ABSTRACT

OUZTS, SANDYE MICHELLE. Response Processes Validity of the Stanford Binet Intelligence Scales, Fifth Edition. (Under the direction of Jeffery Braden).

This study examined the evidence for the response process validity of the Stanford Binet, Fifth Edition (Roid, 2003a). Students from introductory psychology classes (n = 101) were randomly assigned to one of five conditions defined by the five cognitive processes the SB5 intends to measure. Participants responded to items from the Verbal and Nonverbal Scale subtests for their condition. Participants explained how they solved the problems and rated the degree to which they used each of the cognitive processes and verbal mediation. Graduate student raters independently assigned participants' descriptions into categories representing the cognitive process and level of verbal mediation. Results generally provide strong evidence for the response processes validity of the SB5, as well as the measures used in this study.

Response Processes Validity of the Stanford Binet Intelligence Scales,
Fifth Edition

by
Sandye Michelle Ouzts

A thesis submitted to the Graduate Faculty of
North Carolina State University
In partial fulfillment of the
Requirements for the degree of
Master of Science

Psychology

Raleigh, North Carolina

2007

APPROVED BY:

Joan Michael, PhD

John Nietfeld, PhD

Jeffery Braden, PhD
Chair of Advisory Committee

BIOGRAPHY

I was born in Greenwood, South Carolina, the daughter of Lawrence and Brenda Ouzts. I am the older sister of Jason Ouzts. I received my high school degree, graduating third in my class, from Saluda High School in Saluda, South Carolina in 1999. In 2003, I graduated Magna Cum Laude with a Bachelor of Arts and honors in Psychology from the University of South Carolina. I entered North Carolina State University for graduate study in School Psychology in the fall of 2003.

Growing up in my rural neighborhood in South Carolina, the houses were far and few between. I believe the population of cows was greater than the population of people. After graduating from high school, I attended the small campus of the University of South Carolina at Aiken with a student body population of about 3,000. Originally a Communications major, I became interested in psychology after taking Introduction to Psychology with Maureen Carrigan, who later became my advisor. My senior year, I conducted research on child maltreatment under the guidance of Keri Weed, getting my first hands on experience with research and psychological testing and also my first publication.

After arriving in the big city of Raleigh to attend graduate school at North Carolina State University, I experienced a bit of "culture shock." Raleigh was quite a change from the small towns of South Carolina I had known. Even today, I am still not exactly fond of the traffic and find the "inner" and "outer" beltlines confusing. With the help of dear friends to whom I will always be grateful, I successfully navigated through many years of graduate school. These have been the most eventful years of my life. My accomplishments have included co-authoring two articles in the Mental Measurements Yearbook with Jeff Braden, two NCSPA poster presentations, many conferences, a departmental research award, learning about and working with children with disabilities, a successful (and publishable) thesis, and completing my 1200 hour internship.

After graduation, I plan to work in the public school system as a Specialist level school psychologist. Who knows where I will be ten years from now, but I do know that a new chapter of my life is about to begin.

ACKNOWLEDGMENTS

I would like to thank my advisor, Jeff Braden, not only for the help and encouragement along the way, but also for setting an example for myself and others. As I have known Jeff over the years, he has been a leader with integrity and compassion whose pursuit is excellence and truth. I would also like to thank my committee members, John Nietfeld and Joan Michael, for their time and feedback on my thesis.

I would like to thank Lauren, Michelle, Sena, Hailey, and Kelly for scoring responses, reviewing thesis drafts, and giving me feedback. I am also grateful to Danielle and Elizabeth for conducting the participant experiment sessions and helping to develop scoring rubrics. Thanks also to Lynne Myers of Wake County Public Schools for lending me a SB5 kit to use for my study. Thanks to a dear friend, Michelle, for listening and giving feedback during practice runs for my presentations and giving me tips on using Microsoft Word. Thanks to Athena for her friendship and encouragement over the years. I also want to give great thanks to Kevin, who has been an invaluable source of moral support and a great friend.

Finally, I am grateful to my mom and dad for their love and support throughout my graduate career.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Validity and Its Importance

According to the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), validity is the extent to which the interpretation of test scores is supported by both evidence and theory. Test developers must provide evidence to support their claims concerning the appropriate uses and interpretation of test scores, and professionals need this information to ethically base decisions on the results of the test. According to the *Standards*, there are five types of evidence that may be used to support validity: (1) evidence based on test content, (2) response processes, (3) internal structure, (4) relations to other variables, and (5) consequences of testing. Evidence based on response processes consists of evidence that the examinee is actually engaging in the processes that the tasks are intended to measure, as well as evidence that the examiner is evaluating assessment data according to the processes that the tasks are intended to assess. The focus of this paper will be processes used by examinees, not the processes examiners use to interpret data.

Response Processes: Importance and Research

The inclusion of evidence based on response processes in the *Standards* stems from work by Samuel Messick. According to Messick (1994, 1995, 2000), there are six aspects of validity: content, substantive, structural, generalizability, external, and consequential. The "substantive" aspect of validity is identical to the concept of evidence based on response processes. Messick (2000) also argued that evidence that empirically evaluates the meaning and consequences of test scores is essential. Evidence based on response processes relies on empirical analysis of the processes used by examinees and examiners during assessment. Therefore, evidence based on response processes is an important source of validity evidence.

According to Messick (1994, 1995), there are two possible sources of invalidity, construct underrepresentation and construct-irrelevant variance. Construct underrepresentation occurs when the task fails to assess some important aspect of the construct it is intended to measure. Construct-irrelevant variance occurs when the task assesses an irrelevant construct or the examinee's responses are influenced by extraneous variables that the test does not intend to measure. These sources of invalidity are particularly relevant to all cognitive tests because examinees may fail to use the cognitive processes the tests intend to measure (construct underrepresentation), or may use processes that a particular test does not intend to measure (task irrelevant variance). For example, if examinees use inner language, or verbal mediation, in comprehending, mediating, or responding to a nonverbal task, then the test scores should be considered invalid as nonverbal measures of cognitive processes. This would alter the meaning test users might assign to scores relative to the meaning they might assign if examinees did not use inner language. In sum, evidence that examinees do (or do not) use intended processes, and avoid using unintended processes, would help test users better understand the validity of any test.

Cognitive Psychology Research

Evidence based on response processes can be obtained using a variety of methods. However, researchers cannot directly measure cognitive processes because these processes cannot be observed. Researchers can only infer which processes examinees are using when performing a task. Cognitive psychologists utilize a variety of methods to infer information processing, including eye movements, response latencies, error analysis, measurement of brain activity, and examinee interviews (Snow & Lohman, 1989). Using information collected by multiple methods, researchers are able to more accurately draw conclusions about the processes that a particular task elicits.

Eye movement and response latencies were used in one study to develop models of the mental processes involved in a mental rotation task, sentence verification task, and quantitative comparison task (Just & Carpenter, 1976). In another study (Carpenter, Just, & Shell, 1990), eye movement, response latencies, and error analysis were used to develop a model of the processes that examinees used to respond to the Raven Progressive Matrices test (Raven, 1962). Individual differences were examined to determine what processes (e.g., working memory) contributed to participants' ability to answer items correctly on the test (Carpenter, Just, & Shell, 1990).

Cognitive psychologists have also examined brain activity as measured by an electroencephalogram (EEG), positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and an electromyogram (EMG) to draw conclusions about response processes involved in various tasks. An EEG records brainwave patterns, both the PET and the fMRI measure changes in blood flow in the brain, and an EMG measures the electrical activity of a muscle. EEG and EMG activity, number of errors, and reaction time were used in one study to examine how response processes involved in inhibition varied as a function of age (Ridderinkhof & van der Molen, 1995). In another study, participants' reaction time and EEG activity when completing a Sternberg memory task and a Stroop-like task were examined to understand the processes involved in reaction time (Houlihan, Campbell, & Stelmack, 1994). Reaction time and PET activity have been used to study the effects of aging on the processes involved in completing a verbal recognition memory task (Madden et al., 1999). In a recent study, reaction time and fMRI activity across different types of tasks were examined to infer which areas of the brain are involved in response selection (Schumacher & D'Esposito, 2002).

Education and Other Areas of Psychology

In other areas of research, participants are interviewed and asked to complete questionnaires about the cognitive processes they used to complete a task. A questionnaire was used in one study to analyze college students' response strategies for answering personality test items (Gordon & Holden, 1996). College students' cognition and use of learning strategies related to reading and understanding information from a Biology textbook have been assessed by questionnaire, written discussion, and interview discussion (McCrindle & Christensen, 1995). Graduate students' use of networking, a visual-spatial learning strategy, to comprehend novel text has been assessed through interviews. In addition, participants' prior use of learning strategies (e.g., elaboration) has been assessed using a questionnaire (De Simone & Schmid, 2004).

Researchers have also interviewed school-aged children about the cognitive strategies they used during the learning process. Fourth and fifth grade students' knowledge construction during classroom instruction has been assessed. After taking a classroom test, students were interviewed about the process they used to learn the material required for each test item (Nuthall, 2000). Fourth through eighth grade students' metacognition related to reading and understanding science textbooks was measured using a questionnaire and structured interview (Craig & Yore, 1995). Another study assessed the cognitive strategies of seventh grade students while they constructed a model of a tower in groups. Students were videotaped during the task and later participated in an interview in which they described the process the group used to complete the task. In addition, students' statements during the problem solving were analyzed (Welch, 1998).

Each of the methods described could provide useful information about the cognitive processes examinees use when solving test items from an intelligence test. Interviews and questionnaires are relatively easy and convenient methods that could be used to assess response processes validity. Research on the response processes in which examinees engage when taking an intelligence test would provide further insight into the specific cognitive abilities that are measured by that intelligence test. This type of information could be used to guide and refine theories of intelligence, as well as to inform the practice of intellectual assessment.

Validity of Intelligence Tests

Intelligence theory began with Spearman's model of general intelligence, or *g*, after which researchers began to develop hierarchical models of intelligence comprised of many different factors that contributed to intelligence. Carroll (1993) defined these factors as creating a three-stratum model, in which he proposed three different levels of abilities: general ability, broad abilities, and narrow abilities. Clearly, there has been a shift in focus away from measuring general intellectual ability to measuring the many different—and distinct—cognitive processes involved in intellectual behavior. Today, many intelligence tests are based on the three-stratum Cattell-Horn-Carroll (CHC) model (McGrew, 2005) and, therefore, claim to measure different cognitive processes that compose intelligence. Thus, evidence that examinees actually engage in the cognitive processes that the test claims to measure is an important component of validity evidence for intelligence tests (Braden & Niebling, 2005).

Tasks used to assess intelligence vary with respect to the amount of language they require examinees to use when responding to items. In comparison to verbal (or language-loaded) tasks, nonverbal (or language-reduced) tasks reduce the amount of language involved in test administration, item content, and examinee responses. These tests are often used with examinees who are hearing impaired, have language deficits, or who have not been exposed to the dominant language of their culture of residence. Nonverbal tests reduce the likelihood that examinees will use language to understand, mediate, or produce their responses. Although nonverbal tests purport to measure intelligence without the use of language processes, it may be that nonverbal tests do, in fact, elicit language processes in examinees (Braden & Anathasiou, 2005).

There are two distinct schools of thought regarding the cognitive processes nonverbal tests elicit. On the one hand, some intelligence tests are based on the assumption that nonverbal tasks measure the same cognitive processes as verbal, or verbally-loaded, tasks (i.e., intelligence). On the other hand, some tests are developed based on the assumption that nonverbal intelligence tests measure a different (i.e., nonverbal) kind of intelligence. Each of these assumptions lead to somewhat different interpretations of test scores (e.g., does a score represent an ability that can be measured either verbally or nonverbally, or does it represent a cognitive process that is exclusively nonverbal in nature?).

Although many contemporary tests have been developed to reflect the hierarchical, CHC model of intelligence, only one has sought to measure CHC abilities using verbal and nonverbal tests: the Stanford Binet Fifth Edition (SB5; Roid, 2003a). The SB5 purports to measure five different cognitive processes: (1) Fluid Reasoning, (2) Quantitative Reasoning, (3) Visual Spatial Processing, (4) Knowledge, and (5) Working Memory, with verbal and nonverbal subtests to measure each process. For example, the nonverbal subtest for Working Memory requires the examinee to tap blocks in the same order as the examiner, whereas the verbal subtest requires the examinee to repeat the last words of sentences in order (Roid, 2003b; 2003c). In contrast, most other tests confound language loading with the abilities measured (e.g., Knowledge is measured exclusively with verbal tests, whereas Fluid Reasoning is measured primarily with language-reduced tests). Confounding language with ability can restrict the range of cognitive abilities (i.e., nonverbal abilities) a test can measure if it is exclusively verbal or nonverbal. The question of whether nonverbal tests elicit similar processes as verbal tests of intelligence, or whether nonverbal tests elicit processes other than those elicited by verbal tests, is not directly addressed by relevant evidence for most current cognitive test batteries (Braden & Anathasiou, 2005).

Description of the Stanford Binet, Fifth Edition

The subtests in the SB5 battery purport to measure five types of cognitive abilities, providing verbal and nonverbal subtests to measure each of the cognitive abilities. Table 1.1 displays the SB5 subtests, the activities that comprise each subtest, and the abilities they intend to measure. Table 1.2 defines the nature of the five cognitive abilities the SB5 intends

Table 1.1

SB5 Subtests and Activities (Level 3, 4, and 5) According to Cognitive Process and Level of Verbal Mediation

	Nonverbal Subtests	**Verbal Subtests**
Fluid Reasoning	Object Series/Matrices	Early Reasoning Verbal Absurdities Verbal Analogies
Knowledge	Procedural Knowledge, Picture Absurdities	Vocabulary
Quantitative Reasoning	Quantitative Reasoning	Quantitative Reasoning
Visual-Spatial Processing	Form Patterns	Position and Direction
Working Memory	Block Span	Memory for Sentences Last Word

Table1.2

Description of Five Processes Measured by the SB5

Process	**Definition**
Fluid Reasoning	Ability to determine the underlying rules and relationships among pieces of novel information
Knowledge	Range of general knowledge acquired at home, school, or work
Quantitative Reasoning	Ability to reason with numbers or to solve numerical problems
Visual-Spatial Processing	Ability to see patterns, relationships, positions in space, or the whole picture among pieces of a visual display
Working Memory	Ability to temporarily hold in mind and then transform or sort information in one's memory

to measure. These abilities are derived from the Cattell-Horn-Carroll model of intelligence (Roid, 2003b).

The SB5 technical manual (Roid, 2003c; see Chapter 1 and Chapter 2) and examiner's manual (Roid, 2003b; see Chapter 2 and 5) present sufficient evidence for three of the five types of evidence discussed in Chapter 1 of the *Standards*: (1) evidence based on test content, (2) relations to other variables, and (3) internal structure. However, there is little if any evidence relevant to understanding the other two types of validity: *viz.*, test consequences, and response processes (see Braden & Niebling, 2005). In the manuals, the author provides information about how scores from the SB5 change across the developmental lifespan, providing indirect evidence that the processes underlying the scores are different (Braden & Niebling, 2005). However, more direct evidence of response processes is needed to understand to what degree SB5 subtests represent intended constructs, and avoid representing unintended constructs. The current study will examine response processes evidence by asking test-takers to explain what processes they use when solving test items. This approach shifts the unit of analysis away from indirect methods of identifying processes (e.g., factor analyses, developmental differences in scores) to more direct reports from the individual test-taker.

Purposes of the Study

The present study will examine the degree to which claims regarding the type of process, and the degree to which they are verbal or nonverbal, are supported for the SB5. Volunteer participants will be given selected items from the SB5 and then interviewed about the response processes they used to solve the items. First, participants will be asked to discuss how they solved the items. Second, participants will be asked to identify the degree to which they used each of the five cognitive processes the SB5 intends to sample, as well as the degree to which they used verbal mediation, when responding to test items. The results of this study will help to illuminate the degree to which post-assessment interviews may elicit evidence of response processes when applied to cognitive tests, and to supply evidence of response processes examinees use in responding to SB5 subtests and items. The hypotheses of the study, along with the predictions for each hypothesis, are below:

1. Examinees will report using the intended cognitive process (e.g., Working Memory) more than they report using the other four cognitive processes that the subtest is not intended to measure.
    a) Prediction one. The mean rating for the degree to which examinees use the intended cognitive process will be higher than the mean rating of the other four cognitive processes that the subtest is not intended to measure.
    b) Prediction two. Participants' answers to open-ended questions about how they solved the problems on verbal subtests will reflect the cognitive process that the subtest is intended to measure more than the processes the subtest is not intended to measure.
    c) Prediction three. Participants' answers to open-ended questions about how they solved the problems on nonverbal subtests will reflect the cognitive process that the subtest is intended to measure more than the processes the subtest is not intended to measure.

2. Examinees will report using primarily verbal mediation in response to verbal subtests and avoid using verbal mediation in response to nonverbal subtests.
    a) Prediction one. The mean rating for the degree to which examinees use verbal mediation will be higher for verbal subtests than the mean rating for nonverbal subtests.
    b) Prediction two. Participants' answers to open-ended questions about how they solved problems on verbal subtests will reflect the level of verbal mediation (i.e., verbal or nonverbal) that the subtest is intended to measure.
    c) Prediction three. Participants' answers to open-ended questions about how they solved problems on nonverbal subtests will reflect the level of verbal mediation (i.e., verbal or nonverbal) that the subtest is intended to measure.

CHAPTER 2: METHOD

Pilot Studies

Pilot Study One

A pilot study was conducted with a sample of three college students (two females, one male; age range 23 to 25 years; level of education range one year of college to one year of graduate study). Participants were administered Level 3 and Level 5 items from ten subtests: verbal and nonverbal subtests for each of the five of the processes. I asked participants to explain how they solved the items using their own words (i.e., "How did you solve these problems?"). One participant rated the degree to which he used the five processes and inner language for both verbal and nonverbal subtests for three of the five processes.

Participants' answers to the initial open-ended question did not necessarily provide enough information for the interviewer to determine the strategy used to solve the problems. Follow-up questions, such as "How did you know how to (insert participant's words)?" were necessary to elicit sufficiently detailed information. After asking questions, summarizing what the participant stated sometimes elicited additional information as well. Participants often had difficulty remembering which items they were being asked about; showing the stimulus book pages provided participants with a cue so that they could explain how they solved the items. I concluded that follow-up questions and summarizing would be helpful in obtaining detailed information from participants. In addition, I decided that participants may need to review the stimulus pages for some subtests so they can accurately explain how they solved the problems.

Pilot Study Two

A second pilot study was conducted with a sample of four undergraduate students and one graduate student (three females, two males). The purpose of the pilot study was to: (a) provide information about the types of follow-up questions that would elicit the most amount of information, (b) provide information about the best way to phrase definitions of the five processes and inner language, so that participants would understand the task, and (c) explore whether procedures elicited unintended consequences (e.g., Do participants "make up" answers to interview questions or survey items?).

Participants answered selected items from Levels 3, 4, and 5 from the Nonverbal and Verbal Scale subtest for one of the five cognitive process domains (the same procedure used in the actual study). Each participant was administered subtests from a different cognitive process domain (e.g., only one person was administered the two subtests for Working Memory). Before answering test items, participants were given a list that defines the five processes and inner language (see Appendix A) and asked to read the list to cue them to consider which processes they might use in solving the tasks. After each subtest, I asked participants to explain how they solved the items using their own words (i.e., "How did you solve the last nine items you answered?"). Follow-up questions were asked to elicit additional information until the participant gave at least five sentences in response to the question. I also asked participants to rate the degree to which they used each of the five cognitive processes, as well as the degree to which they used "inner language," using the Likert-type rating scale (see Appendix B), when responding to all nine items that were administered.

I asked participants to discuss their thoughts and reactions to the testing and interview including whether they felt the interview questions adequately elicited information about how they solved the problems. Concerning the survey, I asked participants whether they felt: (a) they understood the definitions on the rating scales, (b) their ratings on the survey were accurate, (c) they based their ratings on all nine of the items, and (d) whether easier items elicited different processes than more difficult items. Participants were also invited to discuss any other suggestions they had for improving the procedure and any other comments they had.

When asked whether they felt their ratings reflected all 9 test items, three participants felt that their use of the processes depended on the level of difficulty of the items. For example, when responding to Knowledge items, one participant reported using primarily Knowledge for easy items and primarily Visual Spatial Processing for more difficult items. When given Visual Spatial Processing items, a participant reported that she used Visual Spatial Processing for both easy and difficult items, but that she used Visual Spatial Processing to a greater degree for the difficult items. Further, participants reported that it was difficult to rate the degree to which they used “inner language” because many items could be answered with great ease and required little thought. One participant reported rating “inner language” based only on the last (and most difficult) item because that was the only item she felt required complex thought. Thus, it appears that participants may have had difficulty rating “inner language,” and also may have based these ratings only on select items, because they were not aware of the cognitive processes they used for the easier items.

In some cases, I made slight changes to the testing materials and interview procedure based on participant feedback. I revised the definitions for Fluid Reasoning, Knowledge, Working Memory, and Inner language to make the definitions clearer and easier to apply to the testing situation. In addition, I changed the procedure to allow participants more time to complete items from the nonverbal Visual Spatial Processing subtest (Form Patterns) to facilitate rapport with the participants.

Main Study

Participants

A total of 101 students (100 undergraduate, 1 graduate student) from Psychology 200 classes at North Carolina State University participated in the study. As part of their course requirements, introductory psychology students are given the choice of either writing a research paper or participating in research studies. Participants who elected to participate in research selected and signed up for research studies through a website provided by the university. Students who signed up for this research study received two research credits toward the six-credit research requirement for Psychology 200. The mean age of participants was 20.4 years (range 17 to 53 years). Most participants were female (39.6% male, 60.4% female), Caucasian (75.2%), and did not report having a disability (98%). Descriptive data for participants in each of the five conditions (e.g., Fluid Reasoning, Knowledge) including age, gender, ethnicity, class, and disability status are presented in Table 2.1. None of the participants withdrew or were excluded from the analyses of the results.

Table 2.1

Descriptive Data for Participants Sampled According to Condition

Condition		FR	KN	QR	VSP	WM	All Conditions
Demographic Characteristics							
Age	*M*	18.60	22.14	20.90	21.10	19.30	20.43
	(SD)	(1.39)	(7.70)	(6.81)	(8.31)	(2.18)	(6.06)
Female	*%*	11	14	11	12	13	60
	(n)	(11)	(14)	(11)	(12)	(13)	(61)
Male		9	7	9	8	7	40
		(9)	(7)	(9)	(8)	(7)	(40)
Caucasian		14	16	14	16	16	75
		(14)	(16)	(14)	(16)	(16)	(76)
Asian		3	2	4	1	1	11
		(3)	(2)	(4)	(1)	(1)	(11)
African American		1	2	2	2	2	9
		(1)	(2)	(2)	(2)	(2)	(9)
Hispanic		1	0	0	0	1	2
		(1)	(0)	(0)	(0)	(1)	(2)
American Indian/ Alaskan Native		1	0	0	0	0	1
		(1)	(0)	(0)	(0)	(0)	(1)
Native Hawaiian/ Other Pacific Islander		0	0	0	0	0	0
		(0)	(0)	(0)	(0)	(0)	(0)
[a]Other		0	1	0	1	0	2
		(0)	(1)	(0)	(1)	(0)	(2)

Table 2.1 (continued)

Condition	FR	KN	QR	VSP	WM	All Conditions
Demographic Characteristics						
Level of Education						
Freshman	12 (12)	10 (10)	11 (11)	12 (12)	13 (13)	57 (58)
Sophomore	4 (4)	3 (3)	3 (3)	2 (2)	3 (3)	15 (15)
Junior	2 (2)	3 (3)	4 (4)	3 (3)	0 (0)	12 (12)
Senior	1 (1)	4 (4)	1 (1)	1 (1)	3 (3)	10 (10)
Graduate	0 (0)	1 (1)	0 (0)	0 (0)	0 (0)	1 (1)
[b]Other	1 (1)	0 (0)	1 (1)	2 (2)	1 (1)	5 (5)
Disability Status						
[c]Disability	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	2 (2)
No Disability	20 (20)	20 (20)	20 (20)	19 (19)	20 (20)	98 (99)
Total *N*	20	21	20	20	20	101

Note. FR = Fluid Reasoning, KN = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory. Percentages have been rounded; therefore, the sum of subgroup percentages may not equal the total percentage for All Conditions.
[a]Participants in this category reported their ethnicity as: "multiracial" (KN) and "American Indian/Caucasian" (VSP).
[b]Participants in this category reported their level of education as: senior in high school (FR), Lifelong Education student (QR), Post-Baccalaureate Studies student and Lifelong Education student (VSP), and Post-Baccalaureate Studies student (WM).
[c]Participants in this category reported their disability status as a mild form of Charcot-Marie-Tooth disease (KN), and both partial deafness in one ear and a lazy left eye (VSP).

Study Design

This study used an experimental design in which participants were randomly assigned to one of five conditions defined by the five cognitive processes the SB5 intends to measure (i.e., Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual Spatial Processing, and Working Memory). The study involved a repeated measures design because participants answered selected items from both the Verbal and Nonverbal Scale subtests for their condition (see Table 2.2). Subtests were administered in a counterbalanced order within each condition to avoid order effects. However, due to error, the number of participants was exactly counterbalanced across examiners in only three of the five conditions. Participants answered 9 items from the Nonverbal Scale subtest and 9 items from the Verbal Scale subtest for their condition, for a total of 18 items. The dependent variables were verbal self-report descriptions and survey ratings of the degree to which the participant used the five cognitive processes the SB5 intends to measure, as well as the degree to which the participant used verbal mediation, when solving SB5 items.

Twenty participants were assigned to each condition with the exception of the Knowledge condition, which had 21 participants ($N = 101$). One of two examiners tested each participant. It was intended for the number of participants in each of the five conditions to be counterbalanced across the two examiners. However, due to error, the number of participants was exactly counterbalanced across examiners in only three of the five conditions. Each of the two trained research assistants administered test items and subsequently interviewed 10 participants assigned to the Fluid Reasoning, Quantitative Reasoning, and Working Memory conditions. For the Knowledge condition, each assistant

Table 2.2

SB5 Subtests Administered According to Participant Condition

Between Subjects (Random Assignment)		Repeated Measures (Counterbalanced Order)	
		Nonverbal Subtests	**Verbal Subtests**
	Fluid Reasoning	Object Series/Matrices (routing subtest)	Early Reasoning task Verbal Absurdities Verbal Analogies
	Knowledge	Procedural Knowledge (Level 3) Picture Absurdities (Level 4-5)	Vocabulary (routing subtest)
	Quantitative Reasoning	Quantitative Reasoning (Level 3-5)	Quantitative Reasoning (Level 3-5)
	Visual-Spatial Processing	Form Patterns (Level 3-5)	Position and Direction (Level 3-5)
	Working Memory	Block Span (Level 3-5)	Memory for Sentences (Level 3) Last Word (Level 4-5)

**Note.* The SB5 routing subtests include only one set of items, and thus, do not include levels in their design.

interviewed 8 and 13 participants, respectively. For the Visual Spatial Processing condition, each assistant interviewed 12 and 8 participants, respectively. The total number of participants interviewed by each research assistant was 50 and 51 participants, respectively (see Table 2.3). To examine whether there was an association between examiner and number of participants interviewed in each condition, a Pearson chi square was conducted. There was no difference between the number of participants interviewed in each condition by the two examiners, χ^2 (4, N = 101) = 1.98, p = .74.

Procedures

Two undergraduate students from Psychology 499 (Individual Study in Psychology) who were blind to the study's hypotheses were trained to administer selected items from the

Table 2.3

Number of Participants Interviewed by Each Examiner Within Each Condition

Condition	Examiner 1	Examiner 2
Fluid Reasoning	10	10
Knowledge	8	13
Quantitative Reasoning	10	10
Visual Spatial Processing	12	8
Working Memory	10	10
Total *N*	50	51

SB5. The research assistants were trained in test administration and interview procedures for all five of the cognitive process areas. Each research assistant completed test administration and interview procedures for participants in all five conditions.

As part of their training, assistants critiqued themselves using self-review videos, and I also reviewed the videos and provided feedback. Assistants were provided feedback concerning their use of follow-up questions to elicit detailed information, as well as accuracy of scoring and administration. Interrater agreement for trained examiners who administer the SB5 has been found to range from .74 to .97 with an overall median of .90, indicating good interrater reliability (Roid, 2003c). During the training, research assistants were also provided with specific follow-up questions to use during participant interviews.

Participants were randomly assigned to one of the five conditions defined by the five cognitive processes the SB5 intends to measure (i.e., Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual Spatial Processing, and Working Memory). Research assistants administered selected items from the SB5 and then interviewed examinees about the processes they used to solve the problems. Prior to beginning the study, participants reviewed and signed the Informed Consent Form (see Appendix C). Participants completed a demographic survey including name, age, gender, race/ethnicity, year in college, and disability status (see Appendix D). Participants were given a list that defines the five processes (i.e., Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory) and use of inner language (see Appendix A) and asked to read the list to cue them to consider which processes they might use in solving the tasks. They were encouraged to ask any questions they had about the definitions.

Participants responded to 9 items from each of the two subtests (Verbal and Nonverbal) used to measure the cognitive process in their condition, for a total of 18 items. Subtests were administered in a counterbalanced order within each condition to avoid order effects (except as explained previously). To include items that were representative of a range of difficulty levels, participants answered three items each from levels three, four, and five of each subtest with the exception of Level 3 of the Verbal Fluid Reasoning subtest because it involves a single task. The items administered were determined by the following method: The SB5 includes both routing and nonrouting subtests in its design. For the nonrouting subtests, six levels are clearly identified by the developers of the SB5. For the two routing subtests, which do not have predetermined levels, the routing tables were used to provide a general estimate of the items corresponding to levels three, four, and five. Due to the fact that the highest items corresponding to each level are typically less difficult than the items to which they would route (A.D. Carson, personal communication, November 8, 2005), the range of items for each level was shifted up three items. After items for the three levels of each subtest were identified, three items from each level were selected to represent low, medium, and high levels of difficulty (see Appendix E for further description).

Originally, I proposed that if participants incorrectly answered all three items in the third level, they would be given additional items from level two. Further, I proposed that if participants incorrectly answered all three items in level three or four (or level two if administered), then questions from the next highest level would not be administered. However, none of the participants were administered items from level two, and none of the participants answered all three items in the third or fourth levels incorrectly, so that all participants were administered items from levels three, four, and five. For items that had a time limit (according to standardized administration procedures), participants were given extra time as long as they were actively engaged in solving the problem.

After each subtest (i.e., after the 9 items were administered for one of the two subtests), participants explained in their own words how they solved the problems and then rated the degree to which they used each process (see Interview Protocol in Appendix F). First, participants explained how they solved the problems in their own words. Before asking participants to explain how they solved problems, research assistants reviewed the items by showing participants the test pages and/or stimuli (e.g., blocks), as well as verbally summarizing the items (e.g., “I asked you to tell me what a ‘dog’ means”). Then, the participants were asked, “How did you solve all of the items you just answered?” The research assistant ensured that the participant’s answer included at least five sentences in response to the question before moving on to the next subtest. Research assistants asked specific follow up questions (i.e., “What were you thinking while you were solving the problems?,” “Could you give me more detail?,” and “Did you use a certain strategy help you to solve the problems?”) to elicit additional information, as well as to clarify the meaning of responses that were unclear. If the participant did not provide at least five sentences in response to the follow-up questions, then the participant was asked to talk about how they solved a specific problem to elicit more detail from the participant. The assistant restated the participant’s answer to validate the participant’s answer and to elicit additional information before moving on to the next subtest. The participant’s answers to all of the questions were recorded verbatim.

Second, participants filled out a survey asking them to rate the degree to which they used the five cognitive processes, as well as the degree to which they used “inner language” (language as a mediator of psychological processes), using a Likert-type rating scale (see the following Measures section). The total time period for the testing and interview ranged from approximately 20 to 40 minutes, depending on the participant’s speed of responses during the testing and interview, as well as the type of subtest that was administered (i.e., the condition to which the participant was assigned).

Participants’ names and identifying information did not appear on the testing or interview data, nor were their names included in the database. Each participant was assigned an identification number, and only the participant's identification number appeared on the testing and interview data. Each participant's demographic survey and informed consent form, which included the participant's identification number as well as personal identifying information (i.e., name, gender, age, race/ethnicity, year in college, and disability status), was kept confidential and stored in a file cabinet separate from interview and testing data. See Appendix G for the IRB approval letter.

Measures

Participants reported the processes they used to answer the SB5 items via verbal self-report and structured survey. The following is a description of these measures, as well as the methods by which participants' responses were scored.

Verbal Self-Report

Description. Participants were asked to explain how they solved the SB5 items in their own words. After administering the 9 test items for one subtest, the research assistants reviewed the test items to cue participants' memory by showing them the item content (e.g., stimulus page) and summarizing the items (e.g., "I asked you to tell me what a 'dog' means"). Then, participants provided at least five sentences to answer the question, "How did you solve all of the items you just answered?" Participants' answers to the question were recorded verbatim. Each participant provided two open-ended comments; one comment for the Verbal subtest and one comment of the Nonverbal subtest.

Scoring. Participants' answers to the question, "How did you solve all of the items you just answered?" (N = 202) were printed separately onto half sheets of paper. Three upper-level psychology graduate students scored the answers by assigning them to categories. The raters were not informed of which subtest each participant had been given. It was originally proposed that each rater would score all 202 participant answers (i.e., 101 Nonverbal scale descriptions and 101 Verbal scale descriptions). However, the scoring procedure was changed due to the high frequency (96 of 202 self-reports) with which participants labeled their processes (e.g., reported using "fluid reasoning" when describing their response processes). If the participant used labels, the answer was scored twice (i.e., the version with labels and the version without labels; see description below). Due to this change in the scoring procedure, each participant's answer was scored by two raters instead of three raters. Each of the three raters was assigned one-third of the 202 items to score (Rater 1, *n* = 67; Rater 2, *n* = 68; Rater 3, *n* = 67), and also scored half of the items assigned to each of the other two raters (Rater 1, *n* = 67; Rater 2, *n* = 67; Rater 3, *n* = 68). Thus, each rater scored a total of 134 (Rater 1) or 135 (Rater 2, Rater 3) items (see Table 2.4).

Table 2.4

Items Scored by Each Rater

	Rater 1	Rater 2	Rater 3
Item Numbers	1-67	68-135	136-202
	68-101	1-33	34-67
	136-168	169-202	102-135
Total Items	134	135	135

Raters used a scoring rubric to guide them in assigning participants' responses to categories. Two scoring rubrics were developed; a Cognitive Process Scoring Rubric for the five cognitive processes and a Verbal Mediation Scoring Rubric for the level of verbal mediation (see Appendices H and I). The two undergraduate research assistants and I developed the scoring rubrics. Using the definitions provided to the participants, as well as information provided in the SB5 examiner's and technical manuals (Roid, 2003b; 2003c), the team sorted 20 random participant responses into each of the five cognitive process categories and 20 random participant responses into the nonverbal and verbal categories. After the sorting, the team collaborated to draft the two scoring rubrics by developing examples to serve as guidelines for how to assign responses to categories. The Cognitive Process Scoring Rubric also provided concise definitions of the five cognitive processes, and the Verbal Mediation Scoring Rubric stated that Verbal subtests should require inner language, but that Nonverbal subtests should require little or no inner language (based on information provided in the SB5 examiner's and technical manuals). The scoring rubrics provided the three graduate student raters with objective criteria to guide them during the sorting, and were included to increase the reliability with which the raters assigned responses to categories.

Three graduate student raters independently assigned participants' responses to categories. If the two raters disagreed on the category for the primary cognitive process or the category for the level of verbal mediation, the raters discussed and attempted to reach a consensus in person or via email on the final assignment of the answer to a category. In the case of responses where the two raters did not reach a consensus, I served as the final adjudicator by either agreeing with one of the raters or declaring the response unscorable.

Raters assigned cognitive process and verbal mediation scores using two different procedures. Each of these is described below.

1. *Cognitive process scoring procedure.* Subsequent to my proposal, I made the decision to remove the labels from participant self-report descriptions. A label was defined as using one or more of the following terms: “fluid reasoning,” “knowledge,” “quantitative reasoning” (or “quantitative”), “visual spatial processing” (or “visual spatial”), and “working memory.” The labels were removed in case, due to limited background knowledge in cognitive processes, participants incorrectly labeled the process they used. In other words, I removed the labels in case there was a discrepancy between the processes the participant described in their own words, and the processes the participant actually labeled in their statement. For example, a participant may state using “knowledge,” but then actually describe using a process more likely to be interpreted as a different cognitive process.

With this procedure, the number of times the participant’s answer was scored depended on whether the participant used labels when describing their response processes. If the participant did not use labels, then the raters scored the participant’s answer only once. If the participant labeled one or more cognitive process(es), and also described how the labeled process was used, then raters scored two different versions of the participant’s answer (one version with labels and one version without labels). The raters first scored a version of the description in which the labels were removed and replaced with blanks (e.g., “I used _______ because I used addition and subtraction”). Then, the raters scored the version of the description with all labels included (e.g., “I used quantitative reasoning because I used addition and subtraction."). In the few cases where the participant failed to describe how the labeled process was used (i.e., the participant only made a statement that a certain process was used, such as "I used quantitative reasoning"), the phrase or sentence referencing the label was completely removed from the participant's answer and was not shown to the raters.

The raters were asked to assign each self-report description to one of six cognitive process categories (i.e., "Fluid Reasoning," "Quantitative Reasoning," "Working Memory," "Knowledge," "Visual Spatial Processing," or "Other"). The raters used the following process to score the answers: If the participant used one or more of the five processes, first indicate each process that was used, and second indicate the primary process if it could be determined from the participant's response. If the participant did not use any of the five processes, or used processes that were ambiguous or unclear, then the answer would be scored as "Other."

For the primary cognitive process, scores of 1, 2, 3, 4, and 5, respectively, were assigned to the cognitive categories of *Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual Spatial Processing*, and *Working Memory* (no responses were scored as "Other"). For statements where more than one process was identified (i.e., secondary processes), a score was calculated to reflect the number of secondary processes identified. The total possible number of scores recorded depended on whether the response included label(s). If the response included label(s), the response could receive up to four scores (i.e., a primary and secondary process score for the version without labels, and a primary and secondary process score for the version with the labels included). If the response did not include labels, the response could receive up to two scores (i.e., a primary and secondary score). For the

Nonverbal subtests, 44.5 % (45) of the 101 participants gave one or more cognitive process labels. On the Verbal subtests, 50.49% (51) of the 101 participants gave one or more labels (see Appendix J for further description). Note: Only the score for the primary cognitive process was used in statistical analyses. For self-report descriptions with labels (n=96), in the 5 cases where there was a discrepancy between the primary score for the two versions (i.e., with labels and without labels), the score for the version without labels was included in the analyses. See Appendix K for the Cognitive Process Scoring Sheet.

2. *Level of verbal mediation scoring procedure.* After rating each self-report description according to cognitive process, the pieces of paper with participants' answers were shuffled to randomize the order, and the raters were asked to reassign each participant comment to one of three categories representing the level of verbal mediation ("Verbal," "Nonverbal," or "Other"). The phrase "inner language" was not removed for scoring purposes because the definition of this term is easily understandable even with limited background knowledge in cognitive processes and because of the low frequency with which participants used this label. Of the 101 participants, 5 participants used the label "inner language" for the Nonverbal subtests, and 18 participants used the label on the Verbal subtests (see Appendix J for further description). Each response received one score; scores of 1 and 2, respectively, were assigned to the verbal mediation categories of *Nonverbal* and *Verbal* (no responses were scored as "Other"). See Appendix L for the Verbal Mediation Scoring Sheet.

Inter-rater agreement. Inter-rater agreement was calculated for cognitive process and level of verbal mediation scores. For cognitive process scores, only the primary process score was used to calculate inter-rater agreement. For responses with labels (*n*=96), in the 5 cases where there was a discrepancy between the primary score for the two versions (i.e., with labels and without labels), the score for the version without labels was included in the analysis. The kappa coefficient for the graduate student raters' assignment of responses to categories (prior to adjudication by a third party) was k = .94. This suggests that graduate student raters agreed when assigning responses to categories. For level of verbal mediation scores, each response received one score (either verbal or nonverbal). The kappa coefficient for the graduate student raters' assignment of responses to level of verbal mediation categories (prior to adjudication by a third party) was k = .89. This suggests that raters agreed when assigning responses to categories.

It was originally proposed that the percentage of responses for which the raters assigned the same primary process score independently would be calculated. These data were collected for 135 of the responses, but were not collected for the remaining 67 responses. Thus, the percentage was not calculated because of the missing information.

Survey

Description. Participants rated the degree to which they used specific processes to solve the test items using a Likert-type scale (0-not at all, 1-a bit, 2-somewhat, 3-a lot, and 4-exclusively). Specifically, participants were asked to rate the degree to which they used each of the five cognitive processes, as well as "inner language," or language as a mediator of psychological processes (see Appendix B for the rating scale).

Scoring. Scores of 0, 1, 2, 3, and 4, respectively, were assigned to survey ratings of *Not At All, A Bit, Somewhat, A Lot,* and *Exclusively.* Separate scores were recorded for each process, and for the degree to which participants used “inner language” in developing their responses.

Statistical Analyses

Subsequent to proposing the thesis, I revised the method for evaluating some of the predictions due to the nature of the data and the statistical program (SPSS 14.0) used to run the statistical analyses. The original and revised analyses, as well as the rationale for why these changes were made, are presented below.

For Prediction One, the original analyses proposed were repeated measures MANOVAs, followed by post-hoc ANOVAs and *t*-tests with Bonferroni error correction. The MANOVA results did not allow me to directly test my “a priori” prediction that the survey rating for the intended process would be higher than the ratings for the other four processes (see results in Appendix M). I decided (in consultation with my advisor) to create an index of “dominance” score based on the participant’s ratings for the intended process given the participant’s condition. Each participant’s ratings for the five cognitive process survey items were transformed into a single “dominance” score. The dominance score was equal to the intended process survey score minus the mean of the other four unintended process survey scores. For example, for the Fluid Reasoning condition, the mean of the Knowledge, Quantitative Reasoning, Visual Spatial Processing, and Working Memory scores was subtracted from the participant’s rating for use of Fluid Reasoning processes. Thus, a positive dominance score (i.e., greater than 0) would indicate the participant reported using the intended process more than the other four processes, a dominance score of “0” would indicate no preference for the intended process (i.e., random use), and a negative dominance score (i.e., less than 0) would indicate the participant reported using other processes more than the intended process. The dominance scores were used to conduct one-sample *t*-tests and one-way ANOVAs with post-hoc *t*-tests with Bonferroni error correction.

The dominance scores allowed me to simplify the analyses so that I could easily test my hypotheses, although this method resulted in loss of data. For three of the subtests, the mean survey score for one or more unintended processes is higher than the mean score for the intended process, although the mean dominance score is positive. Whereas the dominance scores do not test the differences between the mean survey scores for each of the processes, they are a useful way to test the study's hypotheses (i.e., that participants will report using intended processes to a greater degree than unintended processes) without increasing the alpha error rate that would be generated by multiple (n = 32) *t*-tests or the 50 (5 x 5x 2) post-hoc tests generated by a repeated measures MANOVA.

For Prediction Four, a repeated measures MANOVA was originally proposed, followed by post-hoc ANOVAs. MANOVA results and effect sizes were sufficient to test these analyses. Effect sizes were calculated for the MANOVA results, as proposed, and additional effect sizes were calculated for each condition to examine the magnitude of the difference between the mean survey scores for the Verbal Scale and Nonverbal Scale subtests. Mean differences for the survey item scores for each condition were divided by the pooled *SD* of the score distribution to create effect sizes.

For Predictions Five and Six, a 3 ("nonverbal," "verbal," or "other" scores) by 2 (Nonverbal Scale subtest, Verbal Scale subtest) chi-square test of single proportions for the

Nonverbal Scale subtest, and then for the Verbal Scale subtests, was proposed. However, this analysis would include two scores from the same participant and, thus, would violate the assumption for chi-square analyses that all observations are independent. Thus, the analyses included a one-sample chi-square test of single proportions for the Nonverbal Scale subtest, and then for the Verbal Scale subtests. The analyses and results for these predictions are discussed in Chapter 3.

Further, it was originally proposed that eta squared (η^2) would be calculated for all significant MANOVAs. Due to limitations of the statistical software (SPSS 14.0), eta squared could not be calculated. Instead, partial eta squared (η_p^2) values provided by SPSS were provided. However, there is a lack of consensus about the interpretation of these values, and readers should use caution when interpreting them. The partial eta squared values reported were interpreted based on descriptors for Cohen's (1988) multiple regression coefficient (R^2), although the authors do recognize the values for these two measures of association do not correspond exactly (Pierce, Block, & Aguinas, 2004).

CHAPTER 3: ANALYSES AND RESULTS

In this chapter, the major hypotheses are evaluated by testing the predictions associated with each hypothesis. The chapter is organized according to the hypotheses proposed.

Hypothesis One: Examinees will report using the intended cognitive process more than they report using the other four cognitive processes that the subtest is not intended to measure.

This hypothesis was tested using *t*-tests and ANOVAs (Prediction One) and chi square analyses (Predictions Two and Three). Descriptive results are presented first, and then results relevant to each prediction follow. Participants' mean survey scores on the five survey items, according to the condition to which they were assigned, are presented in Table 3.1. Values in boxes represent the mean survey ratings for the intended cognitive process.

Prediction one. The mean rating for the degree to which examinees use the intended cognitive process will be higher than the mean rating of the other four cognitive processes that the subtest is not intended to measure. Evidence to support this prediction would manifest in the form of a statistically significant omnibus one-sample *t*-test using the dominance scores for all participants ($n = 101$) for the Verbal Scale subtests, as well as for the Nonverbal Scale subtests, at alpha = .05, assuming a directional (i.e., one-tailed) test. If significant, one-sample *t*-tests using dominance scores for each of the five conditions (between- subjects) for Verbal Scale subtests, and then for Nonverbal Scale subtests, will be conducted at alpha = .01, one-tailed (to reduce Type I error for multiple dependent contrasts). Mean dominance scores were divided by the *SD* of the score distribution to create

Table 3.1

Mean Response Values for Cognitive Process Survey Items for Participants Sampled Within Each Condition

Survey Item		FR	K	QR	VSP	WM
				M (*SD*)		
Level of Verbal Mediation						
				Fluid Reasoning		
	Nonverbal Subtest	2.80 (0.62)	0.85 (0.93)	0.95 (0.95)	3.05 (1.15)	2.15 (1.31)
	Verbal Subtest	2.20 (1.36)	3.15 (0.49)	0.00 (0.00)	2.55 (1.50)	1.90 (1.07)
				Knowledge		
	Nonverbal Subtest	1.05 (1.12)	3.33 (0.80)	0.48 (0.81)	1.48 (1.08)	1.05 (0.97)
	Verbal Subtest	1.67 (1.24)	2.76 (1.04)	0.05 (0.22)	0.90 (1.22)	1.29 (1.27)
				Quantitative Reasoning		
	Nonverbal Subtest	1.60 (1.57)	1.70 (1.22)	2.65 (0.99)	2.45 (0.89)	1.65 (0.81)
	Verbal Subtest	1.10 (1.25)	2.15 (1.18)	3.05 (0.39)	2.40 (0.88)	1.65 (0.88)

Table 3.1 (continued)

Survey Item	FR	K	QR	VSP	WM
			M (*SD*)		
Level of Verbal Mediation					
			Visual Spatial Processing		
Nonverbal Subtest	1.35 (0.93)	1.55 (1.19)	0.30 (0.73)	3.50 (0.61)	2.15 (1.09)
Verbal Subtest	1.30 (0.87)	2.85 (0.81)	0.15 (0.37)	2.25 (0.97)	1.90 (0.85)
			Working Memory		
Nonverbal Subtest	1.00 (1.17)	0.90 (1.29)	0.65 (0.93)	3.25 (0.72)	3.40 (0.68)
Verbal Subtest	0.90 (1.12)	1.20 (1.15)	0.05 (0.23)	1.65 (1.50)	3.45 (0.61)

Note. FR = Fluid Reasoning (*n*=20), K = Knowledge (*n*=21), QR = Quantitative Reasoning (*n*=20), VSP = Visual Spatial Processing (*n*=20), and WM = Working Memory (*n*=20). Mean ratings are based on a scale ranging from 0 to 4. Values in boxes represent the mean ratings/ standard deviations for the intended cognitive process.

effect sizes. Large (*ES* $\geq$ 0.80) effect sizes will be considered strong evidence for the prediction, medium (*ES* $\geq$ 0.50) effect sizes will be considered moderate evidence for the prediction, and small (*ES* $\geq$ 0.20) effect sizes will be considered weak evidence for the prediction.

Next, one-way ANOVAs using the dominance scores for all participants for the Verbal Scale subtests, and then for the Nonverbal Scale subtests, were conducted. When the main effect for condition (between- subjects) was significant, partial eta squared (η_p^2) values were calculated, and post-hoc *t*-tests with Bonferroni error correction were conducted to identify in which conditions participants were more or less likely to report using the intended process to respond to items. Strong evidence that participants' tendency to report using the intended process varies according to the condition (i.e., the cognitive process intended to be measured by the subtest) would be indicated by a large ($\eta_p^2 \geq 0.26$) effect size, moderate evidence would be indicated by a medium ($\eta_p^2 \geq 0.13$) effect size, and weak evidence would be indicated by a small ($\eta_p^2 \geq 0.02$) effect size. *Note:* Partial eta squared (η_p^2) values were interpreted based on descriptors for Cohen's (1988) multiple regression coefficient (R^2). Caution should be used when interpreting these values as the proportion of explained variance exactly (Pierce, Block, & Aguinas, 2004). Table 3.2 presents the mean dominance score for each condition for the Nonverbal and Verbal Scale subtests separately. Bold values indicate the value is significantly greater than 0 (i.e., supportive of Prediction One).

Omnibus one-sample *t*- tests for all participants (*n* = 101) for the Verbal Scale subtests, and then the Nonverbal Scale subtests, were conducted. For Verbal Scale subtests, the mean dominance score was significantly greater than the test value of 0, $t(100) = 9.59$, *p*

Table 3.2

Mean Dominance Scores for Cognitive Process Survey Items for Participants Sampled Within Each Condition

Condition	FR	K	QR *M* *(SD)* *ES*	VSP	WM	Overall Mean
Nonverbal Subtest	**1.05** (1.13) 0.93	**2.32** (0.96) 2.42	**0.80** (1.20) 0.67	**2.16** (0.80) 2.70	**1.95** (0.83) 2.35	**1.67** (1.16) 1.44
Verbal Subtest	0.30 (1.63) 0.18	**1.79** (1.44) 1.24	**1.23** (0.55) 2.24	**0.70** (0.96) 0.73	**2.48** (0.78) 0.78	**1.30** (1.37) 0.95

Note. FR = Fluid Reasoning (*n*=20), K = Knowledge (*n*=21), QR = Quantitative Reasoning (*n*=20), VSP = Visual Spatial Processing (*n*=20), and WM = Working Memory (*n*=20). Results are based on one-sample *t*-tests. Positive values indicate the intended process was rated higher than the unintended processes. Values in bold are significantly greater than 0 ($p < .01$).

$< .0001$ (one-tailed). The mean dominance score was positive (M = 1.30, SD = 1.37) and large (ES = 0.95) indicating that participants reported using the intended cognitive process more than unintended processes. For the Nonverbal Scale subtests, the mean dominance score was significantly greater than 0, $t(100) = 14.47$, $p < .0001$ (one-tailed). The mean dominance score was positive (M = 1.67, SD = 1.16) and large (ES = 1.44), again indicating that participants reported using the intended cognitive process more than unintended processes. These results provide strong support for Prediction One by suggesting that participants reported using the intended cognitive process more than unintended cognitive processes.

One-sample *t*- tests using dominance scores for each of the five conditions were conducted for Nonverbal Scale subtests, and then for the Verbal Scale subtests. Results support Prediction One for all subtests except the Fluid Reasoning Verbal Scale subtest. For 9 of the 10 subtests, the mean dominance score was greater than 0, indicating that participants had a tendency to report using the intended cognitive process more than unintended cognitive processes for those subtests. There was a large effect size (i.e., ES $\geq$ 0.80) for seven of the subtests and a medium effect size (i.e., ES $\geq$ 0.50) for two of the subtests (Quantitative Reasoning Nonverbal Scale and Visual Spatial Processing Verbal Scale subtests). However, for the Fluid Reasoning Verbal Scale subtest, participants did not indicate a clear preference for the intended cognitive process over the other four cognitive processes (*ES* = 0.18).

Next, one-way ANOVAs were conducted using dominance scores for the Verbal Scale subtests, and then for the Nonverbal Scale subtests. Post-hoc comparisons with Bonferroni error correction were also conducted. For the Verbal Scale subtests, there was a significant main effect for condition, $F(4, 96) = 11.43$, $p < .0001$, suggesting that participants' tendency to report using the intended cognitive process varied according to the condition to which they were assigned. There was also a large effect size for condition ($\eta_p^2 =$ 0.32). Post-hoc comparisons with Bonferroni error correction indicated that participants in the Working Memory and Knowledge conditions reported higher survey ratings for the intended cognitive process when compared to participants in other conditions. The mean Working Memory dominance score (*M* = 2.48) was higher than the mean dominance score for the Fluid Reasoning (*M* = 0.30), Quantitative Reasoning (*M* = 1.23), and Visual Spatial

Processing (*M* = 0.70) conditions. The mean Knowledge dominance score (*M* = 1.79) was higher than the mean dominance score for the Fluid Reasoning and Visual Spatial Processing conditions. The mean dominance scores for the Fluid Reasoning, Quantitative Reasoning, and Visual Spatial Processing conditions were not higher than the mean dominance scores for any other conditions.

For the Nonverbal Scale subtests, there was a significant main effect for condition, $F(4, 96) = 9.61$, $p < .0001$, indicating that participants' tendency to report using the intended cognitive process varied according to the condition to which they were assigned. There was also a large effect size for condition ($\eta_p^2 = 0.29$). Post-hoc comparisons with Bonferroni error correction indicated that participants in the Knowledge and Visual Spatial Processing conditions reported higher survey ratings for the intended cognitive process when compared to participants in other conditions. Participant ratings from the Fluid Reasoning and Quantitative Reasoning conditions indicated participants in these conditions were less consistent in reporting using the intended cognitive process than participants in other conditions. The mean Knowledge (*M* = 2.32) and Visual Spatial Processing (*M* = 2.16) dominance scores were both higher than the mean dominance score for the Fluid Reasoning (*M* = 1.05) and Quantitative Reasoning (*M* = 0.80) conditions. The mean Quantitative Reasoning dominance score (*M* = 0.80) was lower than the mean dominance score for the Knowledge, Visual Spatial Processing, and Working Memory (*M* = 1.95) conditions. The mean Fluid Reasoning dominance score (*M* = 1.05) was lower than the mean dominance score for the Knowledge and Visual Spatial Processing conditions.

Evidence suggests that participants' tendency to report using the intended process is strong, both in its statistical significance and effect size. However, the tendency varies according to the subtest/intended cognitive process. Evidence provided strong support for Prediction 1 for the Knowledge factor (both Verbal and Nonverbal Scale subtests), Working Memory Verbal Scale subtest, and the Visual Spatial Processing Nonverbal Scale subtest. Evidence provided moderate support for the validity of the Working Memory Nonverbal Scale subtest. Evidence provided the least support for the Fluid Reasoning subtests (both Verbal and Nonverbal Scale), Quantitative Reasoning subtest (both Verbal and Nonverbal Scale), and Visual Spatial Processing Verbal Scale subtest.

Prediction two. Participants' answers to open-ended questions about how they solved the problems on Nonverbal Scale subtests will reflect the cognitive process that the subtest is intended to measure more than the processes the subtest is not intended to measure. Evidence to support this prediction would manifest in the form of a significant χ^2 value for a 6 (5 cognitive process categories and one "other" category for unclassified responses) x 5 (experimental condition) chi square test of single proportions at alpha = .05. Because there were no responses assigned to the "other" process category, this analysis used a 5 x 5 chi square. Table 3.3 illustrates the number of items assigned by raters to each category according to the cognitive process intended to be elicited by the subtest.

Evidence provided moderate support for the prediction that participants' open-ended comments would reflect the intended cognitive process more than the unintended process for Nonverbal Scale subtests. A 5 x 5 chi-square test of single proportions indicated the proportion of scores assigned to each category was significantly different than would be

Table 3.3

Self-Reported Cognitive Process Scores for Participants Sampled Within Each Condition for Nonverbal Scale Subtests

Intended Process	FR	K	QR	VSP	WM	Row *N*
Identified Process						
FR	9	2	1	2	0	14
K	0	19	1	0	0	20
QR	0	0	13	0	0	13
VSP	11	0	4	18	4	37
WM	0	0	1	0	16	17
N	20	21	20	20	20	101

Note. FR = Fluid Reasoning, K = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory. Values in boxes represent the number of comments assigned to the intended cognitive process category for each condition.

expected based on chance, χ^2 (16, N = 101) = 234.40, $p < .001$. Thus, participants' self-reported cognitive processes varied according to the condition to which they were assigned. Approximately 74% (n = 75) of the participants reported using intended processes, although participants tended to report using Visual Spatial Processing more often than would be expected (n = 37).

Participants reported using the intended process to solve test items from the Nonverbal Scale Knowledge and Visual Spatial Processing subtests. However, participants responding to test items from the Nonverbal Scale Fluid Reasoning (11 participants), Quantitative Reasoning (4 participants), and Working Memory (4 participants) subtests frequently reported using Visual Spatial Processing rather than the intended process.

Prediction three. Participants' answers to open-ended questions about how they solved the problems on Verbal Scale subtests will reflect the cognitive process that the subtest is intended to measure more than the processes the subtest is not intended to measure. Evidence to support this prediction would manifest in the form of a significant χ^2 value for a 6 (5 cognitive process categories and one "other") x 5 (experimental condition) chi square test of single proportions at alpha = .05. One answer from a participant in the Quantitative Reasoning condition was considered unscorable due to lack of consensus among raters, and thus was excluded from the analyses. Because there were no responses assigned to the "other" category, this analysis used a 5 x 5 chi square. Table 3.4 illustrates the number of items assigned by raters to each category according to the cognitive process intended to be elicited by the subtest.

Evidence provided strong support for the prediction that participants' open-ended comments would reflect the intended cognitive process more than the unintended process for Verbal Scale subtests. A 5 x 5 chi-square test of single proportions indicated the proportion of scores assigned to each category was significantly different than would be expected based on chance, $\chi^2(16, N = 100) = 320.50, p < .001$. This suggests that participants' self-reported cognitive processes varied according to the condition to which they were assigned. Ninety percent (n = 90) of the participants reported using intended processes.

Participants reported using the intended process to solve test items from the Verbal Scale Knowledge, Quantitative Reasoning, and Working Memory subtests. However,

Table 3.4

Self-Reported Cognitive Process Scores for Participants Sampled Within Each Condition for Verbal Scale Subtests

Intended Process	FR	K	QR	VSP	WM	Row *N*
Identified Process						
FR	15	0	0	0	0	15
K	5	21	0	3	0	29
QR	0	0	17	0	0	17
VSP	0	0	2	17	0	19
WM	0	0	0	0	20	20
N	20	21	19	20	20	100

Note. FR = Fluid Reasoning, K = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory. One item from the Quantitative Reasoning condition was unscorable due to lack of consensus among raters and thus was excluded from the results. Values in boxes represent the number of comments assigned to the intended cognitive process category for each condition.

participants responding to test items from the Verbal Scale Fluid Reasoning (5 participants) and Visual Spatial Processing (3 participants) subtests frequently reported using unintended Knowledge processes.

Hypothesis Two: Examinees will report using primarily verbal mediation in response to Verbal Scale subtests and avoid using verbal mediation in response to Nonverbal Scale subtests.

First, descriptive results are presented, and then results relevant to each prediction follow. The evidence for this hypothesis was evaluated using a repeated measures MANOVA (Prediction Four) and chi square analyses (Predictions Five and Six). The mean scores for the inner language survey item for Nonverbal and Verbal Scale subtests, as well as the effect sizes for each condition, are presented in Table 3.5. Mean differences for the survey item scores for each condition were divided by the pooled *SD* of the score distribution to create effect sizes.

Table 3.5

Mean Response Values for the Inner Language Survey Item for Participants Sampled Within Each Condition

Condition	FR	K	QR *M* *(SD)*	VSP	WM	Overall Mean
Nonverbal Subtest	0.80 (1.28)	1.10 (1.22)	0.95 (1.19)	0.85 (1.04)	0.75 (1.07)	0.89 (1.15)
Verbal Subtest	2.45 (1.23)	2.19 (1.44)	1.40 (1.19)	1.95 (1.40)	2.70 (1.49)	2.14 (1.40)
Effect Size	1.13	0.70	0.45	0.80	1.52	0.88

Note. FR = Fluid Reasoning, K = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory. Values represent mean ratings on a scale ranging from 0 to 4. Higher scores reflect higher degrees of reported verbal mediation.

Prediction four. The mean rating across all conditions for the degree to which examinees use verbal mediation will be higher for Verbal Scale subtests than the mean rating for Nonverbal Scale subtests. Evidence to support this prediction would manifest in the form of a significant main effect for the type of subtest (i.e., Verbal or Nonverbal) on the survey ratings based on a 5 (between- subjects) x 2 (within- subjects) repeated measures MANOVA at alpha = .05. Should a main effect be significant, partial eta squared (η_p^2) will be calculated. Strong evidence that participants tended to report using more verbal mediation for Verbal Scale subtests than for Nonverbal Scale subtests would be indicated by a large ($\eta_p^2 \geq 0.26$) effect size, moderate evidence would be indicated by a medium ($\eta_p^2 \geq 0.13$) effect size, and weak evidence would be indicated by a small ($\eta_p^2 \geq 0.02$) effect size. *Note:* Partial eta squared (η_p^2) values were interpreted based on descriptors for Cohen's (1988) multiple regression coefficient (R^2). Caution should be used when interpreting these values as the proportion of explained variance exactly (Pierce, Block, & Aguinas, 2004).

To examine the support for this prediction for each condition, effect sizes will be calculated. For each condition, large ($ES \geq 0.80$) effect sizes will be considered strong evidence for the prediction, medium ($ES \geq 0.50$) effect sizes will be considered moderate evidence for the prediction, and small ($ES \geq 0.20$) effect sizes will be considered weak evidence for the prediction. Table 3.6 presents the repeated measures MANOVA results for the inner language survey item scores.

Evidence provided strong support for this prediction. A repeated measures MANOVA showed a significant main effect for inner language use in response to verbal vs. nonverbal subtests (within-subjects), $F(1, 96) = 86.87$, $p < .0001$, and a significant interaction

Table 3.6

Repeated Measures Multivariate Analysis of Variance for Inner Language Survey Item Scores

Source	*df*	*F*	η_p^2	*p*
	Between subjects			
Condition	4	0.89	.04	.48
Condition X Subtest	4	3.69*	.13	.01
S within- group error	96	(2.29)		
	Within subjects			
Subtest	1	3.69*	.48	.00
S within- group error	96	(0.91)		

Note. Values in parentheses represent mean square errors. S = subjects.
* $p < .05$.

between subtest and condition (between-subjects), $F(4, 96) = 3.69$, $p < .01$. The main effect for condition was not significant, $F(4, 96) = 0.89$, $p > .05$. Further, there was a large effect ($\eta_p^2 = 0.48$) for verbal vs. nonverbal subtests, and medium effect ($\eta_p^2 = 0.13$) for the interaction between subtest (within- subjects) and condition (between- subjects). Participants reported using more inner language for the Verbal Scale subtests (*M* = 2.14, *SD* = 1.40) than for the Nonverbal Scale subtests (*M* = 0.89, *SD* = 1.15), as predicted. The level of verbal mediation intended to be measured by the subtest explains a substantial proportion of the variance in scores (which supports Prediction Four), although the interaction indicates this effect is inconsistent across conditions (also see Table 3.5).

To examine the support for this prediction for each condition, effect sizes were calculated (see Table 3.5). There was a large effect size (i.e., $ES \geq 0.80$) for the Fluid Reasoning, Visual Spatial Processing, and Working Memory conditions, indicating that participants in these conditions consistently reported using more verbal mediation for the Verbal Scale subtest than the Nonverbal Scale subtest. There was a medium effect size (i.e., $ES \geq 0.50$) for the Knowledge condition, providing moderate support for the prediction for this condition. There was a small effect size (i.e., $ES \geq 0.20$) for the Quantitative Reasoning condition, suggesting that participants in this condition did not consistently report using more verbal mediation for the Verbal Scale subtest than for the Nonverbal Scale subtest.

Predictions five and six. Participants' answers to open-ended questions about how they solved problems on Verbal and Nonverbal Scale subtests will reflect the level of verbal mediation (i.e., Verbal or Nonverbal) that the subtest is intended to measure. Evidence to support this prediction would manifest in the form of significant chi-square one-sample tests of single proportions for the Nonverbal Scale scores and Verbal Scale scores at alpha = .05. Table 3.7 illustrates the self-reported verbal mediation scores for Nonverbal and Verbal Scale subtests.

Evidence provided moderate to strong support for the prediction that participants' open-ended answers on Nonverbal and Verbal Scale subtests would reflect the intended level of verbal mediation. For Nonverbal Scale subtests, a chi-square one-sample test of single

Table 3.7

Self-Reported Verbal Mediation Scores for Nonverbal and Verbal Scale Subtests

Identified Level Of Verbal Mediation	**Verbal Subtests**	**Nonverbal Subtests**
Verbal	81	6
Nonverbal	20	95
N	101	101

Note. No responses were assigned to the "other" category.

proportions indicated the proportion of scores assigned to each category (Verbal, Nonverbal) was significantly different than would be expected based on chance, χ^2 (1, N = 101) = 78.43, $p < .001$. Approximately 94% (n = 95) of participants reported using the intended level of verbal mediation. For the Verbal Scale subtests, a chi-square one-sample test of single proportions indicated the proportion of scores assigned to each category (Verbal, Nonverbal) was also significantly different than would be expected based on chance, χ^2 (1, N = 101) = 36.84, $p < .001$. Approximately 80% (n = 81) of participants reported using the intended level of verbal mediation.

Given that 20% of participants failed to report using verbal processes in response to Verbal Scale subtests, the number of participants within each condition who reported the intended level of verbal mediation was examined for Verbal Scale subtests. The results for the Nonverbal Scale subtests are presented as well. Table 3.8 reports the degree to which spontaneous comments identified verbal mediation within Verbal Subtests according to condition. Table 3.9 reports the degree to which spontaneous comments identified verbal

Table 3.8

Self-Reported Verbal Mediation Scores for Participants Sampled Within Each Condition for Verbal Scale Subtests

Condition	FR	K	QR	VSP	WM	Row *N*
Identified Level Of Verbal Mediation						
Nonverbal	3	0	10	7	0	20
Verbal	17	21	10	13	20	81
N	20	21	20	20	20	101

Note. FR = Fluid Reasoning, K = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory.

Table 3.9

Self-Reported Verbal Mediation Scores for Participants Sampled Within Each Condition for Nonverbal Scale Subtests

Condition	FR	K	QR	VSP	WM	Row *N*
Identified Level Of Verbal Mediation						
Nonverbal	19	19	20	20	17	95
Verbal	1	2	0	0	3	6
N	20	21	20	20	20	101

Note. FR = Fluid Reasoning, K = Knowledge, QR = Quantitative Reasoning, VSP = Visual Spatial Processing, and WM = Working Memory.

mediation within the Nonverbal Subtests according to condition. As can be seen from the tables, participants who responded to test items from the Verbal Scale Quantitative Reasoning and Visual Spatial Processing subtests frequently failed to report using verbal mediation processes. However, the Nonverbal Scale subtests generally elicited non-language based processes as intended.

Summary

Evidence provided strong support for both hypotheses for this study. Data generally provide evidence showing SB5 subtests elicit the targeted cognitive process, and not other processes, although some subtests are less likely to induce respondents to identify intended processes than others. For Hypothesis One (i.e., examinees will report using the intended cognitive process more than they report using the cognitive processes that the subtest is not intended to measure), there was strong support for Prediction 1, moderate support for Prediction 2, and strong support for Prediction 3. For Hypothesis Two (i.e., examinees will report using primarily verbal mediation in response to Verbal Scale subtests and avoid using verbal mediation in response to Nonverbal Scale subtests), there was strong support for Prediction 4 and moderate to strong support for Predictions 5 and 5. In the following chapter, the evidence for the response processes validity of the SB5 tests, hypotheses for the (relatively) poor results for some subtests, and evidence for the degree to which surveys and interviews can be used as a measure of response processes will be discussed. Directions for future research will also be discussed.

CHAPTER 4: DISCUSSION

This is the first study to examine directly the response process evidence for the validity of the SB5, the only intelligence test developed to measure CHC theory abilities using nonverbal and verbal subtests. Although not without exceptions, data provide evidence for the response processes validity of the SB5 subtests. Survey ratings and participant comments provided moderate to strong validity evidence in suggesting participants generally use intended cognitive processes when responding to test items, and generally avoid using unintended cognitive processes. Further, this is one of the few studies to examine the feasibility of using self-report measures to assess response processes. The generally strong results indicate that the survey and interview used in this study are useful tools to assess cognitive processes and verbal mediation processes.

Results provide strong support for the validity of the SB5 subtests, although results suggest some SB5 subtests are less likely to induce respondents to identify intended processes than others. Data provide less support for the Fluid Reasoning Verbal and Nonverbal Scale subtests, Visual Spatial Processing Verbal Scale subtest, Working Memory Nonverbal Scale subtest, and Quantitative Reasoning Nonverbal Scale subtest. The relatively poor results could be due to limitations of the SB5 (i.e., it does not elicit the constructs intended), limitations of examinees' ability to identify and describe their response processes, or limitations of the method used to identify intended constructs (i.e., the interviews do not elicit the right information). In the following paragraphs, these issues, as well as directions for future research, are discussed.

Limitations of the SB5

One hypothesis for the relatively poor results is that certain SB5 tasks may have elicited unintended processes and/or actually failed to elicit the intended processes. According to Samuel Messick (1994, 1995), these two sources of invalidity would be described as construct irrelevant variance and construct underrepresenation, respectively. The Fluid Reasoning Verbal Scale subtest (Early Reasoning, Verbal Absurdities, and Verbal Analogies activities) and Visual Spatial Processing Verbal Scale subtest (Position and Direction) may have failed to elicit the cognitive process they were intended to measure. These subtests elicited frequent reports of Knowledge processes according to both survey and interview data. The Fluid Reasoning subtest includes the Early Reasoning (card sorting activity with pictures of everyday objects), Verbal Absurdities (examinee describes what is “silly or impossible” about scenarios that could not logically occur in everyday life), and Verbal Analogies activities. The Visual Spatial Processing subtest requires the examinee to demonstrate basic spatial concepts such as “behind” using gestures and to explain spatial orientations and directions (e.g., North and South). It may be that these tasks did actually elicit Knowledge processes because they required examinees to use information they learned from experiences in everyday life at home or school, such as how to give someone directions. If so, then participants accurately reported that these tasks elicited unintended Knowledge processes (i.e., construct irrelevant variance). If participants did use Knowledge processes for the Fluid Reasoning subtest, they may not have perceived the tasks to involve “novel” pieces of information; thus, the subtest may have failed to elicit Fluid Reasoning processes (i.e., construct underrepresentation), instead eliciting an unintended process (i.e., construct-irrelevant variance). Alternatively, another possible explanation is that examinees misinterpreted the processes they used for some tasks, which will be discussed next.

Limitations of Examinees

Another hypothesis that may explain the relatively poor results for some SB5 subtests is that examinees may have difficulty identifying and describing their response processes, as some cognitive processes may be more easily identifiable than other processes. Examinees may have difficulty identifying their response processes for verbal tasks (i.e., tasks that elicit language processes) because they misinterpret these tasks as eliciting other processes such as knowledge-based processes. Perhaps examinees have difficulty monitoring their strategies and reactions when responding to certain (difficult) items because their cognitive resources are taxed by solving these items. Alternatively, examinees may have the cognitive resources necessary, but their cognitive processes on certain (easy) items may be so automatic they are unable to identify and discriminate which processes they are using. In addition, participants may have misunderstood or misinterpreted the definitions of the cognitive constructs they were provided. Specifically, they may have omitted certain parts of the definition when attempting to identify whether a certain process applied to how they responded to the test items.

It may be that participants misinterpreted their cognitive processes as "Knowledge" processes on the Fluid Reasoning Verbal Scale subtest and Visual Spatial Processing Verbal Scale subtest. These Verbal Scale tasks, such as verbal analogies and giving directions, require knowledge of definitions and directional terms. It is true that both knowledge processes and language are learned through everyday experiences. Participants may have misidentified their language-based processes on these Verbal Scale subtests as being knowledge-based processes. The use of highly verbally loaded tasks is common among intelligence tests, and it may be that some degree of knowledge is elicited in most verbally loaded tasks. Alternatively, as discussed earlier, it may be that participants accurately reported that these subtests do actually elicit Knowledge processes. I will suggest future research that could help generate evidence to test these competing explanations.

The Working Memory and Quantitative Reasoning Nonverbal Scale subtests (Block Span and Quantitative Reasoning) elicited frequent reports of Visual Spatial Processing. Participants frequently reported using Visual Spatial Processing for these subtests on the survey, and when asked to describe their response processes, 4 of 20 participants each in the Working Memory and Quantitative Reasoning conditions described using Visual Spatial Processing for the Nonverbal Scale subtest. The tendency to under-report use of the intended cognitive process for these subtests may have occurred because examinees have difficulty identifying and describing the processes they use to solve language-reduced items. Perhaps participants experienced difficulty monitoring their strategies and reactions on difficult test items because their cognitive resources were taxed by solving these items. Additional cognitive resources above and beyond simply solving test items may not have been available when participants attempted to identify their response processes for challenging items.

Another explanation is that examinees may have the cognitive resources necessary, but their cognitive processes on certain items (i.e., items that were not age-appropriate) may have been so automatic they were unable to identify and discriminate which processes they were actually using. It appears that participants consistently reported the processes they *perceived* themselves as using; participants both identified (on the survey) and then described (during the interview) use of Visual Spatial Processing for these subtests. On the less difficult Quantitative Reasoning items where the examinee compares or counts blocks and the Working Memory items where the examinee taps 2-3 blocks in order, participants may have perceived themselves as using Visual Spatial Processing due to the automaticity with which they could solve these items. However, on the more difficult items such as Quantitative Reasoning items where examinees deduce numerical equations underlying illustrations and then identify the last figure in a series, and Working Memory items where the examinee taps 5-7 blocks in order, participants may have used more complex, labored processes that they were able to successfully identify as the intended process. This hypothesis seems plausible, as participants did frequently report using the intended process, in addition to Visual Spatial Processing, according to both survey and interview results. However, future research that systematically examines the processes used on easy versus hard items would help evaluate this hypothesis and rule out the possibility that some items on these subtests may elicit unintended (visual spatial) processes.

The Fluid Reasoning Nonverbal Scale subtest (Object Series/ Matrices) elicited frequent reports of Visual Spatial Processing. On the survey, participants reported using Visual Spatial Processing and Fluid Reasoning processes most frequently. It may be that the Fluid Reasoning subtest elicited reports of Visual Spatial Processing on the survey because participants misunderstood or misinterpreted the definitions of the cognitive constructs they were provided. The definitions for Fluid Reasoning (the ability to determine the underlying rules and relationships among pieces of novel information) and Visual Spatial Processing (the ability to see patterns, relationships, positions in space, or the whole picture among pieces of a visual display) are similar in that they both refer to finding patterns or relationships among pieces of information. The frequent reports of Fluid Reasoning and Visual Spatial Processing suggest that participants did process the patterns or relationships in the figures. However, the reference to the use of "novel information" in the Fluid Reasoning definition may have been overlooked or omitted by participants when they attempted to identify their cognitive processes. This hypothesis is supported by the fact that participants did not report using Knowledge processes, which suggests they may have, in fact, perceived these items to involve "novel" information. Thus, it may be that participants did not distinguish between processes involving "novel" information (or Fluid Reasoning processes), and processes that simply involved finding patterns or relationships among pieces of information.

Limitations of the Method

Relatively poor results for certain subtests may be due to limitations of the interview used to identify the intended constructs in this study (i.e., the interviews did not elicit the right information). Overall, results suggest that the survey used in this study is a useful tool for assessing cognitive processes and verbal mediation processes. Although survey results for some subtests were less consistent with the hypotheses, these findings are likely due to limitations of the SB5 or examinees, rather than a failure to accurately operationalize and describe intended processes. The definitions of the five processes and verbal mediation were taken from the SB5 manuals, and were reviewed by the author of the SB5, Gale Roid. The definitions provided on the survey are therefore likely to be valid descriptions of the processes intended to be measured by SB5 subtests. However, the interview questions used in this study may not have been specific enough to elicit a complete description of all response processes (i.e., cognitive and language-based processes). More specific interview questions may elicit valid descriptions of examinees' response processes.

During the interview, participants reported using little or no verbal mediation processes for Nonverbal Scale subtests. However, when asked to describe their response processes on Verbal Scale subtests, participants had a tendency to under-report the use of language-based processes, particularly for the Quantitative Reasoning and Visual Spatial Processing Verbal Scale subtests. Ten of 20 participants reported using little or no verbal mediation on the Quantitative Reasoning Verbal Scale subtest, and 7 of 20 participants reported using little or no verbal mediation on the Visual Spatial Processing Verbal Scale subtest. One possible explanation for participants' tendency to under-report the use of language processes is that they may have experienced difficulty describing their language-based processes. Through everyday experience, language processes have become automatic, and participants may not have been aware they were using language processes. However, it seems most likely that the tendency to under-report the use of language processes was largely due to limitations of the interview questions used in this study. Participants were asked to provide a general description of how they solved items and were not specifically asked to describe whether they used verbal mediation. It may be that participants were aware of their language processes, but did not report using language perhaps because they were not asked or because they assumed it was obvious language was used. The interview questions may not have been specific enough to elicit descriptions of language-based processes. This explanation seems the most plausible, as participants frequently reported using verbal mediation for these subtests according to a different measure (i.e., the survey). Examinees may be more likely to report use of language processes when interview questions make specific reference to the use of language-based processes. Thus, the interview shows promise as a measure of verbal mediation processes, although future research is needed in this area.

Generalizability is also an issue. This study's sample was limited to mainly Caucasian females in college, and results could differ for individuals from different populations. Likewise, generalization to other tests and processes is not possible.

Finally, it is possible the experimental procedures inappropriately "cued" participants to respond in ways congruent with the hypotheses. I provided participants definitions of the processes prior to interviewing them, which may have affected their responses. The definitions were introduced to participants as processes people sometimes use when solving test items, and were provided to increase awareness of cognitive response processes. This procedure may have cued participants to describe their processes in terms of the definitions they were provided, which may have affected the type and range of processes reported during the interview. In the next section, I will suggest future research to provide important insight into the validity of the SB5 and to extend the methods developed in my study.

Directions for Future Research

The results of this study provide interesting insight into the study's research questions, though there are a number of unresolved issues that lend themselves to future research. Researchers should address how to use interview questions to elicit complete descriptions of examinees' response processes (i.e., all aspects of their processes, as well as their use of language-based processes). Interview questions that explicitly ask examinees to describe whether they used verbal mediation (e.g., Describe whether or not you used inner language to solve the items you just answered) may elicit more valid descriptions of the use of these processes. Specific follow-up questions as they relate to the cognitive constructs should facilitate a complete discussion of important aspects of the examinee's processes. Research should also examine whether providing examinees definitions of processes may cue them to respond in a certain way to interview questions.

In addition, research should examine the feasibility of using other verbal methods such as written discussion, and nonverbal methods such as error analysis, neuropsychological measures (e.g., PET scans, fMRIs), and eye gaze methods, to assess and infer examinees' response processes. Talk aloud protocols may not be feasible, especially for assessing nonverbal processes. Talk aloud approaches increase the demand on examinees' working memory (which may reduce accuracy of selecting and reporting strategies) and may affect strategy use (e.g., demanding participants talk aloud while addressing nonverbal tasks may influence them to use verbal in place of nonverbal processes). Given the limitations of certain self-report measures such as interviews, obtaining information through multiple, and nonverbal, methods should contribute important information on the processes a particular task elicits.

Future research should also examine whether examinee training increases the accuracy of self-report measures. Training examinees to consider all aspects of the definition of a construct and to appreciate the similarities and differences among various constructs may improve the accuracy of survey results. Future research should also examine the effect item difficulty has on reports of response processes. Examiners could elicit self-report (via survey and interview) about response processes for items typically administered to younger individuals to see if examinees report different processes for items that are not developmentally appropriate. Perhaps easier items elicit automatic processes that are difficult for examinees to correctly identify or perhaps these items actually elicit unintended response processes.

Future research should also examine the generalizability of this study's results to a more diverse sample including males, individuals of different ethnicities, school-age children, and individuals who are hearing impaired, have language deficits, or who have not been exposed to the dominant language of their culture of residence. Research with school-age children would provide particularly important insight into the validity of the SB5 given the common use of intelligence tests such as the SB5 to assess learning difficulties in this population. However, extension of these methods to that population would have to assume that middle and high school students have the metacognitive skills necessary to monitor and accurately report their response processes. Given this study's results, it may be difficult for school-age children to report their response processes on verbal subtests, but perhaps they could accurately describe their processes for nonverbal subtests. In addition, research with other tests of intelligence would be useful. Research with nonverbal (or language-reduced) tests of intelligence would be important to provide evidence that these tests do not elicit language processes in examinees. Research could also examine how to measure the response processes of individuals to whom these tests are typically administered including individuals who are hearing impaired, have language deficits, or who have not been exposed to the dominant language of their culture of residence. If the response processes of these individuals can be assessed, research should examine the evidence for the validity of nonverbal tests with this population (see Braden & Anathasiou, 2005). Also, research that extends this study's methods to other intelligence tests purporting to measure the same CHC (Cattell-Horn-Carroll) constructs would be particularly important, as evidence that examinees are using these abilities would further support the validity of the CHC theory of intelligence.

Conclusions

In summary, the data in this study generally provide evidence for the response processes validity of the SB5 subtests. Although not without exceptions, survey ratings and participant comments were congruent with the cognitive and verbal mediation processes targeted by the SB5 subtests. There are many possible explanations for the occasional lack of congruence between the intended and identified processes for certain SB5 subtests. It may be that examinees have difficulty accurately identifying and describing their response processes due to limited cognitive resources, or conversely, the automaticity of cognitive and language processes. Or, poor results could be due to limitations of the measures (interview and survey) used to assess examinees' response processes. Examinees may need more specific questions to cue them to provide a complete report of their response processes. Alternatively, participant ratings and reports may have been accurate, and certain SB5 subtests may elicit unintended processes (construct irrelevant variance), or fail to elicit the intended process (construct underrepresentation).

This is one of the few studies to examine the feasibility of using self-report measures as a form of response processes evidence for the validity of an intelligence test. Further, this is the first study to examine the response process evidence for the validity of the SB5, the only intelligence test developed to measure CHC theory abilities using nonverbal and verbal subtests. These results can be used to guide future investigations of the response processes evidence for the validity of other intelligence tests, as well as to guide research on the use of self-report measures to assess examinees' response processes. This research is an important first step in helping test developers and test users address the paucity of direct evidence showing tests elicit intended response processes, and avoid eliciting unintended processes. Test users and test developers should use research in this area to understand the degree to which test developers' claims about the interpretation and use of test scores are supported, as professionals need this information to ethically base decisions on test results.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Braden, J.P., & Athanasiou, M.S. (2005). A comparative review of nonverbal measures of intelligence. In D.P. Flanagan & P.L. Harrison (Eds.). *Contemporary Intellectual Assessment: Theories, Tests, and Issues, 2nd ed.* (pp. 557-577). New York: Guilford Press.

Braden, J.P., & Niebling, B.C. (2005). Using the joint test standards to evaluate the validity evidence for intelligence tests. In D.P. Flanagan & P.L. Harrison (Eds.). *Contemporary Intellectual Assessment: Theories, Tests, and Issues, 2nd ed.* (pp. 615-630). New York: Guilford Press.

Braden, J. P. (2000). Editor's introduction: Perspectives on the nonverbal assessment of intelligence. *Journal of Psychoeducational Assessment, 18,* 204-210.

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,* 404-431.

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

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Craig, M.T., & Yore, L.D. (1995). Middle school students' metacognitive knowledge about science reading and science text: An interview study. *Reading Psychology, 16,* 169-213.

De Simone, C., & Schmid, R.F. (2004). Networking from the inside out: Understanding learners' processes, activities, and experiences. *Educational Research and Evaluation, 10,* 523- 549.

Gordon, E. D., & Holden, R. R. (1996). Use of item ratings to examine personality test item cognitive response processes. *Personality and Individual Differences, 21,* 897-905.

Houlihan, M., Campbell, K., & Stelmack, R. M. (1994). Reaction time and movement time as measures of stimulus evaluation and response processes. *Intelligence, 18,* 289-307.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8,* 441-480.

Madden, D. J., Gottlob, L. R., Denny, L. L., Turkington, T. G., Provenzale, J. M., Hawk, T. C., et al. (1999). Aging and recognition memory: Changes in regional cerebral blood flow associated with components of reaction time distributions. *Journal of Cognitive Neuroscience, 11,* 511-520.

McCallum, R. S., & Bracken, B. A. (1997). The Universal Nonverbal Intelligence Test. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.). *Contemporary Intellectual Assessment.* (pp. 268-280). New York: Guilford Press.

McCrindle, A.R., & Christensen, C.A. (1995). The impact of learning journals on metacognitive and cognitive processes and learning performance. *Learning and Instruction, 5,* 167-185.

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, 2nd ed.* (pp. 136-181). New York: Guilford Press.

Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment, 10,* 1-9.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Messick, S. (2000). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. In R.D. Goffin, & E. Helmes (Eds.) *Problems and Solutions in Human Assessment.* (pp. 3-20). Boston: Kluwer Academic.

Nuthall, G. (2000). The role of memory in the acquisition and retention of knowledge in science and social studies units. *Cognition and Instruction, 18,* 83-139.

Pierce, C.A., Block, R.A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64,* 916-924.

Ridderinkhof, K. R., Van der Molen, M. W. (1995). A psychophysiological analysis of developmental differences in the ability to resist interference. *Child Development, 66,* 1040-1056.

Roid, G.H. (2003a). *Stanford Binet Intelligence Scales, Fifth Edition.* Itasca, IL: Riverside Publishing.

Roid, G.H. (2003b). *Stanford Binet Intelligence Scales, Fifth Edition, examiner's manual*. Itasca, IL: Riverside Publishing.

Roid, G.H. (2003c). *Stanford Binet Intelligence Scales, Fifth Edition, technical manual*. Itasca, IL : Riverside Publishing.

Schumacher, E. H., & D'Esposito, M. (2002). Neural implementation of response selection in humans as revealed by localized effects of stimulus-response compatibility on brain activation. *Human Brain Mapping, 17,* 193-201.

Snow, R. E., & Lohman, D. F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.) *Educational Measurement (3rd ed.)* (pp. 263-331). New York: American Council on Education.

Welch, M. (1998). Students' use of three-dimensional modeling while designing and making a solution to a technological problem. *International Journal of Technology and Design Education, 8,* 241-260.

Appendixes

Appendix A

Participant Interview Form Definitions

Fluid Reasoning: I determined the underlying rules or relationships among pieces of information I had not seen before

Quantitative Reasoning: I reasoned with numbers or solved numerical problems

Knowledge: I applied or used the knowledge that I have previously acquired at home, school, or work

Working Memory: I temporarily held information in mind *and* then transformed or sorted it in my memory

Visual-Spatial Processing: I saw or imagined patterns, relationships, positions in space, or the "whole picture" among pieces of a visual display

Inner Language: I thought to myself using words or sentences to help me solve the problems

Appendix B

Participant Interview Form

Participant ID _____
Subtest ___________

Directions: Using the descriptions below, please rate the degree to which you used each process in responding to test items. Circle the number of the category that best describes the degree to which you used the process on the left.

	0 Not at all	1 A bit	2 Somewhat	3 A lot	4 Exclusively
Fluid Reasoning: I determined the underlying rules or relationships among pieces of information I had not seen before	0	1	2	3	4
Quantitative Reasoning: I reasoned with numbers or solved numerical problems	0	1	2	3	4
Knowledge: I applied or used the knowledge that I have previously acquired at home, school, or work	0	1	2	3	4
Working Memory: I temporarily held information in mind *and* then transformed or sorted it in my memory	0	1	2	3	4
Visual-Spatial Processing: I saw or imagined patterns, relationships, positions in space, or the "whole picture" among pieces of a visual display	0	1	2	3	4
Inner Language: I thought to myself using words or sentences to help me solve the problems	0	1	2	3	4

Appendix C

North Carolina State University
INFORMED CONSENT FORM for RESEARCH

Title of Study: Response Processes Validity of the Stanford-Binet Intelligence Scales, Fifth Edition

Principal Investigator: Sandye Ouzts Faculty Supervisor: Jeffery Braden, Ph.D.

You are invited to participate in a research study. The purpose of this study is to learn about the validity, or usefulness, of the Stanford-Binet Intelligence Scales, Fifth Edition. We hope to learn whether the tasks from the Stanford-Binet test really measure the cognitive skills that the test authors claim they measure.

INFORMATION
If you choose to participate in the study, you will first be asked to review and sign the informed consent form. You will then be asked to complete a demographics survey. Next, you will be asked to solve a set of problems from a standard psychoeducational test. The tasks represent thinking skills such as math, reasoning or puzzle-like problems. You will be asked to solve a total of 18 test items and to answer questions about how you solved the problems. First, we will ask you to explain how you solved the problems in your own words. Second, we will ask you to complete a questionnaire about how you solved the problems.

It will take approximately 1 hour for you to complete this study.

RISKS
Risks of participation are minimal, and likely to be limited to the loss of your time. Some participants may experience mild stress or discomfort when taking the test because some of the tasks are difficult. However, the purpose of the study is to see *how* you solve the problems, not *how well* you solve them. If you experience more than mild discomfort, you should tell the research assistant. You may stop and withdraw from participation at any time without penalty or loss of benefits. In addition, we can refer you to a counselor at the university if you would like further assistance.

BENEFITS
You will be given credit for participating in the study (see Compensation below). Additionally, you will learn more about how psychological research is conducted. Another benefit of participating is that you will help psychologists to better understand how individuals respond to this psychological test. Psychologists use tests to help their clients and others make important decisions about their client's lives, so we want to better understand what test takers are thinking when they take the test.

CONFIDENTIALITY
Your responses and personal identifying information will be kept confidential and will be made available only to researchers conducting the study. Your responses to questions will be stored securely on CD, and the CD will be retained for future research purposes after the completion of the study. Personal identifying information will be kept separate from response data and will be destroyed after the completion of the study. Only the code that I have assigned to you will be written on the documents with your responses to the questions. No reference will be made in oral or written reports

which could link you to the study. Data will be reported as group numbers (e.g., averages); any individual responses to questions will be presented anonymously so that the individual making the comment cannot be identified.

COMPENSATION
For participating in this study you will receive 2 research credits toward your requirements for PSY 200. Other ways to earn the same amount of credit are to participate in an alternative research project or to write a review paper for a research article. In the event that you withdraw from this study prior to its completion, you will receive one credit for participation. You have a right to withdraw at any point in the data collection process, and withdrawing will not affect your relationship with anyone at NC State University.

CONTACT
If you have questions at any time about the study or its procedures, you may contact Jeffery Braden (625 Poe Hall, jeff_braden@ncsu.edu, or 919-513-7393) or Sandye Ouzts (smouzts@ncsu.edu or 919-280-9099). If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may contact Dr. Matthew Zingraff, Chair of the NCSU IRB for the Use of Human Subjects in Research Committee, Box 7514, NCSU Campus (919-513-1834) or Mr. Matthew Ronning, Assistant Vice Chancellor, Research Administration, Box 7514, NCSU Campus (919-513-2148).

PARTICIPATION
Your participation in this study is voluntary; you may decline to participate without penalty. If you decide to participate, you may withdraw from the study at any time without penalty and without loss of benefits to which you are otherwise entitled. If you withdraw from the study before data collection is completed your data will be returned to you or destroyed at your request.

CONSENT
If have read and understand the above information. I have received a copy of this form. I agree to participate in this study.

______________________________________ **Participant ID __________**
Participant's Name- Please Print

______________________________________ ____________________
Participant's Signature **Date**

______________________________________ ____________________
Investigator's Signature **Date**

Appendix D

Demographics Form

Participant ID ______

First Name ___________________________
Last Name ________________________

Age _____

Gender (Circle one) Male Female

Race/Ethnicity
African American
American Indian/Alaskan Native
Asian American
Caucasian
Hispanic American
Native Hawaiian/Other Pacific Islander
Other (Please specify)

____

Year in College (Circle one)
Freshman
Sophomore
Junior
Senior
Graduate
Other (Please specify) _________________

Do you have a disability? If yes, please specify______________________

____

Appendix E

Method Used to Determine SB5 Items Administered

The SB5 includes both routing and nonrouting subtests in its design. The two routing subtests are administered prior to any other subtests to determine which level (Level 1 through 6) of nonrouting subtests should be administered to the examinee. This process is facilitated by the use of a routing table, which provides the examiner with the number of points the examinee must earn in order to route to a particular level on the nonrouting subtests.

For the nonrouting subtests, six levels are clearly identified by the developers of the SB5. To include items that were representative of a range of difficulty within each subtest, items from Levels 3, 4, and 5 were administered in the study. To ensure that items were representative of all levels of difficulty within each level, items of low, medium, and high difficulty were selected for each level. The following process was used to identify items of low, medium, and high difficulty: If the level included three items, then those three items were administered. If the level included six items, then the first, third, and sixth items were administered to represent low, medium, and high levels of difficulty, respectively. For the third level of the Verbal Fluid Reasoning, there was only one level of difficulty because this subtest only includes one activity.

However, the routing subtests (i.e., Object Series/Matrices and Vocabulary) are not divided into distinct levels. The levels within these subtests were identified by the following method: For routing subtests, the routing table for each routing subtest was used to determine which items would roughly correspond to items from levels three, four, and five in nonrouting subtests. Given the examinee’s score on the routing subtest, the routing table informs the examiner at which level to begin testing on the non-routing subtests. The ranges of items provided in the routing tables for each level were used to decide the item numbers that would roughly correspond to items in the nonrouting subtests. For example, for the Vocabulary subtest, if the examinee scores between 18 and 27 points, then the examiner administers items from Level 3 for the nonrouting subtests. Assuming the examinee receives maximum credit for all items, items 16 through 20 would correspond to Level 3 items in the nonrouting subtests. However, this method necessitates shifting up the range of items provided by the routing tables by a few items (A.D. Carson, personal communication, November 8, 2005). This is due to the fact that the highest items that are identified for each level will typically be less difficult than the items to which they would route (A.D. Carson, personal communication, November 8, 2005). To shift the range of items up, the first two items in each subtest were excluded, so that the range of items for each level was shifted up three items.

For routing subtests, the following process was used to identify items of low, medium, and high levels of difficulty: The first and last items included in the range for each level were selected. If only three items were included in the range, then those items were selected. The item representing medium level of difficulty was selected to ensure there was an equal number (if possible) of items between the three selected items. For example, for “Level 3” in the Vocabulary subtest, items 18, 20, and 22 were administered, so that there was one item between each item administered. For levels in which there were an even number of items (and thus, it is not possible to have an equal number of items between each level), the medium item was chosen so that there were fewer items between the low and medium difficulty items, than between the medium and high difficulty items.

Appendix F

Interview Protocol

When Participant Arrives

1. Go over what the participant will do

"If you choose to participate in the study, you will first be asked to review and sign the informed consent form. You will then be asked to complete a demographics survey. Next, you will be asked to solve a set of problems from a psychoeducational test with tasks such as math, reasoning, or puzzle-like problems. After the items, you will be asked to explain how you solved the problems in your own words and then to complete a questionnaire about how you solved the problems. All information and answers you give will be kept confidential."

2. Consent Form

"Here is the consent form for the study. Please look over this and let me know if you have any questions. When you are done, please print your name and sign and date it on the second page."

- After they sign it, give them a copy.

3. Demographics Survey: *"Now I'd like you to fill out the demographics survey."* Hand them survey.

4. Definition Form

"Before we begin, I want to point out that the purpose of the study is to see ***how*** *you solve the problems,* ***not how well*** *you solve them. It is important to do your best on the items, but it's more important to focus on how you go about solving the problems as you do them.* Give them Definitions. *Here is a list of processes that people sometimes use to solve problems. I will be asking you to rate how much you used each of these processes after we finish a set of items. Please read these and let me know if you have any questions."*

<u>If they ask a question (read verbatim)</u>

- Fluid Reasoning: *Fluid Reasoning involves inferring rules or looking at relationships in order to solve "on the spot" problems.*
- Quantitative Reasoning: *Quantitative Reasoning involves reasoning with numbers or numerical concepts.*
- Knowledge: *Knowledge involves using information you have learned through general life experiences and formal and informal educational experiences.*
- Working Memory: *Working Memory involves retaining information in memory temporarily, performing some operation or manipulation on it, and then producing a response.*
- Visual-Spatial Processing: *Visual-Spatial Processing involves visually perceiving or mentally imagining patterns or relationships, the whole picture, or positions in space. It involves using a visual image.*

- Inner Language: *Inner Language involves using one more words to think about or solve the problems.*

After You Give Each Subtest

1. Interview

- *"I would like you to explain in your own words how you solved* ***all*** *of the items you just answered. I'm going to review the items so you can think about how you solved all of them*." Review all items to jog participant's memory. *"How did you solve* ***all*** *of the items you just answered?*"
- Follow up questions
  Could you give me more detail? What were you thinking while you were solving the problems? How did that strategy help you to solve the problems?
- If you do not get five sentences even after the follow up questions, then ask the participant to choose a certain item and explain how they solved that item.

- **After you get 5 sentences, paraphrase participant's answer**.
  *Is there anything else you would like to add? Now I'm going to summarize what you just told me so I can make sure I have it correct and also to see if there is anything else you would like to add. You said …*Paraphrase. *Is there anything else you would like to add?*

2. Questionnaire

- Give them the Questionnaire marked "1" if you just gave a Nonverbal Subtest.
- Give them the Questionnaire marked "2" if you just gave a Verbal Subtest.

"Now, I'd like you to rate the degree to which you used each of these processes (point) *when you solved the test items. 0 is not at all* (point) *and 4 is Exclusively* (point). *Please rate each process based on* ***all*** *of the items you just answered."*

- If you forget to give the questionnaire: Finish all items for the subtest you are giving and do the interview and questionnaire. Go back and say, *"Now I'd like you to rate the degree to which you used each of these processes when you solved the problems I gave you earlier. I'm going to review the items so you can think about how you solved all of them.* Review. *0 is not at all and 4 is Exclusively. Please rate each process based on all of the items I just reviewed."*

End of Study

1. Debriefing (does not have to be read verbatim)

Thanks for your participation in this study! The items that you answered today came from the Stanford-Binet Intelligence Scales, which is an intelligence test used by psychologists to look at both child and adult intelligence. Psychologists use intelligence scores to make diagnoses so that children can be given appropriate educational placements and accommodations in the schools. Intelligence tests are also used in research, forensic, and employment settings. In this study, we are looking at the validity of the test and are interested in whether people report solving the items using the same cognitive skills the test developers intend for the items to measure. If you have any questions about the study, please contact the researchers listed in the Contact section on your consent form.

- In case they ask: People use intelligence tests to do research on cognitive abilities, for example, in the older adult population. In the forensic setting, federal law mandates that people with mental retardation cannot get the death penalty, and psychologists often do the testing. In the employment setting, intelligence tests are sometimes used for screening for employment.

Appendix G

IRB Approval Letter

North Carolina State University is a land-grant university and a constituent institution of The University of North Carolina

Office of Research and Graduate Studies

NC STATE UNIVERSITY

Sponsored Programs and
Regulatory Compliance
Campus Box 7514
1 Leazar Hall
Raleigh, NC 27695-7514

919.515.7200
919.515.7721 (fax)

From: Debra A. Paxton, Regulatory Compliance Administrator
North Carolina State University
Institutional Review Board

Date: February 24, 2006

Project Title: Response processes validity of the Stanford-Binet Intelligence Scales, Fifth Edition

IRB#: 071-06-2

Dear Ms. Ouzts:

The research proposal named above has received administrative review and has been approved as exempt from the policy as outlined in the Code of Federal Regulations (Exemption: 46.101.b.2). Provided that the only participation of the subjects is as described in the proposal narrative, this project is exempt from further review.

NOTE:
1. This committee complies with requirements found in Title 45 part 46 of The Code of Federal Regulations. For NCSU projects, the Assurance Number is: M1263; the IRB Number is: 01XM.

2. Review de novo of this proposal is necessary if any significant alterations/additions are made.

Please provide your faculty sponsor with a copy of this letter. Thank you.

Sincerely,

Debra Paxton
NCSU IRB

Appendix H

Cognitive Process Scoring Rubric

Fluid Reasoning

- Solving verbal or nonverbal problems using **inductive or deductive reasoning**
  - Deductive reasoning: examinee is given general information and is required to infer a conclusion, implication, or specific example
  - Inductive reasoning: reasoning from
    - -the part to the whole
    - -specific to general OR
    - -individual instance to the universal principle
- Includes determining the **underlying rules or relationships** among pieces of information the examinee has not seen before (i.e., **novel information**)
- Does **not require knowledge from school or previous experience**, but rather involves **understanding figural or verbal relationships** (figural relationships includes relationships between shapes, patterns, or series)
- **Examples**
  - Make a *relationship* between two things (words or parts of something) → understanding verbal or figural relationships
  - How two things are "*similar*" or finding "similarities" (i.e., understanding the relationship b/*t* two things)
  - Deductive reasoning where examinee
    - finds a general *pattern* and then infers a conclusion
    - applies a pattern to a specific example/ instance

Quantitative Reasoning

- Reasoning with numbers and numerical problem solving
  - Includes both word problems and problems using pictured relationships
  - Includes a variety of mathematical concepts (number concepts, estimation, problem solving, measurement, etc.)
  - May require applied problem solving and/or specific mathematical knowledge
- **Examples**
  - Numbers, counting
  - Operations such as addition, subtraction, multiplication, division
  - Mathematical problem-solving (may include formulas, equations)
  - Using math skills/ reasoning

<u>Knowledge</u>

- Applying or using the general information previously acquired at home, school, or work
  - Involves learned material, such as vocabulary, that has been acquired and stored in long-term memory

- **Examples**
  - Knowing things from everyday experience/ observation: "common knowledge" or "common sense"
  - "Learned" from school, home (parents), work (e.g., common laws)
  - Vocabulary (definitions) and parts of words (suffix, prefix, root words)

<u>Working Memory</u>

- Information is temporarily stored in short-term memory and then is inspected, sorted, or transformed (in memory)

- **Examples**
  - Uses words such as "remember," "memory," or "stored" information
  - Repeat words/sentences in your head
  - Remember things in a certain order/sequence (e.g., sorting into a certain order)
  - Focus on remembering certain parts of the information/ stimulus you are presented

---

<u>Visual Spatial Processing</u>

- Seeing patterns, relationships, spatial orientations (positions in space), or the gestalt whole among pieces of a visual display

- Includes visualization

- **Examples**
  - Visualize: form a visual image in your mind
    - Visualize positions in space
    - Visual patterns (e.g., shapes in a picture)

    - Use words like "look at," "visualize," "picture" in my head/mind, "envision," "imagine"
  - Look at whole picture, then analyze patterns/relationships among pieces or certain parts of the whole picture (i.e., how pieces go together to make a whole picture)

Appendix I

Verbal Mediation Scoring Rubric

Verbal

- Used inner language (thought to self in words, phrases, or sentences) to understand, think about, or solve the problems

- **Examples**
  - Use words/sentences
  - Think about words (e.g., vocabulary, listen to what someone says)
  - Mention "words," "sentences," or "statements"
  - Word problems
  - Repeat words/ sentences to self (rehearsing)
  - Use other languages
  - Report using "inner language" or self-talk

Nonverbal

- Used **little or no** inner language (words, phrases, or sentences) in response to the test questions

- **Examples**
  - May include visualizing, observation, analyzing patterns or pictures, looking at pictures or images, nonverbal problem solving (solving just by looking/watching)
  - Report using little or no "inner language"

Appendix J

Participants' Use of Labels in Interview Descriptions

Nonverbal Subtests

For the Nonverbal subtests, 44.5 % (45) of the 101 participants gave one or more cognitive process labels (e.g., "working memory"). Of the 45 participants who gave cognitive process labels, 68.9 % (31) labeled the matching process (i.e., the process intended to be elicited by the subtest), whereas 31.1% (14) participants gave only non-matching processes. Of the 31 participants who labeled the matching cognitive process, 41.93% (13) gave the matching label and no other labels, whereas the remaining 58.07 % (18) labeled one or more non-matching processes in addition to the matching process. In addition, a total of 5 participants used the label "inner language." Of these participants, 3 participants reported that they did not use "inner language" (i.e., provided a matching label), whereas 2 participants reported using "inner language" (i.e., provided a non-matching label).

Verbal Subtests

For the Verbal subtests, 50.49% (51) of the 101 participants gave one or more cognitive process labels (e.g., "working memory"). Of the 51 participants who gave cognitive process labels, 49.02% (25) gave the matching process (i.e., the process intended to be elicited by the subtest), whereas 50.98% (26) gave only non-matching processes. Of the 25 people who labeled the matching process, only 17.6% (9) gave the matching label and no other labels, whereas the remaining 82.4% (16) labeled one or more non-matching processes in addition to the matching process. In addition, a total of 18 participants used the label "inner language." Of these participants, 17 participants reported that they used "inner language” (i.e., provided a matching label), whereas 1 participant reported not using “inner language” (i.e., provided a non-matching label).

Both Subtests

Of the 101 participants, 35.65% (35) gave cognitive process labels for both the verbal and nonverbal subtest they completed. Of these 35 participants, 42.86% (15) gave the matching label for both subtests, whereas 57.14% (20) gave only non-matching labels. Of the 15 people who gave the matching label for both subtests, 20% (3) gave the matching labels and no other labels, whereas the remaining 80% (12) labeled one or more non-matching processes in addition to the matching process.

Appendix K

Cognitive Process Scoring Sheet
(Rater 1)

ID	Version 1	Version 2
1	FR QR KN WM VSP Other Primary __________	
2	FR QR KN WM VSP Other Primary __________	
3	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
4	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
5	**FR QR KN WM VSP Other** **Primary __________**	
6	FR QR KN WM VSP Other Primary __________	
7	FR QR KN WM VSP Other Primary __________	
8	FR QR KN WM VSP Other Primary __________	
9	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
10	**FR QR KN WM VSP Other** **Primary __________**	
11	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
12	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
13	FR QR KN WM VSP Other Primary __________	
14	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
15	**FR QR KN WM VSP Other** **Primary __________**	
16	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
17	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
18	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________

19	FR QR KN WM VSP Other Primary __________	
20	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
21	FR QR KN WM VSP Other Primary __________	
22	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
23	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
24	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
25	**FR QR KN WM VSP Other** **Primary __________**	
26	FR QR KN WM VSP Other Primary __________	
27	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
28	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
29	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
30	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
31	FR QR KN WM VSP Other Primary __________	
32	FR QR KN WM VSP Other Primary __________	
33	FR QR KN WM VSP Other Primary __________	
34	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
35	**FR QR KN WM VSP Other** **Primary __________**	
36	FR QR KN WM VSP Other Primary __________	
37	FR QR KN WM VSP Other Primary __________	
38	FR QR KN WM VSP Other Primary __________	
39	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
40	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**

41	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
42	FR QR KN WM VSP Other Primary __________	
43	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
44	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
45	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
46	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
47	FR QR KN WM VSP Other Primary __________	
48	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
49	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
50	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
51	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
52	FR QR KN WM VSP Other Primary __________	
53	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
54	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
55	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
56	FR QR KN WM VSP Other Primary __________	
57	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
58	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
59	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
60	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**

61	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
62	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
63	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
64	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
65	**FR QR KN WM VSP Other Primary __________**	**FR QR KN WM VSP Other Primary __________**
66	FR QR KN WM VSP Other Primary __________	
67	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
68	FR QR KN WM VSP Other Primary __________	
69	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
70	**FR QR KN WM VSP Other Primary __________**	**FR QR KN WM VSP Other Primary __________**
71	FR QR KN WM VSP Other Primary __________	
72	FR QR KN WM VSP Other Primary __________	
73	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
74	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
75	**FR QR KN WM VSP Other Primary __________**	**FR QR KN WM VSP Other Primary __________**
76	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
77	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
78	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
79	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
80	**FR QR KN WM VSP Other Primary __________**	**FR QR KN WM VSP Other Primary __________**

81	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
82	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
83	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
84	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
85	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
86	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
87	FR QR KN WM VSP Other Primary __________	
88	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
89	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
90	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
91	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
92	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
93	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
94	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
95	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
96	FR QR KN WM VSP Other Primary __________	
97	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
98	FR QR KN WM VSP Other Primary __________	
99	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
100	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
101	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________

136	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
137	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
138	FR QR KN WM VSP Other Primary __________	
139	FR QR KN WM VSP Other Primary __________	
140	**FR QR KN WM VSP Other** **Primary __________**	
141	FR QR KN WM VSP Other Primary __________	
142	FR QR KN WM VSP Other Primary __________	
143	FR QR KN WM VSP Other Primary __________	
144	FR QR KN WM VSP Other Primary __________	
145	**FR QR KN WM VSP Other** **Primary __________**	
146	FR QR KN WM VSP Other Primary __________	
147	FR QR KN WM VSP Other Primary __________	
148	FR QR KN WM VSP Other Primary __________	
149	FR QR KN WM VSP Other Primary __________	
150	**FR QR KN WM VSP Other** **Primary __________**	**FR QR KN WM VSP Other** **Primary __________**
151	FR QR KN WM VSP Other Primary __________	
152	FR QR KN WM VSP Other Primary __________	
153	FR QR KN WM VSP Other Primary __________	
154	FR QR KN WM VSP Other Primary __________	
155	**FR QR KN WM VSP Other** **Primary __________**	
156	FR QR KN WM VSP Other Primary __________	

157	FR QR KN WM VSP Other Primary __________	
158	FR QR KN WM VSP Other Primary __________	
159	FR QR KN WM VSP Other Primary __________	
160	**FR QR KN WM VSP Other** **Primary __________**	
161	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
162	FR QR KN WM VSP Other Primary __________	FR QR KN WM VSP Other Primary __________
163	FR QR KN WM VSP Other Primary __________	

Appendix L

Verbal Mediation Scoring Sheet
(Rater 1)

ID	
1	Verbal Nonverbal Other
2	Verbal Nonverbal Other
3	Verbal Nonverbal Other
4	Verbal Nonverbal Other
5	**Verbal Nonverbal Other**
6	Verbal Nonverbal Other
7	Verbal Nonverbal Other
8	Verbal Nonverbal Other
9	Verbal Nonverbal Other
10	**Verbal Nonverbal Other**
11	Verbal Nonverbal Other
12	Verbal Nonverbal Other
13	Verbal Nonverbal Other
14	Verbal Nonverbal Other
15	**Verbal Nonverbal Other**
16	Verbal Nonverbal Other
17	Verbal Nonverbal Other
18	Verbal Nonverbal Other
19	Verbal Nonverbal Other
20	**Verbal Nonverbal Other**
21	Verbal Nonverbal Other
22	Verbal Nonverbal Other
23	Verbal Nonverbal Other
24	Verbal Nonverbal Other
25	**Verbal Nonverbal Other**
26	Verbal Nonverbal Other
27	Verbal Nonverbal Other
28	Verbal Nonverbal Other
29	Verbal Nonverbal Other
30	**Verbal Nonverbal Other**
31	Verbal Nonverbal Other
32	Verbal Nonverbal Other

33	Verbal Nonverbal Other
34	Verbal Nonverbal Other
35	**Verbal Nonverbal Other**
36	Verbal Nonverbal Other
37	Verbal Nonverbal Other
38	Verbal Nonverbal Other
39	Verbal Nonverbal Other
40	Verbal Nonverbal Other
41	Verbal Nonverbal Other
42	Verbal Nonverbal Other
43	Verbal Nonverbal Other
44	Verbal Nonverbal Other
45	**Verbal Nonverbal Other**
46	Verbal Nonverbal Other
47	Verbal Nonverbal Other
48	Verbal Nonverbal Other
49	Verbal Nonverbal Other
50	Verbal Nonverbal Other
51	Verbal Nonverbal Other
52	Verbal Nonverbal Other
53	Verbal Nonverbal Other
54	Verbal Nonverbal Other
55	**Verbal Nonverbal Other**
56	Verbal Nonverbal Other
57	Verbal Nonverbal Other
58	Verbal Nonverbal Other
59	Verbal Nonverbal Other
60	Verbal Nonverbal Other
61	Verbal Nonverbal Other
62	Verbal Nonverbal Other
63	Verbal Nonverbal Other
64	Verbal Nonverbal Other
65	**Verbal Nonverbal Other**
66	Verbal Nonverbal Other
67	Verbal Nonverbal Other
68	Verbal Nonverbal Other
69	Verbal Nonverbal Other
70	**Verbal Nonverbal Other**

71	Verbal Nonverbal Other
72	Verbal Nonverbal Other
73	Verbal Nonverbal Other
74	Verbal Nonverbal Other
75	**Verbal Nonverbal Other**
76	Verbal Nonverbal Other
77	Verbal Nonverbal Other
78	Verbal Nonverbal Other
79	Verbal Nonverbal Other
80	**Verbal Nonverbal Other**
81	Verbal Nonverbal Other
82	Verbal Nonverbal Other
83	Verbal Nonverbal Other
84	Verbal Nonverbal Other
85	**Verbal Nonverbal Other**
86	Verbal Nonverbal Other
87	Verbal Nonverbal Other
88	Verbal Nonverbal Other
89	Verbal Nonverbal Other
90	**Verbal Nonverbal Other**
91	Verbal Nonverbal Other
92	Verbal Nonverbal Other
93	Verbal Nonverbal Other
94	Verbal Nonverbal Other
95	**Verbal Nonverbal Other**
96	Verbal Nonverbal Other
97	Verbal Nonverbal Other
98	Verbal Nonverbal Other
99	Verbal Nonverbal Other
100	**Verbal Nonverbal Other**
101	Verbal Nonverbal Other
136	Verbal Nonverbal Other
137	Verbal Nonverbal Other
138	Verbal Nonverbal Other
139	Verbal Nonverbal Other
140	Verbal Nonverbal Other
141	Verbal Nonverbal Other

142	Verbal Nonverbal Other
143	Verbal Nonverbal Other
144	Verbal Nonverbal Other
145	**Verbal Nonverbal Other**
146	Verbal Nonverbal Other
147	Verbal Nonverbal Other
148	Verbal Nonverbal Other
149	Verbal Nonverbal Other
150	Verbal Nonverbal Other
151	Verbal Nonverbal Other
152	Verbal Nonverbal Other
153	Verbal Nonverbal Other
154	Verbal Nonverbal Other
155	**Verbal Nonverbal Other**
156	Verbal Nonverbal Other
157	Verbal Nonverbal Other
158	Verbal Nonverbal Other
159	Verbal Nonverbal Other
160	Verbal Nonverbal Other
161	Verbal Nonverbal Other
162	Verbal Nonverbal Other
163	Verbal Nonverbal Other

Appendix M

Results of Original Statistical Analyses

The original statistical analyses for Prediction One included an omnibus repeated measures MANOVA, followed by a repeated measures MANOVA for each cognitive process condition separately. For the repeated measures analyses, the survey items (i.e., To what degree did you use X process?) were considered a (within-subjects) independent variable due to the nature of the statistical program used (SPSS). Therefore, both main and interaction effects were calculated. Next, one way MANOVAs for the verbal and nonverbal subtests separately, across all five cognitive conditions were conducted, followed by paired samples *t*-tests ($p < .05$) for nonverbal and verbal subtests from each of the five conditions separately.

First, a repeated measures MANOVA showed a significant interaction between the cognitive condition and survey item, $F(16,380) = 30.84$, $p < .0001$. These results indicate that scores on survey items vary according to the participant's condition. There was also a significant three-way interaction between condition, test (i.e., nonverbal or verbal), and survey item, $F(16, 281.70) = 7.50$, $p < .0001$. These results indicate that scores on the survey vary according to the type of cognitive condition, as well as the type of verbal mediation intended to be elicited by the subtest.

Second, a repeated measures MANOVA was calculated for each condition to examine the survey scores for each condition separately. For the Fluid Reasoning condition, there was a significant main effect for survey item: $F(4, 16) = 94.10$, $p < .0001$, and a significant interaction between subtest (i.e., nonverbal or verbal) and survey item: $F(4, 16) = 45.74$, $p < .0001$. These results indicate that the survey scores are different across each of the five processes rated, and the scores are also different for nonverbal and verbal subtests. Further, there was a strong effect size for survey item (partial $\eta^2 = 0.96$) and the interaction between subtest and survey item (partial $\eta^2 = 0.92$).

For the Knowledge condition, the main effect for survey item was not significant: $F(1, 20) = 2.81$, $p = 0.11$, but there was a significant interaction between subtest (i.e., nonverbal or verbal) and survey item: $F(4, 17) = 3.23$, $p < .05$. These results indicate that scores are not different across each of the five processes rated (for nonverbal and verbal subtests combined), but that the scores are different for nonverbal and verbal subtests. Further, there was a medium effect size for the interaction (partial $\eta^2 = 0.44$).

For the Quantitative Reasoning condition, neither the main effect: $F(1, 19) = 0.33$, $p = 0.57$, nor the interaction: $F(4, 16) = 1.52$, $p = 0.24$, were significant. These results indicate that, across nonverbal and verbal subtests combined, the survey scores are not significantly different for each of the five processes rated and also are not different for nonverbal and verbal subtests.

For the Visual Spatial Processing condition, the main effect for survey item was not significant: $F(1, 19) = 0.86$, $p = 0.37$, but there was a significant interaction between subtest (i.e., nonverbal or verbal) and survey item: $F(4, 16) = 11.17$, $p < .0001$. These results indicate that scores are not different across each of the five processes rated (for nonverbal and verbal subtests combined), but that the scores are different for nonverbal and verbal subtests. Further, there was a strong effect size for the interaction (partial $\eta^2 = 0.74$).

For the Working Memory condition, there was a significant main effect for survey item: $F(4, 20) = 94.10$, $p < .0001$, and a significant interaction between subtest (i.e., nonverbal or verbal) and survey item: $F(4, 20) = 45.74$, $p < .0001$. These results indicate that the survey scores are different across each of the five processes rated, and the scores are also different for nonverbal and verbal subtests. Further, there was a medium effect size for survey item (partial $\eta^2 = 0.29$) and a strong effect size for the interaction between subtest (i.e., nonverbal or verbal) and survey item (partial $\eta^2 = 0.69$).

Third, survey scores from the verbal and nonverbal subtests separately, across all five cognitive conditions, were examined using one-way MANOVAs. MANOVA results showed that across all nonverbal subtests there was a significant main effect for condition, $F(20, 306) = 15.36$, $p < .0001$, and across all verbal subtests there was a significant main effect for condition, *F*(20, 302 = 40.32, $p < .0001$. These results indicate that the survey scores for each of the five processes for the verbal and nonverbal subtests, separately, are different depending on the participant's condition (i.e., the cognitive process intended to be measured by the subtest). Further, there was a medium effect size for condition for nonverbal subtests (partial $\eta^2 = 0.44$) and a strong effect size for condition for verbal subtests (partial $\eta^2 = 0.66$).

Last, paired samples *t*-tests ($p < .05$) for nonverbal and verbal subtests from each of the five conditions separately were conducted. The results for the Nonverbal subtests are presented, followed by the results for the Verbal subtests.

Nonverbal Subtests

For the Fluid Reasoning condition, the mean rating for the Fluid Reasoning survey item ($M = 2.80$, $SD = 0.62$) was significantly higher than the Quantitative Reasoning ($M =$

0.95, *SD* = 0.95) and Knowledge (*M* = 0.85, *SD* = 0.93) mean ratings. However, the mean rating for Fluid Reasoning was not significantly higher than the Visual Spatial Processing (*M* = 3.05, *SD* = 1.15) or Working Memory (*M* = 2.15, *SD* = 1.31) ratings.

For the Knowledge condition, the mean rating for the intended cognitive process (Knowledge; *M* = 3.33, *SD* = 0.80) was significantly higher than the mean ratings for each of the other four unintended process as predicted.

For the Quantitative Reasoning condition, mean ratings for Quantitative Reasoning (*M* = 2.65, *SD* = 0.99) and Visual Spatial Processing (*M* = 2.45, *SD* = 0.89) were the highest ratings. The mean Quantitative Reasoning rating was significantly higher than the mean ratings for Fluid Reasoning (*M* = 1.60, *SD* = 1.57), Knowledge (*M* = 1.70, *SD* = 1.22), and Working Memory (*M* = 1.65, *SD* = 0.81).

For the Visual Spatial Processing condition, the mean rating for the intended cognitive process (Visual Spatial Processing; *M* = 3.50, *SD* = 0.61) was significantly higher than the mean ratings for each of the other four unintended process as predicted.

For the Working Memory condition, mean ratings for Working Memory (*M* = 3.40, *SD* = 0.68) and Visual Spatial Processing (*M* = 3.25, *SD* = 0.72) were the two highest ratings. The mean Working Memory rating was significantly higher than the mean ratings for Fluid Reasoning (*M* = 1.00, *SD* = 1.17), Knowledge (*M* = 0.90, *SD* = 1.29), and Quantitative Reasoning (*M* = 0.65, *SD* = 0.93).

Verbal Subtests

For the Fluid Reasoning condition, the mean rating for the Fluid Reasoning survey item (*M* = 2.20, *SD* = 1.36) was significantly higher than Quantitative Reasoning (*M* = 0.00, *SD* = 0.00) mean rating, but not higher than the Knowledge (*M* = 3.15, *SD* = 0.49), Visual Spatial Processing (*M* = 2.55, *SD* = 1.50), and Working Memory (*M* = 1.90, *SD* = 1.07) ratings.

For the Knowledge condition, the mean rating for the intended cognitive process (Knowledge; *M* = 2.76, *SD* = 1.04) was significantly higher than the mean ratings for each of the other four unintended process as predicted.

For the Quantitative Reasoning condition, the mean rating for the intended cognitive process (Quantitative Reasoning; *M* = 3.05, *SD* = 0.39) was significantly higher than the mean ratings for each of the other four unintended process as predicted.

For the Visual Spatial Processing condition, the mean rating for Visual Spatial Processing (*M* = 2.25, *SD* = 0.97) was significantly higher than the ratings for Fluid Reasoning (*M* = 1.30, *SD* = 0.87) and Quantitative Reasoning (*M* = 0.15, *SD* = 0.37), but not higher than the ratings for Knowledge (*M* = 2.85, *SD* = 0.81) or Working Memory (*M* = 1.90, *SD* = 0.85).

For the Working Memory condition, the mean rating for the intended cognitive process (Working Memory; *M* = 3.45, *SD* = 0.61) was significantly higher than the mean ratings for each of the other four unintended process as predicted.