (TIMSS) Science Assessment take this approach. Each test results in an overall achievement score, but their frameworks are also broken into content topics and subtopics; treating science knowledge as a multidimensional scale. The most recently released NAEP science framework contains three broad content areas: physical science, life science, and earth/space science (National Assessment Governing Board, 2008). Moreover, the NAEP operates under the assertion that crosscutting content, defined as content knowledge that spans across two or more of physical, biological, or earth sciences, may be particularly important for identifying greater depth of science knowledge. For example, “the theory of plate tectonics and the evolution of Earth’s surface are inextricably linked with environmental pressures (such as geographic barriers), speciation, and the evolution of life” (National Assessment Governing Board, 2008), is one circumstance the NAEP framework provides, illustrating the ways in which content knowledge between fields of science
interact. NAEP subscales are highly correlated (National Research Council, 2000). Therefore, though the measurement of science content knowledge is treated as a multidimensional scale, close relations among scientific fields may also support the notion of science content knowledge as a unitary factor. Similarly, TIMSS-2011 also assesses science in three categories: life science, physical science, and earth science, at the 4th grade level, and four areas: biology, chemistry, physics, and earth science at the 8th grade level (Mullis, Martin, Ruddock, O’Sullivan, & Preuschoff, 2009). The same framework is planned for the 2015 assessment (Mullis & Martin, 2013).

Science as a process of knowing

In comparison to the content facts of the body of scientific knowledge, scientific reasoning ability refers to one’s ability to utilize methods of scientific investigation. Scientific reasoning processes have received considerably more attention in the literature compared to content knowledge. Unlike content knowledge, science reasoning processes are typically characterized as domain-general (Zimmerman, 2000). This claim is based on the assertion that science inquiry seems to follow general problem-solving rules or strategies that could be transferred to reasoning situations in a variety of domains.

Scientific reasoning comprises numerous reasoning processes including control of variables (Chen & Klahr, 1999; Inhelder & Piaget, 1958), evaluation of hypotheses, data, and scientific arguments (Koslowski, Marasia, Chelenza, & Dublin, 2008; Kuhn, 2010; Kuhn & Pease, 2008), as well as general deductive and inductive reasoning processes (for review, see Dunbar & Klahr, 2012). Tests of science reasoning may also include items that on the surface seem to be more mathematical reasoning than purely scientific
reasoning, including reasoning about proportions and probabilities (Lawson 1979; Roadrangka, Yeany, & Padilla, 1983). Inhelder & Piaget (1958) suggested that scientific reasoning was not possible until adolescence, which had a substantial effect on early research on science achievement (for review, see Lawson, 1985). However, more recent work has suggested younger children do have basic science reasoning competencies, though in many cases their use of science reasoning strategies are less sophisticated than adults’ reasoning capabilities (Koerber, Sodian, Thoermer, & Nett, 2005; Morris, Croker, Masnick, & Zimmerman, 2012; Siegler & Alibali, 2005).

Population-based tests have also been developed to assess science reasoning skills. The Programme for International Student Assessment (PISA) has focused its testing of science achievement almost exclusively on scientific reasoning skills. The framework for the PISA science assessment focuses on science competencies that are “grounded in logic, reasoning, and critical analysis” (OECD, 2013). In comparison, TIMSS and NAEP assess science reasoning to a lesser extent. TIMSS is designed to have approximately 20% reasoning-based questions in 4th grade and 30% in 8th grade (Mullis, Martin, Ruddock, O’Sullivan, & Preuschoff, 2009). Likewise, the NAEP contained Practical Reasoning and Scientific Investigation in its 1996-2005 framework, and from 2009 on has included Using Scientific Inquiry as a key science practice to be assessed (National Assessment Governing Board, 2004, 2008, 2010).

**Interactions between scientific content knowledge and reasoning**

As much as science as a body of knowledge and science as a process of knowing are often separate standards within education, and characterized by separate literatures of
research, there are notable ways in which content knowledge and scientific reasoning may interact. For one, domain knowledge is known to shape the reasoning process within a scientific context. For example, Klahr, Fay, & Dunbar (1993), as well as Penner & Klahr (1996) showed how participants’ conceptual content knowledge related to differences in experimentation and interpretation of experimental results. Secondly, the process of doing science may build more coherent and factually correct scientific knowledge (Schauble, 1996). Moreover, science reasoning processes may lead to the discovery of new science concepts (Dunbar, 1993; Klahr, 2000; Koslowski, 1996; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Lehrer & Schauble, 2006). Thus, while they are often discussed separately, the distinctiveness of science content knowledge and scientific reasoning skills remains unclear.

**Scientific achievement and other academic skills in early adulthood**

Moving beyond how to best define and measure science as a construct, achievement in the sciences may also be related to skills in other academic domains. In general, most research relating achievement in the sciences to other academic domains has been specific to a particular area within the sciences (i.e. chemistry, physics, biology) and much of it has been carried out by science educators in their particular field of interest (e.g. Capizzo, Nuzzo, & Zarcone, 2006; Tai, Sadler, & Loehr, 2005; Wagner, Sasser, & DiBiase, 2002). These studies have used an array of mathematics, visuo-spatial, reading, and other relevant cognitive skills, often using scores from college entrance exams to predict achievement in a particular area of science. The following
sections discuss findings in regard to mathematics, visuo-spatial, and reading skills separately.

**Association with math skills**

The role of mathematics skill has been perhaps most extensively studied for the impact it has in physics achievement (Blumenthal, 1961; Bolte 1966; Capizzo, Nuzzo, & Zarcone, 2006; Champagne & Klopfer; 1982; Champagne, Klopfer, & Anderson, 1980; Cilliers, Kruger, Basson, & Kirschner 1996; Cohen, Hillman, & Agne, 1978, Colette, Phillips, & Steinert, 2007; Griffith, 1985; Harlow, Harrison, & Meyertholen, 2014; Hart & Cottle, 1993; Hudson, 1986; Hudson & Liberman, 1982; Hudson & McIntire, 1977; Hudson & Rottmann, 1981; McCammon, Golden, & Wuensch, 1988; Meltzer, 2002; Sadler & Tai, 2001). Most of these studies have been carried out by physics educators within their universities, and aim to examine the role of prior mathematics skill on course grades, final exam scores, course completion, or performance on specific physics-related inventories. Overall, there is a moderate and consistent relationship between mathematics and physics course grades, whether mathematics was assessed with the quantitative section of the SAT (Cohen, Hillman, & Agne, 1978), other college entrance examinations (Blumenthal, 1961), or investigator-developed measures of mathematical skills (Champagne, Klopfer, & Anderson, 1980; Hudson & Liberman, 1982; McCammon, Golden, & Wuensch, 1988). Research investigating the role of math in growth of physics knowledge is less consistent. Whereas Meltzer (2002) and Coletta, Phillips, & Steinert (2007) independently found a significant relationship between math achievement and learning gains in introductory physics, Capizzo, Nuzzo, & Zarcone (2006) failed to find
the same relationship. There is also debate as to the role of mathematics once variance in scientific skills is accounted for. Champagne & Klopfer (1982) found a mathematics factor predicted exam scores independently from prior knowledge of Newtonian physics. In contrast, Griffith (1985) suggested that mathematics was not a significant predictor of course performance after controlling for scientific reasoning skills.

Similar studies have been carried out in chemistry (Andrews & Andrews, 1979; Coley, 1973; House, 1995; Kunhart, Olsen, & Gammons 1958; Leopold & Edgar, 2008; Lewis & Lewis, 2007; Oszogomonyan & Loftus, 1979; Park & Choi, 2013; Lopez, Shavelson, Nandagopal, Szu, & Penn, 2014; Pickering 1975; Schlar, Cluff, & Roth, 1963; Scott, 2012; Sieveking & Larson, 1969; Spencer, 1996; Tai, Sadler, & Loeher, 2005; Wagner, Sasser, & DiBiase, 2002; Xu, Villafane, & Lewis, 2013) and to a lesser extent in biology (Johnsten, 1967, Marsh & Anderson, 1989). These studies have often relied on the quantitative section of the SAT, showing that mathematics is consistently moderately related to course grades (Andrews & Andrews, 1979; Lewis & Lewis, 2007; Oszogomonyan & Loftus, 1979; Pickering, 1979; Spencer, 1996; Tai, Sadler, & Loeher, 2005). However, like in physics, it is unclear within these domains if mathematics contributes to the prediction of science outcomes when accounting for prior science knowledge, with some finding no significant effects (Coley, 1973), while others suggest mathematics independently predicts performance (Xu, Villafane, & Lewis, 2013).

Moreover, research with highly gifted students demonstrated that even within the top 1% of mathematical ability, those in the top quarter of the sample tend to take slightly more science coursework in school and score higher on science achievement tests in
biology, chemistry, and physics in high school than those in the bottom quarter (Benbow, 1992). These students also go on to more advanced science education, participate in more math/science contests, win more awards in math/science, work on research projects more frequently, and possess careers in math or science more often than those in the bottom quarter. Therefore, across a range of contexts and abilities, there is a consistent relationship between mathematics skills and science outcomes.

**Association with visuo-spatial skills**

While the literature suggests a moderate relationship between math and science skills, previous research has also suggested a close relationship between mathematics and visuo-spatial skills (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Mazzocco, Singh Bhatia, & Lesniak-Karpiak, 2006, Reuhkala, 2001, Tosto et al., 2014). Therefore, one might expect evidence for a link between visuo-spatial skills and science achievement. However, findings regarding the correlation between visuo-spatial skills and science achievement have been somewhat mixed. Piburn (1980), using two tasks from the ETS test kit found that Surface Development, but not Card Rotation significantly correlated with a science examination score in children. Kozhevnikov, Hegarty, & Mayer (2002) suggest these differences in whether a spatial task correlates with science may be due to the speeded nature of some visuo-spatial tasks. In college students, Carter, LaRussa, & Bodner (1987) found spatial skills to be significantly correlated with chemistry performance. In addition, their data had a trend which suggested spatial skills may be more highly correlated with chemistry problem solving than rote memory, a finding that may suggest spatial skills are more closely related to
science reasoning than science content knowledge. Similarly, Kozhevnikov & Thornton (2006) showed that spatial visualization skill was related to individual differences in understanding of mechanics prior to instruction, but not after a lecture series related to mechanics. Moreover, in population-based testing, Hamilton (1998) showed that the National Educational Longitudinal Study 1988 science assessment items loaded onto three factors, one of which was a spatial-mechanical reasoning factor. She also demonstrated that differences on questions loaded on this spatial-mechanical factor were primarily responsible for gender differences in the total score.

Beyond success in science courses, visuo-spatial skills may also be related to course choices. Lord & Nicely (1997) showed that 3rd graders with higher levels of spatial skills were more likely to endorse math or science as their favorite class in school. Similarly, Shea, Lubinski, & Benbow (2001) found that among gifted students, those with stronger spatial skills tended to favor science and mathematics courses in high school. However, when considering conferred undergraduate degrees among the same students, spatial ability was higher in those in the physical sciences, but not in biology, where verbal ability tended to be relatively higher.

Association with reading skills

Compared to links with math and visuo-spatial skills, the literature on the links between reading and science achievement is less developed. This is perhaps in part because early results suggested verbal scores did not predict physics grades (Cohen, Hillman, & Agne, 1978), the field of science that was the focus of much of the early research. In addition, the relative importance of mathematics as compared to verbal
abilities may have resonated with educators’ view of what skills seemed primarily important for success in the sciences. Despite this, there is a developing literature that suggests reading may also be related to science achievement. Lewis & Lewis (2007) demonstrated that verbal SAT scores were significantly correlated with chemistry final exam scores. Likewise, Tai, Sadler, & Loehr (2005) suggested that SAT verbal scores were a small but significant predictor of chemistry grades when accounting for numerous other related factors, including demographic variables, mathematics, and high school academic experiences. Moreover, results from PISA show a strong association between scientific literacy and reading literacy (Cromley, 2009).

Particularly influential is the work of O’Reilly & McNamara (2007), which examined reading skill and science knowledge in relation to science grades and state test scores. The authors demonstrated that reading skill helped students compensate for lower levels of science knowledge, but also gave a boost to students with high levels of science knowledge. Similarly, Pyburn, Pazicni, Benassi, & Tappin (2013) showed that reading comprehension skills were related to chemistry performance and could partially compensate for lower levels of prior knowledge in collegiate introductory chemistry courses. Therefore, while literature supporting a link between reading skills and science is generally more recent than investigations that include math or visuo-spatial skills, there too appears to be a moderate relationship between reading and science outcomes.

**Purpose of the present study & hypotheses**

Building upon the previous literature, the purpose of the present study was to examine both science content knowledge and science reasoning skill using a more
holistic test battery than has been previously assessed. The test battery was assembled to assess science, mathematics, visuo-spatial, and reading skills in one session. The present study works to advance understanding of the dimensionality of science achievement and its relation to other academic domains through two specific aims:

**Specific Aim 1: Examine the internal factor structure of science assessments.**

In measurement, single items are noisy while summed scores act to increase the ability to detect signal amid noise. However, a summed score has numerous assumptions about the item level data used to calculate the sum. Importantly, a summed total score implicitly assumes a unidimensional underlying factor structure of the items. Given that there are numerous subdomains that make up both science content knowledge and science reasoning skills, a group of questions assessing each could be multidimensional. Thus, factor analyses of each science achievement measure were assessed to better understand whether that assumption held. It was hypothesized that a first factor might account for a sufficient portion of the variance, but since the tests were multifaceted measures of both constructs, there may be a discernable multi-factor structure to each of the measures.

**Specific Aim 2: Examine the relationships between individual differences in science content knowledge and science reasoning skill in relation to mathematics, visuo-spatial, and reading skills.**

Compared to earlier investigations, the present study sought to combine the assessment of science, mathematics, visuo-spatial, and reading skills into one test battery. Importantly, the relationship between science content knowledge versus science
reasoning was examined in relation to these academic skills. It was hypothesized that
despite all outcomes likely being positively correlated, as success in one academic
domain often supports achievement in others, if content knowledge is domain-specific
then it would not be expected to be as closely related to mathematics and visuo-spatial
skills. However, reading may still be correlated with content knowledge, because the
content knowledge task requires reading prompts, so more advanced reading
comprehension or even decoding skills may support one’s ability to complete science
content knowledge tasks. Conversely, if science reasoning skill is domain-general, then it
would be expected to correlate more generally with mathematics and visuo-spatial skills,
especially in regard to more fluid tasks, such as those assessing visuo-spatial sketchpad
capacity. The science reasoning tasks also requires reading and comprehending the
prompts, so reading skills may also correlate with science reasoning skills.

Relations among science, math, visuo-spatial skills and reading were examined
further using structural equation modeling. To the extent that science content knowledge
and science reasoning skill are more similar to each other than they are to mathematics,
visuo-spatial, or reading skills, they would be expected to make an interpretable factor
when assessing the relationships among academic domains. Five different models were
conceptualized based on the previous literature to be fit using structural equation
modeling. Model 1 examined if a general cognitive factor best fit the data. The other
extreme of a general factor being sufficient, was that each of science, math, visuo-spatial,
and reading skills form their own separable factors. In addition, because of the
relationships noted between math and science, a third model was fit with math and
science scores loading on a single factor, while leaving visuo-spatial and reading skills separate. Alternatively, as mathematics and visuo-spatial skills are known to be closely related, these measures may make a factor, while science and reading remain separate, tested in the fourth model. The final model took the logic that if science and mathematics were closely related, and the same was true for mathematics and visuo-spatial skills, then all three might be combined into a single factor opposite of reading skills. Despite these varying possibilities, it was expected that if each of these domains is meaningfully separate and measures within each domain of science, math, visuo-spatial, and reading skills are more similar within domain than across, then a factor structure consistent with Model 2 would be supported by the data.
Chapter 2: Method

Sample

Participants (N=124) were recruited from the Research Experience Program at The Ohio State University, a program in which students were given the choice of participating in psychological research as part of the Introductory Psychology course. Participants received course credit for their participation. Participants completed a 90 minute battery which assessed a broad range of science, mathematics, visuo-spatial, and reading skills. 5 participants were not included in analyses due to a failure to complete over half the test battery or other patterns of non-response. 19 participants indicated on a questionnaire that they were non-native speakers of English. Analyses (see Table 1) indicated several significant differences which would potentially affect the appropriateness of the tasks administered and the interpretation of findings, leaving a sample of 100 native English speakers for subsequent analyses.

Of the 100 participants, 57 were male. Participants were on average 20.60 years old (SD = 1.25), and had completed an average of 13.05 years of education (SD = 2.76), meaning the average participant was a current sophomore in college. The sample was 81% Caucasian, 10% Black or African-American, 2% Native Hawaiian or other Pacific Islander, 2% Asian, 1% American Indian or Alaskan Native, and 4% identified as being more than one race; which was representative of the campus population.
Science Measures

Scientific content knowledge was assessed with the Science Content Knowledge Task. This task was a set of 23 questions taken from released items of the 12th grade NAEP Science Assessment (US Dept of Education, 2013). Table 2 shows the list of selected items and how they are categorized within the NAEP framework. In addition, the items themselves are contained in Appendix 1. Items from the 2000 and 2005 tests were selected from the pool of conceptual understanding items within the life science, physical science, and earth/space science content areas. Items from the 2009 and 2011 assessments were selected from the pool of available items if they were classified as either identifying science principles or using science principles within the life science, physical science, and earth/space science content areas. Questions classified as practical reasoning or scientific investigation on the 2000 and 2005 assessments or using scientific inquiry or using technological design on the 2009 and 2011 assessments were not included in the item pool. Beyond content area assessed, the NAEP also classifies problems as “easy”, “medium”, or “hard”. Two easy, three medium, and three hard questions were chosen from each content area. There were only two hard earth/space science questions included in the science content assessment because only two items of this type had been released in the question tool at the time of measure construction. Difficulty maps onto the items’ pass rate; national pass rates are shown in the 2nd column from the right of Table 2.

Scientific reasoning was assessed with the Classroom Test of Science Reasoning (Lawson, 1978). This task has been previously used to assess the science inquiry ability of high school and undergraduate students (Lawson, 1978; Bao et al., 2009). Participants
responded to 24 multiple-choice questions which covered a range of science reasoning topics, including conservation of mass, control of variables, interpretation of data and conclusions, experimental design, and proportional and probabilistic reasoning.

**Math Measures**

Math fluency was assessed using the Woodcock-Johnson III Tests of Achievement (WJ III ACH; Woodcock, McGrew, & Mather, 2001). Participants solved single digit addition, subtraction, and multiplication problems as quickly and accurately as possible. Participants had a time limit of 3 minutes in which to complete 160 items. Math Fluency has a reported median reliability of 0.90 (Schrank, McGrew, & Woodcock, 2001).

Math concept knowledge was measured with the Quantitative Concepts subtest of the WJ III ACH (Woodcock, McGrew, & Mather, 2001). The task has two parts. In part A the participant was asked a variety of math-related conceptual questions. In part B participants completed number series in which one of the numbers in the series was missing. Quantitative Concepts has a reported median reliability of 0.91 (Schrank, McGrew, & Woodcock, 2001).

**Spatial Measures**

Visuo-spatial sketchpad was assessed with the Corsi Block tapping task (Corsi, 1972) using a standardized form of administration (Pagulayan, Busch, Medina, Bartok, & Krikorian, 2006). This pattern was quasi-random, in that each block was tapped only once in a sequence. When the experimenter was done indicating the block sequence the participant reproduced the sequence. There were five trials in each level and nine levels
of difficulty. Levels corresponded to the number of blocks tapped in the sequence. The total raw score out of 45 was utilized in analyses.

Mental Rotation was assessed with the redrawn version of the mental rotation task (Peters et al., 1995; Vandenberg & Kuse, 1978). Participants had 3 minutes to complete 12 trials in which participants examined a target figure and identified two images of a set of four that were rotated images of the target. Participants were then given a four minute break. Following the break, participants completed another 12 trials within another 3 minute time limit. Total score correct out of 24 was used in analyses.

**Reading Measures**

Real-word decoding was assessed with the WJ III ACH Letter-Word Identification subtest (Woodcock, McGrew, & Mather, 2001). Participants read real words aloud to the experimenter, words were progressively more difficult as the participant moved through the task. The task required the participant only to decode the word, which did not necessitate knowing its meaning. Letter-Word Identification has a reported median reliability of 0.94 (Schrank, McGrew, & Woodcock, 2001).

Non-word decoding was assessed with the WJ III ACH Word Attack subtest. Participants read aloud non-words to the experimenter. Word Attack has a reported median reliability of 0.87 (Schrank, McGrew, & Woodcock, 2001).

Reading comprehension was assessed with the WJ III ACH Passage Comprehension subtest. The task was a cloze form test. Participants read a 1-2 sentence passage and provided the missing word. Passage Comprehension has a median reliability of 0.88 (Schrank, McGrew, & Woodcock, 2001).
Chapter 3. Results

Measurement of Scientific Skills

The first aim of the present study was to examine the dimensionality of the science content and science reasoning measures. Before investigating factor structure, the internal consistency of each measure was examined. The percent of participants that passed each of the items of the science content task are presented in Table 1. Participants tended to score higher than the national average. Cronbach’s alpha was calculated for the items as a measure of internal consistency. For the science content measure, alpha was 0.75, indicating adequate internal consistency among the items in the measure. However, it should be noted that based on the percentage of the sample passing each item, the items themselves were not tau-equivalent. Tau-equivalency requires observed item means and error variances to be equal across the entirety of the test. One reason the items were not tau-equivalent was that the items were selected to vary in item difficulty. Importantly, Cronbach’s alpha assumes tau-equivalent scores. When the assumption of tau-equivalency is not met, as was true in this case, then alpha can be considered a lower bound of reliability for the measure.

Similarly, Cronbach’s alpha was used to calculate the internal consistency of the science reasoning measure for this sample. Cronbach’s alpha for the full scale was 0.85,
suggesting a good level of internal consistency. However, if the proportional and probability items were not included in the scale then alpha decreased to 0.77. This suggests alpha may be inflated by math-related items giving a boost of stable variance to the task in comparison to the non-math items.

**Internal Factor Structure of Science**

Factor analysis was completed to examine the factor structure of items on the science content measure. An initial factor extraction was completed using Principal Axis Factoring with no rotation. The scree plot of the resulting eigenvalues is shown in Figure 1. There were 9 eigenvalues above 1; but many of those were below the “elbow” of the scree plot. This first factor accounted for 17.83% of the variance in science content scores. What becomes immediately evident in looking at these basic outputs is that science content knowledge was potentially not unidimensional. However, with a sample of N = 100 participants, an interpretable factor structure of the items was unable to be determined.

Similarly, factor analysis was completed with items on the science reasoning measure. Again, initial factor extraction was done using Principal Axis Factoring with no rotation. Item 2 of the measure had to be left out of the analysis because it was perfectly correlated with item 1, causing a non positive definite correlation matrix, as did item 4 which was perfectly uncorrelated with item 1, causing a communality to exceed 1. The scree plot of the resulting eigenvalues is shown in Figure 2. There were 8 eigenvalues above 1. Similar to content knowledge, the “elbow” of the scree plot occurred after the first factor. The first factor accounted for 24.15% of the variance in scores, no other
factors accounted for more than 10% of the variance. Like the science content measure, science reasoning may not be a unidimensional construct, though a multi-factor structure was uninterpretable.

**Descriptives**

The second aim of the study was to examine science skills in relation to other academic domains. Analyses from aim 1 called into question the use of summed scores for the science measures; however, without a meaningful factor structure to build subscales, a sum score for each task was examined in order to move to correlational analyses. Descriptive statistics of the measures are presented in Table 3. Standard scores were available for the WJ III tasks, for all other tasks raw scores are presented. Scores generally covered the range of possible scores. WJ III standard scores were expected to have a population mean of 100 and a standard deviation of 15. Scores from the sample indicated that participants were less variable than the population as a whole, which would be expected given the selection bias of a college population. To control for possible sex, age, and schooling effects in the raw scores on all the tasks, variance due to these variables were removed using a regression procedure. Standardized residuals were used in all subsequent analyses.

**Correlation among cognitive skills.**

Correlations between the tasks are found in Table 4. Observed correlations ranged from $r=-0.01$ between math fluency and mental rotation to $r=0.62$ between science reasoning and quantitative concepts. There was a general trend of positive manifold, in that participants who score higher in one domain tended to score higher on all other
academic domains. The science content measure was most highly correlated with the science reasoning measure \((r=0.56)\), however, this correlation was not significantly different than the correlations of science content with passage comprehension, letter-word identification, and quantitative concepts. In comparison, the science reasoning measure was most highly related to the quantitative concepts measure \((r = 0.62)\), which may be partially due to the fact that it contained explicitly math-related items. However, this correlation was not significantly different from the correlation between science reasoning and science content or science reasoning and Corsi blocks.

For the most part, science content knowledge and science reasoning skill were more highly correlated with one another than with other cognitive skills. However, there were some notable differences that make reasoning skills different from content knowledge. Using Fisher’s \(r\) to \(z\) transformations to compare dependent correlation coefficients across the cognitive skills assessed (Lee & Preacher, 2013), science reasoning skills were more highly correlated with visuo-spatial sketchpad \((Z=2.417, p =0.016)\), math fluency \((Z=2.214, p=0.027)\), and math concept knowledge \((Z=2.461, p=0.014)\) than science content knowledge was with these skills. Therefore, as science relates to mathematical competencies, it may be useful to think of content knowledge and reasoning skills as separable constructs. However, in regard to links with reading, there were no significant differences between the two components of science.

Importantly, these differences must be considered in light of the fact that the science reasoning measure included proportional and probability items. These math items may act to differentiate science reasoning from content knowledge in their link to visuo-
spatial and mathematics skills simply by including math items in the measure. The 8 math items were significantly correlated with the 16 other science reasoning items (r=0.61). However, when considering only the non-proportional and non-probability items and their link to other academic domains, the correlation between science reasoning and visuo-spatial sketchpad, math fluency, and math concept knowledge were reduced to r=0.42, r=0.23, and r=0.57, respectively. Furthermore, differences between science content knowledge and science reasoning skills in correlation to these domains were erased (Fisher’s r-to-z for dependent correlation coefficients, relation to visuo-spatial sketchpad Z=1.559, p=0.119, relation to math fluency, Z=1.181, p=0.237, and relation to math concept knowledge, Z=1.597 p=0.110). Thus, if mathematical reasoning items are not considered as part of scientific reasoning, then science content knowledge and science reasoning skill did not differ in their relationships with other academic domains.

**Factor Structure of Cognitive Skills.**

Given the moderate and significant correlation among science, mathematics, visuo-spatial, and reading measures in the present study, there were several hypothetical ways in which these skills might be interrelated, a problem investigated with structural equation modeling. Five separate models were fit using SPSS Amos (Arbuckle, 2013). The first model, shown in Figure 3, was a general factor model which assumed that a general cognitive factor accounted for the variance in all of the measures. Next, the other extreme was fit, supposing that each of science, math, visuo-spatial, and reading skills were each separate, but potentially correlated factors (see Figure 4). Because science and math have been shown to be more closely interrelated, a third model combined math and
science into one factor, see Figure 5. However, visuo-spatial skills and mathematics have also been demonstrated to be linked in past work, so instead in the fourth model (see Figure 6) visuo-spatial skills and mathematics were combined into a single factor. Finally, the last model combined science, math, and visuo-spatial skills into one factor while leaving reading separate (see Figure 7).

Model fit statistics are found in Table 5. Chi-square values suggested all models were better than a null model; however comparative fit indices (CFI) and root mean square error of approximation (RMSEA) both suggested that none of the models resulted in an adequate fit of the data. Despite this lack of fit, chi-square difference tests were calculated to compare the models. Model 2, in which each academic domain was separated, was superior to Model 1, (p<0.001), Model 4, (p=0.005), and Model 5 (p=0.013). There was no significant difference between Model 2 and Model 3 (p=0.097), where science and math were combined into one factor. However, these differences must be tempered by the fact that none of the models resulted in particularly good fit. This suggests that many of the traditional ways we might think about combining science, math, reading, and visuo-spatial skills need further scrutiny.
Chapter 4. Discussion

The present study sought to assess both the dimensionality of scientific skills measures and science’s relationship with other cognitive skills. In addressing the first aim of the study, a clear simple structure failed to emerge from the science measures, which begs the question as to why this may be. One reason for this lack of factorial clarity may be that the tasks themselves were tapping into more than one thing, or perhaps not measuring anything coherent at all. Simple structure of the items for both measures of science was unattainable, and the 8 or 9 eigenvalues over 1 resulted in an uninterpretable structure of the data. This is a fact that would be hidden by only assessing their adequate to good internal consistency ($\alpha = 0.75$ for science content knowledge and $\alpha = 0.85$ for the science reasoning measure in this sample). Given the complex but uninterpretable factor structure, it can only be concluded that science content knowledge and science reasoning skill is perhaps not one thing, but science as many things would require, at the very least, larger sample sizes if one hopes to ascertain an interpretable factor structure.

In considering the content measure, in which items were drawn from the NAEP, the hope would be that if a multidimensional structure best describes this data, then the structure would be interpretable within the framework of scientific field or even item difficulty on which the items were categorized; however, this was not the case. There are multiple reasons why the current scale may have failed to have an interpretable
underlying factor structure. Importantly, the results of the current study could reflect lack of interpretability due to the shortened length of the assessment and a smaller sample size. The NAEP Science Assessment in its full form is much longer and administered to thousands of students. The first factor is reported to account for approximately 33% of the variance in scores on the 1996 science assessment (National Research Council, 2000). In comparison, the first factor of the science content knowledge measure constructed for the present study only accounted for 17.83% of the variance in this sample. In addition, the items have been used in 12th grade students previously. Pass rates suggest that the sample in the present study, on average college sophomores, were often much more successful than the average 12th grader.

Turning to science reasoning, the Classroom Test of Science Reasoning, chosen in part because of its relative prominence in the literature, blurs the line between science and mathematics in an obvious way by including proportional reasoning items. To the extent that the measures of science are a mix of questions that also include domains that are actually separable from science, such as reading or mathematics, this may make it difficult to obtain a coherent factor structure.

If science content knowledge and science reasoning skills are both multi-dimensional what does that mean for the use of science as a construct? Parts of this question were addressed in aim two, which examined science in relation to other academic skills. The results of the present study suggested modest to moderate correlation between both science content knowledge and science reasoning skills and a variety of cognitive skills covering the domains of reading, mathematics, and space.
However, after eliminating proportional and probability items from the science reasoning total score, there were no significant differences between science content knowledge and science reasoning skills in their relationship to mathematics, visuo-spatial, and reading skills. Thus, while framing content knowledge as domain-specific and science reasoning as domain-general may seem to align with some of the differences between science as a body of knowledge and science as a process of knowing, there were not meaningful differences in how these scientific skills were related to other academic domains. In addition, many of the ways we might think about the relationship between science, math, visuo-spatial, and reading skills failed to result in a measurement model that fit the data. Therefore, while there is some evidence to suggest that science content knowledge and science reasoning skill may make up science in a factorial sense, more work is needed to address how science achievement is supported by skills in other academic domains.

The question remains as to other ways future research should examine the development of science-related skills. One suggestion may be to focus on particular pieces of content knowledge or reasoning skills, an approach taken by many who have done previous work in the field of science development. Rather than including multiple fields of science or multiple reasoning skills within one comprehensive measure, an approach which examines one particular aspect of science reasoning or content knowledge likely allows for deeper understanding of the cognitive underpinning of these particular processes.

However, there are a few caveats to adopting such an approach. First, content knowledge and reasoning processes do not operate in isolation within the scientific
enterprise. This means that isolating reasoning processes in a research setting often requires using novel science concepts or completely artificial “science” in order to isolate the scientific reasoning process. But, in practice, children and adults come to science with a wealth of science experiences gained simply through life experience, thus interactions between prior knowledge or biases based on these experiences and learning formal science remain an important area of research. Secondly, there is clearly practical utility in being able to assess multiple dimensions of science within one test, and to be able to do so in a relatively brief measure. Many of the measures made by science educators for use in their particular field are similar to this goal. However, the psychometric properties underlying such measures are often not well-understood. Therefore, future research should work on validating brief science-related measures, which has practical value for both researchers and educators.

The present study also has implications for educational practices. Within schooling in the US, science classes tend to cover a broad array of science topics and skills throughout primary school before breaking into subject-specific coursework by high school and college. The present study cannot speak to this arrangement, but it can speak to the way in which achievement is measured. In regard to the science content measure, it may be informative to consider the original purpose of the NAEP Science Assessment from which the questions were taken. The NAEP is constructed as a progress monitoring tool, meant to assess nation-level trends in science achievement over time. As such, the questions themselves may be better-suited to examining diffuse long-term trends rather than specific individual differences in content knowledge. In addition, the
present study chose only 23 items, in part for practical reason because testing on each
science subtest could not go beyond 20 minutes and still have time to complete the
battery in 90 minutes. Therefore, it is possible that these items are not as representative of
science content knowledge as compared to a longer test battery. However, if science
content knowledge is important for things like long-term economic success (US Dept of
Labor, 2007), refining a shortened instrument that validly assesses individual differences
in content knowledge remains a need in potentially guiding future educational practices.

Turning to the science reasoning measure, the task was meant to capture the
breadth of science reasoning skills and has been previously used in individual differences
research, with some prior suggestions that the Classroom Test of Science Reasoning, or a
similar instrument be administered at the beginning of college science courses to identify
students who may need extra help developing science reasoning skills or at risk for
falling behind in the course due to less advanced science reasoning skills (Coletta,
Phillips, & Steinert, 2007; Costenson & Lawson, 1986). However, given that a summed
score collapses across the various types of reasoning skills, this may not be an effective
strategy. For example, a student may get a lower score for poor proportional reasoning
skills but another student get a similarly low score by failing the experimental design
items. Remediating these differences would likely take different forms of intervention, so
simply taking the summed score to identify at-risk students would be combining across
dimensions in a way that is less informative.

Lastly, a developmental question not able to be addressed by the current study,
which examined only collegiate level science content knowledge and reasoning skills, is
what developmental changes are expected over the course of development. If the multidimensionality of science holds, one question of particular interest would be the developmental trajectory of the dimensions of science. However, this work is hinged on building a valid multidimensional scale to measure science achievement.

Another developmental trend worth considering further is when and through what processes sex differences emerge in science content knowledge and science reasoning skills. Though not a focus of the present study, because the correlational analyses controlled for sex, age, and schooling effects, it is important to note that consistent with prior research that has found males tend to score higher on science achievement tests (National Center for Educational Statistics, 2011), sex alone accounted for 10% of the variance in science reasoning skills and 19% of the variance in content knowledge. Though not surprising in an undergraduate sample, the mechanisms through which such circumstances come to fruition is important to addressing causes and correlates of STEM achievement.

In sum, the present study analyzed the dimensionality of science content knowledge and science reasoning skill and their relation with other cognitive skills. The results suggest that science was broadly related to reading, math concept knowledge, and visuo-spatial skills. Science as domain-specific content knowledge and domain-general scientific reasoning strategies dichotomizes scientific thinking in a way that does not clearly separate similarities and differences between the two. In addition, the tests of content knowledge and reasoning skills utilized in the present study were not unidimensional. Future research should work to further refine this notion of science as
many things. Subcomponents of scientific thinking likely interact with one another in addition to being linked with achievement in other academic domains. Thus, furthering understanding of the factorial structure of scientific understanding is crucially important to designing valid tests that elucidate the causes and consequences of individual differences in scientific skills.
References


U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP),


Appendix A: Tables and Figures

<table>
<thead>
<tr>
<th>Measure</th>
<th>df</th>
<th>t</th>
<th>sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Letter-Word Identification - Standard</td>
<td>117</td>
<td>-4.63</td>
<td>.000</td>
</tr>
<tr>
<td>Word Attack - Standard</td>
<td>117</td>
<td>-3.64</td>
<td>.000</td>
</tr>
<tr>
<td>Passage Comprehension - Standard</td>
<td>109</td>
<td>-10.82</td>
<td>.000</td>
</tr>
<tr>
<td>Math Fluency – Standard</td>
<td>117</td>
<td>0.61</td>
<td>.545</td>
</tr>
<tr>
<td>Quantitative Concepts – Standard</td>
<td>114</td>
<td>-4.18</td>
<td>.000</td>
</tr>
<tr>
<td>Mental Rotation – Raw</td>
<td>117</td>
<td>-2.66</td>
<td>.009</td>
</tr>
<tr>
<td>Corsi Blocks – Raw</td>
<td>105</td>
<td>-2.79</td>
<td>.006</td>
</tr>
<tr>
<td>Science Content – Raw</td>
<td>117</td>
<td>-1.30</td>
<td>.195</td>
</tr>
<tr>
<td>Science Reasoning - Raw</td>
<td>117</td>
<td>-1.88</td>
<td>.062</td>
</tr>
</tbody>
</table>

Table 1. *Differences between native and non-native English speakers*
<table>
<thead>
<tr>
<th>Item</th>
<th>Science Area</th>
<th>Year</th>
<th>Framework</th>
<th>Difficulty</th>
<th>National Pass %</th>
<th>Sample Pass %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Physical</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Easy</td>
<td>73</td>
<td>84</td>
</tr>
<tr>
<td>2</td>
<td>Physical</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Easy</td>
<td>74</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>Earth/Space</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Easy</td>
<td>70</td>
<td>82</td>
</tr>
<tr>
<td>4</td>
<td>Earth/Space</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Easy</td>
<td>60</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>Life</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Easy</td>
<td>70</td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>Life</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Easy</td>
<td>84</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>Physical</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Medium</td>
<td>44</td>
<td>78</td>
</tr>
<tr>
<td>8</td>
<td>Physical</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Medium</td>
<td>41</td>
<td>56</td>
</tr>
<tr>
<td>9</td>
<td>Physical</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Medium</td>
<td>59</td>
<td>61</td>
</tr>
<tr>
<td>10</td>
<td>Earth/Space</td>
<td>2009</td>
<td>Using Science Principles</td>
<td>Medium</td>
<td>48</td>
<td>52</td>
</tr>
<tr>
<td>11</td>
<td>Earth/Space</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Medium</td>
<td>54</td>
<td>73</td>
</tr>
<tr>
<td>12</td>
<td>Earth/Space</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Medium</td>
<td>42</td>
<td>49</td>
</tr>
<tr>
<td>13</td>
<td>Life</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Medium</td>
<td>41</td>
<td>61</td>
</tr>
<tr>
<td>14</td>
<td>Life</td>
<td>2000</td>
<td>Conceptual Understanding</td>
<td>Medium</td>
<td>51</td>
<td>87</td>
</tr>
<tr>
<td>15</td>
<td>Life</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Medium</td>
<td>46</td>
<td>70</td>
</tr>
<tr>
<td>16</td>
<td>Physical</td>
<td>2009</td>
<td>Identifying Science Principles</td>
<td>Hard</td>
<td>18</td>
<td>31</td>
</tr>
<tr>
<td>17</td>
<td>Physical</td>
<td>2009</td>
<td>Using Science Principles</td>
<td>Hard</td>
<td>38</td>
<td>66</td>
</tr>
<tr>
<td>18</td>
<td>Physical</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>25</td>
<td>58</td>
</tr>
<tr>
<td>19</td>
<td>Earth/Space</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>23</td>
<td>37</td>
</tr>
<tr>
<td>20</td>
<td>Earth/Space</td>
<td>2000</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>37</td>
<td>56</td>
</tr>
<tr>
<td>21</td>
<td>Life</td>
<td>2000</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>34</td>
<td>47</td>
</tr>
<tr>
<td>22</td>
<td>Life</td>
<td>2000</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>28</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>Life</td>
<td>2005</td>
<td>Conceptual Understanding</td>
<td>Hard</td>
<td>20</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2. NAEP items included in the science content measure
<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science Content</td>
<td>100</td>
<td>3</td>
<td>22</td>
<td>14.62</td>
<td>4.00</td>
</tr>
<tr>
<td>Science Reasoning</td>
<td>100</td>
<td>4</td>
<td>24</td>
<td>16.35</td>
<td>4.76</td>
</tr>
<tr>
<td>Mental Rotation</td>
<td>100</td>
<td>0</td>
<td>22</td>
<td>10.68</td>
<td>5.21</td>
</tr>
<tr>
<td>Corsi Blocks</td>
<td>94</td>
<td>23</td>
<td>43</td>
<td>33.96</td>
<td>4.34</td>
</tr>
<tr>
<td>Letter-Word Identification</td>
<td>100</td>
<td>86</td>
<td>119</td>
<td>100.82</td>
<td>6.86</td>
</tr>
<tr>
<td>Word Attack</td>
<td>100</td>
<td>77</td>
<td>120</td>
<td>96.47</td>
<td>8.16</td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td>98</td>
<td>86</td>
<td>126</td>
<td>104.33</td>
<td>7.84</td>
</tr>
<tr>
<td>Math Fluency</td>
<td>100</td>
<td>67</td>
<td>138</td>
<td>107.05</td>
<td>13.68</td>
</tr>
<tr>
<td>Quantitative Concepts</td>
<td>99</td>
<td>75</td>
<td>127</td>
<td>104.84</td>
<td>10.90</td>
</tr>
</tbody>
</table>

Table 3. Descriptive statistics
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Science Content</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Science Reasoning</td>
<td>0.56**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Mental Rotation</td>
<td>0.23*</td>
<td>0.29**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Corsi Blocks</td>
<td>0.27**</td>
<td>0.49**</td>
<td>0.38**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Word Identification</td>
<td>0.44**</td>
<td>0.32**</td>
<td>0.14</td>
<td>0.08</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Word Attack</td>
<td>0.33**</td>
<td>0.20*</td>
<td>0.17</td>
<td>0.15</td>
<td>0.63**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Passage Comprehension</td>
<td>0.55**</td>
<td>0.44**</td>
<td>0.23*</td>
<td>0.29**</td>
<td>0.43**</td>
<td>0.33**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8. Math Fluency</td>
<td>0.12</td>
<td>0.32**</td>
<td>-0.01</td>
<td>0.32**</td>
<td>0.25*</td>
<td>0.21*</td>
<td>0.06</td>
<td>-</td>
</tr>
<tr>
<td>9. Quantitative Concepts</td>
<td>0.44**</td>
<td>0.62**</td>
<td>0.21*</td>
<td>0.34**</td>
<td>0.40**</td>
<td>0.33**</td>
<td>0.54**</td>
<td>0.40**</td>
</tr>
</tbody>
</table>

*Note. *p < .05, **p<.001

Table 4. Observed correlations between cognitive skills
<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>$X^2$</th>
<th>CFI</th>
<th>RMSEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27</td>
<td>90.760</td>
<td>0.755</td>
<td>0.154</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>59.896</td>
<td>0.851</td>
<td>0.137</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>66.224</td>
<td>0.838</td>
<td>0.133</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>72.690</td>
<td>0.813</td>
<td>0.143</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>74.257</td>
<td>0.815</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Table 5. *Model fit statistics*
Figure 1. Scree plot - Science content
Figure 2. *Scree plot - Science reasoning*
Figure 3. Model 1: General factor model
Figure 4. Model 2: All separate factors
Figure 5. Model 3: Science and math combined
Figure 6. Model 4: Math and space combined
Figure 7. Model 5: Science, math, and space combined
Appendix B. Science Content Measure

1. Each figure below shows a force measured in newtons pushing on a block. If there are no other forces pushing on the block, in which case is the acceleration of the block greatest?

A.  
B.  
C.  
D.  

2. The figure above shows some ocean waves. Which of the labeled distances represents the wavelength?

A.  
B.  
C.  
D.  

3. The two most abundant elements in the Solar System are

A. hydrogen and helium  
B. hydrogen and calcium  
C. calcium and iron  
D. helium and iron
4. The following question refers to the carbon cycle diagram below, which shows some of the ways that carbon moves through different parts of Earth's environment.

![Carbon Cycle Diagram]

In the diagram, which portion of the carbon cycle is driven directly by Earth's internal heat energy?
A. The movement of carbon between the ocean and the atmosphere
B. The emission of carbon dioxide from oil-burning power plants
C. The release of carbon dioxide during volcanic eruptions
D. The exhalation of carbon dioxide by animals

5. Which statement about the offspring that result from sexual reproduction is generally true?
A. The offspring show genetic variation from the parents.
B. The offspring have genetic material identical to that of one another.
C. The offspring have genetic material identical to that of one of the parents.
D. The offspring have twice as much genetic material as each parent.
6. The bird shown above would be most likely to eat
   A. insects from the ground
   B. nectar from tube-shaped flowers
   C. fish from the ocean
   D. amphibians from a pond

7. What is the main reason that water has the ability to dissolve many different substances?
   A. Water has a lower molecular mass than many substances.
   B. Water molecules attract ions and the charged parts of molecules.
   C. Water molecules are larger than the ions or molecules they dissolve.
   D. Water is more dense in the liquid phase than in the solid phase.

8. Which of the following observations about a certain pure solid would indicate most strongly that the solid is ionic?
   A. Its water solution is a good conductor of electricity.
   B. It is composed of small white crystals.
   C. It has a density greater than 1.0 gram/cm$^3$.
   D. It has a high melting point.

9. When sulfuric acid, H$_2$SO$_4$, is broken down into separate elements, how many different elements result?
   A. Two
   B. Three
   C. Six
   D. Seven

10. Why does radioactive dating of meteorites give a more accurate age for Earth than radioactive dating of rocks at Earth's surface?
    A. Most rocks that first formed on Earth's surface have since undergone major geologic changes.
    B. Most rocks found on Earth's surface are older than most meteorites.
    C. Most meteorites contain minerals that are the same age as those found on Earth's surface.
    D. Most meteorites are made of the same type of iron that is found in Earth's core.
11. Coal, petroleum, and natural gas found underground in certain parts of Earth are primarily formed from which process?
A. Decay of radioactive elements
B. Collision of tectonic plates in earthquakes
C. Transformation of dead plants and animals under heat and pressure
D. Intrusion of water into the soil that breaks up rocks and minerals

12. What is a property of all galaxies?
A. All galaxies have a spiral shape.
B. All galaxies are the same size.
C. All galaxies contain a large number of stars.
D. All galaxies rotate around a central star.

13. What is the correct order for the levels of organization in living systems from the simplest to the most complex? (Note that not all levels of organization are included.)
A. Elements → molecules → cells → tissues → organs
B. Molecules → tissues → cells → organs → organisms
C. Molecules → elements → tissues → organs → organisms
D. Cells → organs → tissues → organisms → molecules

14. Which of the following is NOT a part of Darwin's theory of evolution by natural selection?
A. Individuals in a population vary in many ways.
B. Some individuals possess features that enable them to survive better than individuals lacking those features.
C. More offspring are produced than can generally survive.
D. Changes in an individual's genetic material are usually harmful.

15. Which pair of systems regulate and coordinate body functions?
A. Excretory and digestive
B. Nervous and endocrine
C. Skeletal and muscular
D. Immune and respiratory

16. Which of the following is an example of a nuclear fission reaction?
A. $^8_4\text{Be} + ^4_2\text{He} \rightarrow ^{12}_6\text{C}$
B. $^{234}_{90}\text{Th} \rightarrow ^{234}_{91}\text{Pa} + ^0_{-1}\text{e}$
C. $^{27}_{13}\text{Al} + ^4_2\text{He} \rightarrow ^{30}_{15}\text{P} + ^1_0\text{n}$
D. $^{239}_{94}\text{Pu} + ^1_0\text{n} \rightarrow ^{137}_{52}\text{Te} + ^{100}_{42}\text{Mo} + 3^1_0\text{n}$
17. When a cork is added to a glass of water, the cork floats at the top of the water instead of sinking to the bottom of the glass. Which statement helps explain why this happens?
   A. The cork absorbs the water, so it cannot sink.
   B. The water is exerting an upward force on the cork.
   C. The water stops the force of gravity from acting on the cork.
   D. The density of the cork is greater than that of the water.

   \[ \ldots \text{C}_3\text{H}_6 + \ldots \text{O}_2 \rightarrow \ldots \text{CO}_2 + \ldots \text{H}_2\text{O} \]

18. When the equation above is balanced and all coefficients are reduced to their lowest whole-number values, the coefficient for \( \text{H}_2\text{O} \) is
   A. 2
   B. 3
   C. 4
   D. 6

19. What two gases make up most of the Earth's atmosphere?
   A. Hydrogen and oxygen
   B. Hydrogen and nitrogen
   C. Oxygen and carbon dioxide
   D. Oxygen and nitrogen

20. Air in the atmosphere continuously moves by convection. At the equator, air rises; at the poles, it sinks. This occurs because
   A. the Earth's ozone layer is thinner at the equator than at the poles
   B. the Earth's magnetic field is stronger at the poles than at the equator
   C. warm air can hold less water vapor than can cold air
   D. warm air is less dense than cold air

21. During the time in the Earth's history when the first amphibians appeared, which of the following was one of the major groups of plants that dominated the land habitats?
   A. Cone-bearing trees (gymnosperms)
   B. Flowering plants (angiosperms)
   C. Ferns
   D. Algae
22. According to evolutionary theory, which of the following evolutionary trees best describes the relationship between groups of vertebrates?

A. 

B. 

C. 

D.
23. Each diagram below shows the same front view of a human heart. Which diagram has arrows that correctly show the path of blood flow through the heart and the blood vessels leading to and from the heart?

A. 

B. 

C. 

D.