Technical Report 1222

Development of a Test Battery to Assess Mental Flexibility Based on Sternberg's Theory of Successful Intelligence

Cynthia T. Matthew
Yale University

Jens F. Beckman
University of Sydney

Robert J. Sternberg
Tufts University

January 2008

United States Army Research Institute
for the Behavioral and Social Sciences

20080229014

U.S Army Research Institute for the Behavioral and Social Sciences

A Directorate of the Department of the Army Deputy Chief of Staff, G1

Authorized and approved for distribution:

Michelle Sams

MICHELLE SAMS, Ph.D.
Director

Research accomplished under contract for the Department of the Army

Yale University

Technical reviews by:

Theresa Taylor, U.S. Army Research Institute
Michelle Zbylut, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This TechnicalReport may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) January 2008	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) January 2003 -March 2006
4. TITLE AND SUBTITLE Development of a Test Battery to Assess Mental Flexibility Based on Sternberg's Theory of Successful Intelligence		5a. CONTRACT OR GRANT NUMBER DASW01-03-K-0001
		5b. PROGRAM ELEMENT NUMBER 611102
6. AUTHOR(S) Cynthia T. Matthew, Yale University Jens F. Beckman, University of Sydney Robert J. Sternberg, Tufts University		5c. PROJECT NUMBER B74F
		5d. TASK NUMBER 4901
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Yale University PACE Center 340 Edwards Street New Haven, CT -6511-3720		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Institute for the Behavioral & Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER Technical Report 1222

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Contractor Officer's Representative and Subject Matter POC: Dr. Paul A. Gade

14. ABSTRACT *(Maximum 200 words)*:
A test battery to assess mental flexibility was developed based on Sternberg's theory successful intelligence (1985). New mental flexibility assessment instruments were developed and underwent formative and summative evaluation. The newly developed mental flexibility tests showed adequate reliability, and preliminary evidence of construct- and criterion-related validity. One mental flexibility factor explained 70% of variance in the test battery and was differentiated from the latent factor underlying divergent and convergent measures of fluid intelligence. Preliminary evidence of incremental criterion-related validity was found, suggesting that the mental flexibility test battery explains variance above and beyond divergent and convergent measures of fluid intelligence in criterion measures.

15. SUBJECT TERMS
Intelligence, cognitive ability, mental flexibility, pattern recognition.

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	103	Ellen Kinzer Technical Publication Specialist 703-602-8047

Standard Form 298

Technical Report 1222

Development of a Test Battery to Assess Mental Flexibility Based on Sternberg's Theory of Successful Intelligence

Cynthia T. Matthew
Yale University

Jens F. Beckman
University of Sydney

Robert J. Sternberg
Tufts University

Basic Research Unit
Paul A. Gade, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

January 2008

Army Project Number
611102B74F

Personnel, Performance and Training

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution made by our able team of research assistants: Heidi Soxman, Laima Broderick, Mark Bluemling, Shrim Bathey, and Robyn Rissman. We thank Ewa Szymanska, an intern who inspired and contributed to FlexArt and Dr. Robert Sorrell for collaborating on this aspect of the project.

DEVELOPMENT OF A TEST BATTERY TO ASSESS MENTAL FLEXIBILITY BASED ON STERNBERG'S THEORY OF SUCCESSFUL INTELLIGENCE

EXECUTIVE SUMMARY

Research Requirement:

The primary purpose of this investigation is to develop and evaluate a test battery that assesses mental flexibility based on the theory of successful intelligence (Sternberg, 1985). Mental flexibility is defined as the ability to cope with novelty and establish automatized levels of information processing. This research simultaneously provides a means for assessing mental flexibility and validating the experiential subtheory of the theory of successful intelligence.

Procedure:

Five new mental flexibility assessment instruments were developed and underwent formative and summative evaluation. An initial item pool for each mental flexibility test was first developed and reviewed. Initial tests were piloted via a combination of paper-and-pencil and computerized on-site administration to a sample of college undergraduates. The pilot data were analyzed and revised accordingly. A revised and expanded mental flexibility battery and validation measures were administered to a larger sample of college undergraduates and analyzed.

Findings:

The newly developed mental flexibility tests showed adequate reliability, and preliminary evidence of construct- and criterion-related validity as measures of the ability to cope with novelty. One mental flexibility factor explained 70% of variance in the test battery and was differentiated from the latent factor underlying divergent and convergent measures of fluid intelligence. Preliminary evidence of incremental criterion-related validity was found, suggesting that the mental flexibility test battery explains variance above and beyond divergent and convergent measures of fluid intelligence in criterion measures. Newly developed mental flexibility tests showed a consistent and strong pattern of association with measures of pattern recognition, suggesting it may be an important predictor of mental flexibility.

Utilization and Dissemination of Findings:

Findings suggest that the newly developed test battery may measure flexible cognitive ability outside the framework used by conventional tests of fluid intelligence, supporting the validity of the experiential subtheory of successful intelligence. The mental flexibility test battery developed for purposes of this researc represents an initial stage in the development of a test battery that could potentially be used for selection and placement in educational and occupational settings. Given the importance of mental flexibility in a rapidly changing world, and the fact that it is not currently assessed within the framework of conventional psychometric tests, these tests seems to have practical utility as well as theoretical justification.

DEVELOPMENT OF A TEST BATTERY TO ASSESS MENTAL FLEXIBILITY BASED ON STERNBERG'S THEORY OF SUCCESSFUL INTELLIGENCE

CONTENTS

INTRODUCTION 1
  Purpose 2

BACKGROUND 2
  Traditional Approaches 2
  Successful Intelligence 4

FRAMEWORK FOR TEST DEVELOPMENT 6
  Test Format 8
  Predictions 8

TEST DEVELOPMENT 9
  Overview 9
  Flexible Inference 9
  Flexible Mapping 14
  Counterfactual Analogies 18
    Counterfactual Analogies–Figural (CFAF) 18
    Counterfactual Analogies–Verbal (CFAV) 19
  FlexArt Test 20

FIELD TESTING 23
  Investigation 1: Formative Evaluation 23
    Purpose 23
    Method 23
      Participants 23
      Reference Measures 23
        Berlin Model of Intelligence Structure (BIS) 23
        French Kit of Factor-Referenced Cognitive Tests (F-Kit) 24
        Cognitive Flexibility Scale 24
        NEO-Personality Inventory Revised 24
    Procedure 25
    Results 28
      Item Analyses: Test of Effects of Different Presentation Modes 28
        Flexible Inference (FI) 28
        Flexible Mapping (FM) 28
      Test of Effects of Different Presentation Modes 41
        Flexible Inference 41
        Flexible Mapping 41
      Factorial Structure of the Mental Flexibility Tests 42
        Flexible Inference 42
        Flexible Mapping 42

Preliminary Construct Validation 43
Summary 44

Investigation 2: Summative Evaluation 44
Purpose 44
Method 45
Participants 45
Measures—Mental Flexibility 45
Mental Flexibility 45
Flexible Inference 45
Flexible Mapping 46
Counterfactual Analogies (CFA) 46
CFA-Verbal 46
CFA-Figural 46
Insight Test 46
Measures—Cognitive Ability 47
Berlin Model of Intelligence (BIS) 47
French Kit of Factor-Referenced Cognitive Tests 47
Measures—Pattern Recognition 47
Soluble/Insoluble Analogy Test 47
Group Embedded Figures Test (GEFT) 48
Revised Minnesota Paper Form Board Test 48
Minnesota Clerical Test 48
Measures—Personality 48
Cognitive Flexibility Scale 48
NEO-Personality Inventory Revised 48
Measures—Criterion 49
College GPA 49
Creative Awards 49
Self-Report Flexible Thinking 49
Self-Report Flexible Behavior 49
Procedure 49
Results 51
Overview 51
Descriptive Statistics and Correlation Analyses 51
Individual Test Analyses 55
Flexible Inference 55
Flexible Mapping 57
Counterfactual Analogies 59
Figural (CFAF) 59
Internal Test Analyses 59
External Validation Analyses 63

Verbal (CFAV) 65
Internal Test Analyses 65
External Validation Analyses 70
Insight 72
Internal Test Analyses 72

External Validation Analyses 73
Test Battery Analysis 75
Convergent and Discriminant Validity 75
Incremental Validity 77
Pattern Recognition 78
Discussion 81
Unexpected Results 82
Pattern Recognition 82

GENERAL DISCUSSION 83

CONCLUSION 84

REFERENCES 85

APPENDIX A 92

APPENDIX B 93

LIST OF TABLES

TABLE 1 DESIGN OF DIFFERENT TESTING CONDITIONS FOR THE TEST FLEXIBLE INFERENCE 13

TABLE 2 ADMINISTRATION DESIGN FOR FLEXIBLE MAPPING IN NUMERICAL DOMAIN 17

TABLE 3 ADMINISTRATION DESIGNS 26

TABLE 4 FLEXIBLE INFERENCE, DOMAIN-TYPICAL INFERENCE ITEM SUBSET A (1 TO 5) 29

TABLE 5 FLEXIBLE INFERENCE, DOMAIN-ATYPICAL INFERENCE ITEM SUBSET A (1 TO 5) 30

TABLE 6 FLEXIBLE INFERENCE, DOMAIN-TYPICAL INFERENCE ITEM SUBSET B (6 TO 10) 31

TABLE 7 FLEXIBLE INFERENCE, DOMAIN-ATYPICAL INFERENCE ITEM SUBSET B (6 TO 10) 32

TABLE 8 FLEXIBLE INFERENCE, DOMAIN-TYPICAL INFERENCE ITEM SUBSET C (11 TO 15) 33

TABLE 9 FLEXIBLE INFERENCE, DOMAIN-ATYPICAL INFERENCE ITEM SUBSET C (11 TO 15) 34

TABLE 10 FLEXIBLE MAPPING, DOMAIN-HOMOGENEOUS

ANALOGIES, ITEM SUBSET A (1 TO 5) 35

TABLE 11 FLEXIBLE MAPPING, DOMAIN-HETEROGENEOUS ANALOGIES, ITEM SUBSET A (1 TO 5) 36

TABLE 12 FLEXIBLE MAPPING, DOMAIN-HOMOGENEOUS ANALOGIES, ITEM SUBSET B (6 TO 10) 37

TABLE 13 FLEXIBLE MAPPING, DOMAIN-HETEROGENEOUS ANALOGIES, ITEM SUBSET B (6 TO 10) 38

TABLE 14 FLEXIBLE MAPPING, DOMAIN-HOMOGENEOUS ANALOGIES, ITEM SUBSET C (11 TO 15) 39

TABLE 15 FLEXIBLE MAPPING, DOMAIN-HETEROGENEOUS ANALOGIES, ITEM SUBSET C (10 TO 15) 40

TABLE 16 FLEXIBLE INFERENCE, PCA WITH VARIMAX ROTATION, COMPONENTS MATRIX 42

TABLE 17 FLEXIBLE MAPPING, PCA WITH VARIMAX ROTATION, COMPONENTS MATRIX 43

TABLE 18 REFERENCE TESTS, PCA WITH VARIMAX ROTATION, COMPONENTS MATRIX 43

TABLE 19 SECOND-ORDER PCA WITH VARIMAX ROTATION, COMPONENT MATRIX 44

TABLE 20 SUB-INVESTIGATION TEST ADMINISTRATION SCHEDULE 50

TABLE 21 DESCRIPTIVE STATISTICS FOR ALL MEASURES 53

TABLE 22 INTERCORRELATIONS AMONG MEASURES 54

TABLE 23 FI: CORRELATIONS WITH COGNITIVE ABILITY MEASURES 55

TABLE 24 FI: CORRELATIONS WITH PERSONALITY MEASURES 56

TABLE 25 FI: CORRELATIONS WITH CRITERION MEASURES 56

TABLE 26 FM: CORRELATIONS WITH COGNITIVE ABILITY MEASURES 57

TABLE 27 FM: CORRELATIONS WITH PERSONALITY MEASURES 58

TABLE 28 FM: CORRELATIONS WITH CRITERION MEASURES 58

TABLE 29 CFAF: DIFFICULTY ESTIMATES AND DISCRIMINATION INDICES 60

TABLE 30 CFAF: SUMMARY OF DIFFICULTY, DISCRIMINATION, AND INTERNAL CONSISTENCY ESTIMATES 61

TABLE 31 CFAF: PRINCIPAL-COMPONENT FACTOR ANALYSIS WITH TWO FACTORS IMPOSED AND VARIMAX ROTATION PRESENTED BY ITEM TYPE 62

TABLE 32 CFAF: CORRELATIONS WITH COGNITIVE ABILITY CORRELATES 64

TABLE 33 CFAF: CORRELATIONS WITH PERSONALITY MEASURES 64

TABLE 34 CFAF: CORRELATIONS WITH CRITERION MEASURES 65

TABLE 35 CFAV: DIFFICULTY ESTIMATES AND DISCRIMINATION INDICES PRESENTED BY ITEM TYPE 66

TABLE 36 CFAV: SUMMARY OF DIFFICULTY, DISCRIMINATION, AND INTERNAL CONSISTENCY ESTIMATES 67

TABLE 37 CFAV: PRINCIPAL-COMPONENT FACTOR ANALYSIS, TWO-FACTORS, VARIMAX ROTATION 68

TABLE 38 CFAV: CORRELATIONS WITH COGNITIVE ABILITY MEASURES 71

TABLE 39 CFAV: CORRELATIONS WITH PERSONALITY MEASURES 71

TABLE 40 CFAV: CORRELATIONS WITH CRITERION MEASURES 72

TABLE 41 INSIGHT TEST: ESTIMATES OF ITEM DIFFICULTY AND DISCRIMINATION 72

TABLE 42 INSIGHT TEST: PRINCIPAL-COMPONENTS FACTOR ANALYSIS WITH VARIMAX ROTATION 73

TABLE 43 INSIGHT TEST: CORRELATIONS WITH COGNITIVE ABILITY MEASURES 74

TABLE 44 INSIGHT TEST: CORRELATIONS WITH PERSONALITY MEASURES 74

TABLE 45 INSIGHT TEST: CORRELATION WITH CRITERION MEASURES...75

TABLE 46 CORRELATIONS AMONG MENTAL FLEXIBILITY TESTS 76

TABLE 47 RESULTS OF PCA OF MENTAL FLEXIBILITY TEST BATTERY 76

TABLE 48 RESULTS OF PCA WITH VARIMAX ROTATION ON MENTAL FLEXIBILITY AND COGNITIVE ABILITY TESTS 77

TABLE 49 MENTAL FLEXIBILITY TEST VARIANCE EXPLAINED BY PATTERN RECOGNITION MEASURES 78

TABLE 50 CORRELATIONS AMONG MENTAL FLEXIBILITY AND PATTERN RECOGNITION MEASURES..80

LIST OF FIGURES

FIGURE 1 MENTAL FLEXIBILITY AND SUBTHEORIES OF SUCCESSFUL INTELLIGENCE..7

FIGURE 2 EXAMPLE OF A CLASSIFICATION TASK CREATED FOR THE FLEXIBLE INFERENCE TEST (PART 1 OF THE GIVEN ITEM TRIPLET) 10

FIGURE 3 EXAMPLE OF A CLASSIFICATION TASK CREATED FOR THE FLEXIBLE INFERENCE TEST (PART 2 OF THE GIVEN ITEM TRIPLET) 10

FIGURE 4 EXAMPLE OF A CLASSIFICATION TASK CREATED FOR THE FLEXIBLE INFERENCE TEST (PART 3 OF THE GIVEN ITEM TRIPLET) 11

FIGURE 5 EXAMPLE OF AN ANALOGY TASK CREATED FOR THE FLEXIBLE MAPPING TEST (A–C: PART 1 TO 3 OF THE ITEM TRIPLET). 15

FIGURE 6 EXAMPLE CFAF ITEM...19

FIGURE 7 CATEGORIES OF ITEMS IN CFAV...20

FIGURE 8 EXAMPLE OF AN “ADAPT” ITEM IN FLEXART..............................21

FIGURE 9 EXAMPLE OF A “SHAPE” IN FLEXART. 22

FIGURE 10 EXAMPLE OF A “SELECT” IN FLEXART. 22

INTRODUCTION

Given the rapid rate of technological advancement and the increase in social complexity brought about by globalization, military leaders must deal with more novelty and change than ever before. In this environment, problems and situations can fundamentally differ from those experienced in the past, and traditional approaches to problem solving based on lessons learned from experience and training do not readily apply. To be effective, military leaders must possess and develop the ability to think about problems and situations in new ways in real time. In particular, they must develop the capacity to shift paradigms or "think outside the box" and restructure problems so that useful and adaptive solutions can be found.

It is difficult to account for the ability to think flexibly solely in terms of conventional conceptions of intelligence. Although many different definitions of intelligence have been proposed over the years (see, e.g., "Intelligence and its measurement: A symposium," 1921; Sternberg & Detterman, 1986), the conventional notion is built around a loosely consensual definition of intelligence in terms of generalized adaptation to the environment. Some theories of intelligence extend this definition by suggesting that there is a general factor of intelligence, often labeled *g*, which underlies all adaptive behavior (Brand, 1996; Jensen, 1998; see essays in Sternberg & Grigorenko, 1997. In many theories, including the theories most widely accepted today (e.g., Carroll, 1993; Gustafsson, 1994; Horn, 1989, other mental abilities are hierarchically nested under this general factor at successively greater levels of specificity. For example, Carroll has suggested that three levels can nicely capture the hierarchy of abilities, whereas Cattell (1971) and Vernon (1971) suggested two levels were especially important. According to Cattell, nested under general ability are fluid abilities of the kind needed to solve abstract reasoning problems such as figural matrices or series completions and crystallized abilities of the kind needed to solve problems of vocabulary and general information. According to Vernon, the two levels corresponded to verbal–educational and practical–mechanical abilities. These theories, and others like them, are described in more detail elsewhere (Brody, 2000; Carroll, 1993; Embretson & McCollam, 2000; Herrnstein & Murray, 1994; Jensen, 1998).

Sternberg's theory of successful intelligence (1983, 1985, 1988) and its three subtheories on the componential level, the experiential level, and the contextual level represent a broader conceptualization of intelligence than is provided by traditional approaches. Assessment procedures that are based on conventional theories do not represent adequate indicators for the ability to deal with novelty, to adjust to changes, and to break out of routine ways of thinking when necessary. Therefore, it can be claimed that test scores on traditional intelligence test procedures, by neglecting the aspect of mental flexibility, are limited in their power to predict a person's capacity to deal effectively within an environment in a rapidly changing world. The theory of successful intelligence provides an expanded conceptual framework for assessing mental flexibility, which has the potential to predict to a greater extent the capacity to cope with novelty.

Purpose

The purpose of this project was to develop a new assessment instrument based on the theory of successful intelligence to measure mental flexibility. Broadly speaking, the concept of successful intelligence recognizes that social and culture factors in an environment and multiple personal capabilities ultimately determine success. This conception stands in contrast to traditional views of intelligence, which posit a single personal capability that determines success across performance situations. Successful intelligence is the balancing of analytical, creative, and practical abilities to achieve success in a particular setting. In other words, success within a particular socio-cultural context is determined by one's ability to capitalize on one's strengths and compensate for one's weaknesses to enact strategies for achieving success.

Mental flexibility, as we conceptualize it, is a part of creative thinking, but not the only part. Aspects of personality (e.g., openness to experience, need for cognition) and motivation (e.g., goal orientation) also contribute to creative thinking. As a sub-construct, mental flexibility manifests itself at every level of the theory of successful intelligence (componential, experiential, and contextual). Accordingly, we have created a multifaceted test of flexible thinking derived from the theory to measure how well one can apply the components of intelligence to relatively novel tasks and situations. Given the importance of flexibility in a rapidly changing world, and the fact that mental flexibility is not currently assessed within the framework of conventional psychometric tests (Sternberg, 1981), such a test seems to have practical utility as well as theoretical justification.

A more general goal of this project was to gain further insight into the construct validity of the theory of successful intelligence. The ability to deal with novelty is captured primarily within the experiential subtheory. However, the componential subtheory specifies information-processing components and the contextual subtheory specifies behavioral strategies (adapt, shape, and select) that play an important role in mentally flexible behavior. This research has been designed to further our understanding of mental processes that underlie effectively dealing with novelty within the framework of successful intelligence.

In this investigation, pattern recognition also is examined as an alternative framework for understanding the information-processing components that give rise to flexible thinking. Pattern recognition is defined here as a dynamic cognitive process of connecting cues to form meaningful configurations (patterns) in a given context (Margolis, 1987). Measures of pattern recognition have been shown to be associated with fluid thinking (Bal, 1988; Witkin, Oltman, Raskin, & Karp, 2002).

BACKGROUND

Traditional Approaches

Examining the link between coping with novel kinds of tasks or situations and intelligence is not new. Psychometric tests of intelligence often include items that measure a person's ability to cope with novelty. Spearman's (1923, 1927) factor analytic approach places coping with novelty squarely in the general factor of intelligence or *g*. Spearman postulated three qualitative cognitive principles—apprehension of experience, education of relations, and education of correlates—to account for *g*. Apprehension of experience involves the ability to recognize attributes of objects and ideas. Education of relations (abstract reasoning) involves the ability to infer relations between two or more objects and ideas, and education of correlates (analogic reasoning) involves the ability to link objects or ideas with a relation. The ability to cope with novelty places emphasis on the latter two principles, education of relations and education of correlates.

Guilford's (1956, 1967, 1982, 1985) structure of intellect (SOI) model classifies intellectual functioning in terms of operations, contents, and products. According to this framework, flexible thinking can be linked to two types of productive-thinking operations, through which new information is generated from known and remembered information. The first are divergent-thinking operations, which involve thinking in different directions, sometimes searching, sometimes seeking variety, as with trial-and-error thinking. The second are convergent-thinking operations, which involve integrating information to find one right answer.

Cattell and Horn (Cattell, 1963; Horn & Cattell, 1967) organized abilities according to a hierarchical structure and divided *g* into fluid intelligence (gf) and crystallized intelligence (gc). Fluid intelligence has been predominantly associated with reasoning, whereas crystallized intelligence has been predominantly associated with knowledge (Horn, 1988). Within this framework, coping with novelty is part of fluid intelligence.

The Berlin Intelligence Model (Jäger, 1982, 1984) represents an attempt to integrate several models of intelligence. It is a faceted model with a content facet for verbal, numerical, and figural abilities that is differentiated from an operation facet for processing speed, memory, creativity, and processing capacity, resulting in two facets that form 12 "structuples" (4 operations x 3 contents). Within this framework, coping with novelty would involve creativity operations, which are measured in three types of content (figural, numerical, and verbal), and processing capacity. Creativity operations have been shown to be moderately related to fluid intelligence and to have a stronger relationship with fluid as compared to crystallized intelligence. Processing capacity, which is very close to reasoning, has been shown to be strongly related to fluid intelligence (Beauducel & Kersting, 2002).

Intelligence tests have been subject to heavy criticism because of their lack of precision in predicting educational and occupational success (e.g., Sternberg, 1981; Sternberg & Williams, 1997). Despite the fact that intelligence tests include measures of flexible thinking, these tests fall short in predicting real-world manifestations of mental flexibility or creativity. One attempt to overcome the failure of intelligence tests to explain individual differences in cognitive task performance on such tests as the Embedded Figures Test (EFT) (Witkin et al., 1971) or the Rod and Frame Test (RFT) (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962) is the concept of Field dependence-independence (Witkin, 1975). Witkin et al. (2002) suggest that the field dependence/independence dimension of the Group Embedded Figures Test is the same as the adaptive flexibility dimension of Guilford and his associates (1952, 1955a, 1955b, 1957) and the flexibility of closure dimension of Thurstone (1944). Witkin claimed that it was identical with one of the three main factors of the Wechsler that is centered on Block Design, Object Assembly, and Picture Completion (Witkin, 1973). Field independence may be at least in part a "fluid ability" as defined by Cattell (1963) but it also may be an indicator of other aspects of intellectual functioning (Grigorenko & Sternberg, 1995). The role of field independence, specifically, and pattern recognition, broadly, in flexible thinking is still not well understood.

Sternberg's theory of successful intelligence provides a more developed theoretical framework as compared with traditional theories. When used as a basis for test development, the theory of successful intelligence has been shown to have incremental criterion-related validity (Sternberg, 1999).

Successful Intelligence

The theory of successful intelligence specifies the kinds of broad abilities (analytical, creative, and practical) that play a role in achieving success, the cognitive processes required to apply these abilities, and the problem-solving strategies to achieve success. The theory recognizes a dynamic aspect of successful performance—that success requires not simply applying acquired knowledge, but also coping with novelty and transforming novel experiences into automatic information processing.

Successful intelligence is conceptualized in the form of three subtheories: componential, experiential, and contextual. With regard to mental flexibility, the componential subtheory focuses on the flexible interaction of cognitive components as the elementary and universal units of information processing. The experiential subtheory focuses on the flexible application of information-processing components in novel situations. The contextual subtheory focuses on the flexible application of strategies for success in novel environments.

The theory of successful intelligence differs somewhat from conventional theories of intelligence in its conceptualization of what mental flexibility is and where mental flexibility belongs in a theory of intelligence. We have used all of the various elements of mental flexibility in different aspects of our own research investigating creative intelligence (e.g., Sternberg, 1981, 1982; Tetewsky & Sternberg, 1986). Within this framework, mental flexibility links creative intelligence to the experiential subtheory. Creative intelligence allows the individual to apply information-processing components to generate novel and interesting ideas or to build on novel concepts. Mental flexibility is the capacity to apply creative intelligence to novel experience. Creative intelligence is involved when the components of intelligence are applied to integrating seemingly disparate pieces of information in unusual ways. It typically is involved when components are applied to generating novel and interesting ideas or to build on novel concepts. According to the theory of successful intelligence, creative intelligence is particularly well measured by problems assessing how well an individual can cope with relative novelty to employ convergent or divergent thinking. In some of their componential work, Sternberg and his colleagues (Sternberg & Gardner, 1982, 1983; Sternberg & Gastel, 1989a, 1989b, 1989c, 1989d) have shown that when one goes beyond the range of novelty present in the items of conventional tests of intelligence, one starts to tap sources of individual differences measured little or not at all by such tests. Thus, when assessing intelligence, it is important to include in a battery of tests problems that are relatively novel in nature. These problems can be either convergent or divergent in nature. Convergent problems are of particular interest here because they represent the aspect of creativity that this proposal focuses on: flexibility in thinking.

In work with convergent problems, Sternberg and his colleagues presented 80 individuals with novel kinds of reasoning problems that had a single best answer. For example, they might be told that some objects are green and others blue; but still other objects might be grue, meaning green until the year 2000 and blue thereafter, or bleen, meaning blue until the year 2000 and green thereafter. Or they might be told of four kinds of people on the planet Kyron: blens, who are born young and die young; kwefs, who are born old and die old; balts, who are born young and die old; and prosses, who are born old and die young (Sternberg, 1982; Tetewsky & Sternberg, 1986). Their task was to predict future states from past states, given incomplete information. In another set of studies, 60 people were given more conventional kinds of inductive reasoning problems, such as analogies, series completions, and classifications, and told to solve them. But the problems had premises preceding them that were either conventional (dancers wear shoes) or novel (dancers eat shoes). The participants had to solve the problems as though the counterfactuals were true (Sternberg & Gastel, 1989a, 1989b).

In these studies, Sternberg and his colleagues found that correlations with conventional kinds of tests depended on how novel the conventional tests were. The more novel the items on the conventional tests, the higher the correlations with our tests. Thus, the components isolated for relatively novel items would tend to correlate more highly with more unusual tests of fluid abilities (e.g., that of Cattell & Cattell, 1963) than with tests of crystallized abilities. In other words, the more tests of both kinds measure flexible thinking, the more highly they correlate with each other. Sternberg and his colleagues also found that when response times on the relatively novel problems were componentially analyzed, some components better measured the creative aspect of intelligence than did others. For example, in the "grue–bleen" task mentioned above, the performance component requiring people to switch from conventional green–blue thinking to grue–bleen thinking and then back to green–blue thinking again was a particularly good measure of the ability to cope flexibly with novelty.

In work with divergent reasoning problems that have no one best answer, the investigators asked 63 people to create various kinds of products (Lubart & Sternberg, 1995; Sternberg & Lubart, 1991, 1995, 1996) where an infinite variety of responses was possible. Individuals were asked to create products in the realms of writing, art, advertising, and science. In writing, they were asked to write very short stories for which the investigators would give them a choice of titles, such as "Beyond the Edge" or "The Octopus's Sneakers." In art, the participants were asked to produce art compositions with titles such as "The Beginning of Time" or "Earth from an Insect's Point of View." In advertising, they were asked to produce advertisements for products such as a brand of bow tie or a brand of doorknob. In science, they were asked to solve problems such as one asking how people might detect extraterrestrial aliens among us who are seeking to escape detection. Participants created two types of products in each domain.

Sternberg and Lubart found, first, that creativity is composed of the elements proposed by their investment model of creativity: intelligence, knowledge, thinking styles, personality, and motivation. Second, they found that creativity is relatively, although not wholly, domain specific. Correlations of ratings of the creative quality of the products across domains were lower than correlations of ratings within domains and generally were at about the 0.4 level. Thus, there was some degree of relation across domains; at the same time there was plenty of room for someone to be strong in one or more domains but not in others. Third, Sternberg and Lubart found a range of correlations of measures of creative performance with conventional tests of abilities. As was the case for the correlations obtained with convergent problems, correlations were higher to the extent that problems on the conventional tests were non-entrenched. For example, correlations were higher with fluid than with crystallized ability tests, and correlations were higher the more novel the fluid test was. These results suggest that tests of creative intelligence have some overlap with conventional tests (e.g., in requiring verbal skills or the ability to analyze one's own ideas—Sternberg & Lubart, 1995) but also tap skills beyond those measured even by relatively novel kinds of items on conventional tests of intelligence.

FRAMEWORK FOR TEST DEVELOPMENT

According to the theoretical assumptions of the theory of successful intelligence, mental flexibility reflects the ability to deal with novelty and to establish automatized levels of information processing. To assess this ability, we used an assessment approach that belongs to the category of dynamic testing. In contrast to static measures that depend on prior knowledge or skill acquisition, dynamic testing involves procedures designed to assess the test-taker's ability to adapt or modify his or her performance in the testing session (Embretson & Prenovost, 2000). As has been shown in other areas (e.g., for the assessment of learning ability, see Guthke & Beckmann, 2000, 2003), this diagnostic approach represents a more appropriate way to assess intellectual abilities such as mental flexibility. In contrast to traditional approaches, the focus here is on the person's ability to deal with standardized variations of test conditions. There are two contemporary trends that characterize the dynamic testing approach. The first trend pertains to assessing responsiveness to intervening conditions in the testing session. The second, employed here, assesses response time or efficiency in cognitive processing.

According to the theory of successful intelligence, mental flexibility should be manifested at every level of the theory. In other words, mental flexibility as an ability construct needs to be indicated within every subtheory. On the *experiential level*, mental flexibility is defined as the ability to effectively cope or deal with novelty and to establish automatized levels of processing. Therefore, flexible use of information processing components (performance components, knowledge-acquisition components, and metacomponents), as they are defined within the *componential subtheory,* is necessary. On the *contextual level*, strategies (adapt, shape, select) must be flexibly applied to successfully manage one's environment. The ability to perceive novel aspects of a given environment, analyze observations from novel perspectives, generate novel and useful solutions to problems in situations, and use novel strategies in these environments requires mental flexibility. Figure 1 illustrates the relation of mental flexibility to subtheories specified by the theory of successful intelligence.

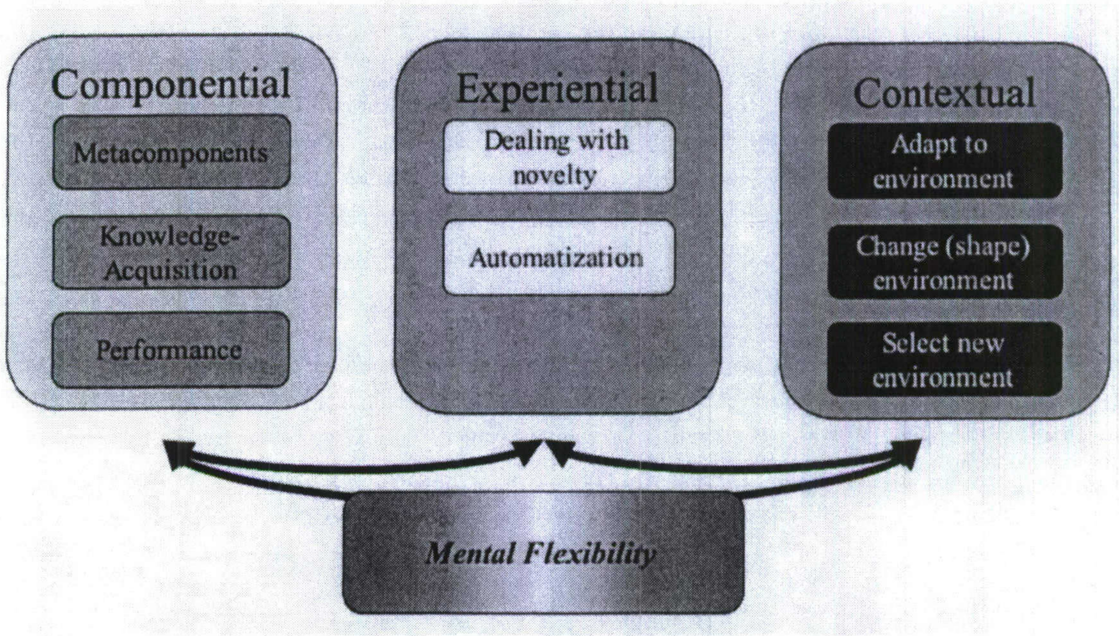

Figure 1: *Mental flexibility and subtheories of successful intelligence.*

Mental flexibility tests were designed to be consistent with each subtheory of successful intelligence. At the componential level, the capacity for flexible "inference" and "mapping" were expected to be of special relevance. Accordingly, tests were developed to measure flexible inference and flexible mapping of performance components using a dynamic testing approach. With dynamic testing procedures, experimentally controlled variations in test conditions and systematic assessment during the course of a test provide a more sensitive measure of intra-individual variability than do traditional test procedures. To assess flexible application of performance components, items quite similar to those used in traditional reasoning tests were developed using classification and analogy paradigms in multiple content domains (verbal, numerical, and figural).

Classification problems were designed to assess task performance in a context where frames of reference were manipulated, which requires test takers to shift their mindset, and/or inhibit the mental set evoked by a previous task. The classification task is to infer different relations between stimuli, balanced over three domains: verbal, numerical, and figural, within a constant set of stimuli. Although the set remains constant, arrangements of stimuli and rules governing their relationship vary. The focus of this test is the ability to flexibly infer relations between the stimuli in the given set.

Analogy problems were designed to assess task performance in a context and inferred rules must be applied to different domains. In traditional analogies tasks, the relation between the elements of the analogy stem has to be inferred, and the rule has to be mapped to other elements in the same domain. In our novel tasks, we broaden the mapping distance by introducing a domain switch within the same analogy. Our goal is to create an indicator of the ability to bridge different mapping distances. The focus here is on the ability to map flexibly.

At the experiential level of assessment, the capacity to work with changing assumptions was expected to be relevant to mental flexibility. Two types of tests were developed for this purpose, analogy and insight. In regard to analogies, two tests of counterfactual analogies were developed in verbal and figural domains, in which premises were manipulated to measure the test takers' ability to shift between familiar and novel premises to solve items on the test. With regard to insight, a test of mind puzzles in verbal, numerical, and figural domains was developed to test novel reasoning, or the capacity to restructure the elements of a problem to find a fitting solution.

Reproductions of visual images from paintings, drawings, and photographs were used to design a classification test within the framework of the contextual subtheory. It was expected that novel natural images provide stimuli that are contextually rich and ecologically valid. The test measures the ability to recognize the relationship among art images and appropriately use different strategies (adapt, shape, and select) based on the perceived nature of the relationship.

Test Format

Our theory-based mental flexibility test battery has been designed to be multifaceted, to provide multiple measures of performance, and to be practical to administer. Pilot data exist for all item types. Items cover verbal, numerical, and figural domains to ensure that measures are not confounded by any one domain. Tests in the battery were designed to be relatively resistant to the differential effects of previous experience. The vast majority of tests are multiple-choice, with the exception of the Insight test, which has an open-response format.

The test battery can be administered via computer for ease and standardization of conditions and data entry and processing. Each test is scored for average response time and response accuracy. Paper-and-pencil administration also is possible but with the obvious limitation to accuracy scoring. Administration of the full test battery varies but takes on average 1.5 to 2 hours. The test battery is suitable for adults in the normal to superior range of abilities.

Predictions

Our general expectation is that the intra-individual variability in performance scores on the full test battery will be indicative of people's ability to use their cognitive resources flexibly. Our research design varies item sequence and presentation mode. We expect that total scores on test performance will reflect the ability to cope with novelty introduced by test procedure as well as test content.

Conventional divergent and convergent tests of fluid ability were selected to establish convergent and discriminant validity. Scores on a test of mental flexibility are expected to be related to these tests but not strongly. The test battery is expected to provide incremental prediction of success criteria beyond these conventional tests.

Selected tests of pattern recognition were examined as predictors of mental flexibility test performance. This aspect of the research is exploratory and is expected to shed light on its relation to mental flexibility.

TEST DEVELOPMENT

Overview

Six tests were developed to assess aspects of mental flexibility according to three subtheories (componential, experiential, and contextual) of the theory of successful intelligence. Flexible Inference and Flexible Mapping tests were developed consistent with the componential subtheory such that components or elements of an object or concept are altered and the test taker must modify inferences and analogies accordingly. Counterfactual Analogies–Figural, Counterfactual Analogies–Verbal, and Insight tests assess the capacity to respond to novelty at a more complex level of cognition consistent with the experiential subtheory such that problems are presented to solve that contain unfamiliar or counterintuitive assumptions. FlexArt was designed to employ a more complex stimulus that more closely simulates the complexity of everyday life experience and, consequently, requires the practical processing consistent with the contextual subtheory. It was designed with more natural concepts in the form of a multifaceted stimulus that favors analysis from versatile perspectives.

The tests are described below. All tests underwent initial pre-pilot testing and review. Final test versions, along with answer keys and scoring rubrics, can be found in a test manual supplement to this report.

Flexible Inference

We developed a test procedure where classification items containing the same set of stimuli but in different arrangements are presented so that different relations have to be inferred and different rules have to be applied to find the correct answer to seemingly identical problems. Traditional-type analogy test items were used as a springboard for item generation. Two researchers adapted and modified items to represent different relations as illustrated below. Items underwent an iterative process of generation, analysis, piloting, and review by four members of the research team to ensure they conformed to the conceptual structure of the test.

Instructions were as follows:

> Select the pair of answer choices that constitutes the best match to the target on the left side, based on their common properties.

Illustrative examples and practice questions were provided at the beginning of the test.

Figure 2 gives an example for an item using shapes for a classification task. Here, the participant must select the pair of answer choices that constitutes the best match to the target on the left side, based on their common properties.

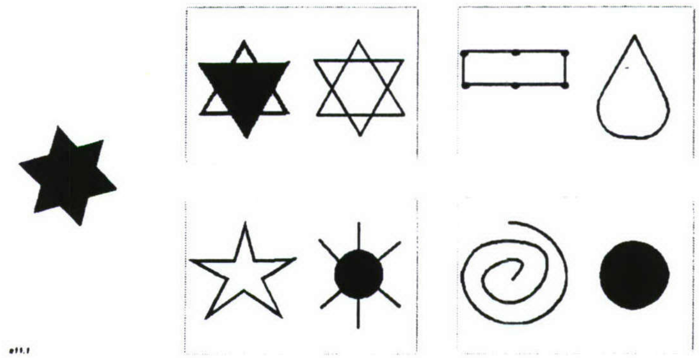

Figure 2. *Example of a classification task created for the Flexible Inference test (part 1 of the given item triplet).*

The correct answer to this item would be the upper left pair, referring to the overall shape the elements of this pair have in common with the target on the left. This paradigm, which is typically used to assess fluid intelligence, was adapted to test mental flexibility. The participant is next presented with the same target (on the left side of the screen) and even with the same set of stimuli on the right side, but which are now rearranged (see Figure 3). Because the rule of inference used for the previous problem (star-shape) is no longer valid, to find the correct solution the participant must attend to other characteristics of the target stimulus. Now, consideration of the number of attributes (dots, spikes, or rays) will lead to the correct answer in this item (lower left pair).

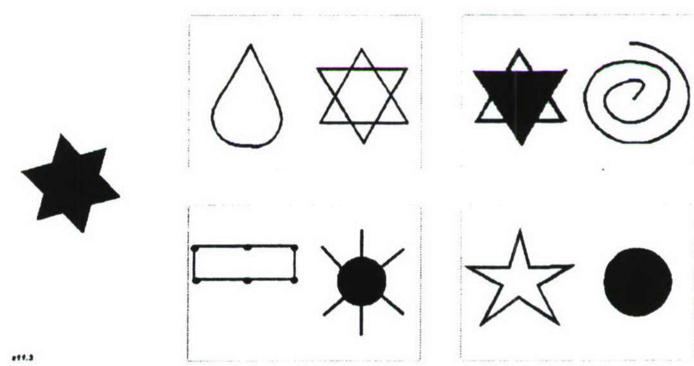

Figure 3. *Example of a classification task created for the Flexible Inference test (part 2 of the given item triplet).*

In the third part of the task (all items are presented in item triplets), the shapes are then presented in yet another arrangement. The previously inferred rules must be inhibited and the participant once again must infer the relationship that links the target together with one of the pairs of answer choices (see Figure 3).

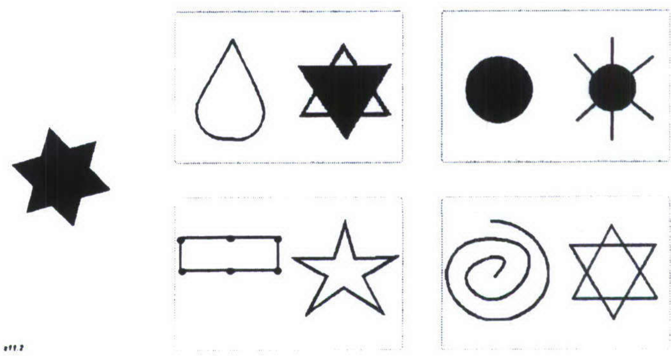

Figure 4. *Example of a classification task created for the Flexible Inference test (part 3 of the given item triplet).*

In this example (Figure 4), the correct answer would be the upper right pair, because of the solidness the target has in common with the elements of this pair.

All items in the Flexible Inference test are thus arranged in item triplets. The stimuli fall into one of three categories, figural (as in this example), numbers, or words, to balance out potential domain-related variance in dealing with classification problems. To be successful on these items, flexible use of different frames of reference for familiar stimuli is necessary. The ability to inhibit experience gained on previous items is the prerequisite for using different cognitive approaches to the same set of stimuli. It is expected that the intra-individual variability in performance scores within each item triplet will be indicative of the person's ability to use his or her cognitive resources flexibly.

Our approach is based on the assumption that we can use the performance differences between two different item classes we have combined in the test. In the Flexible Inference test, items that ask for the inference of domain-typical classification rules (e.g., focusing on numerical characteristics in numbers) represent one class. Items that require the inference of classification rules based on domain-atypical characteristics of the stimuli (e.g., number of vowels in words) represent the other item class. Every item triplet consists of items from each class.

In the case of the classification tasks (Flexible Inference), we can assume that it is harder to find domain-atypical than domain-typical classification rules. We also expect that it will be more difficult to identify rules when an item (as a part of an item triplet) is preceded by another item using the same target and set of stimuli in which domain-typical characteristics were relevant for its solution. In other words, the unfamiliarity or novelty effect (domain-atypical characteristics) will be complemented by a transition effect caused by the inhibition costs for previous perspectives on the same set of stimuli.

In terms of a componential analysis of the task requirements to solve classification problems in the Flexible Inference tests, one needs to:

- *encode* the terms (number, figures, words).
- *infer* the relation between the two terms in each pair. (What do they have in common? Does this make them a unique pair in comparison to the others?)

- *map* this relation onto the set of characteristics of the target encoded. (Is this uniqueness—higher-order differences—of the given pair relevant to the target?)
- *decide* on this basis to which of the pairs the target belongs.
- *apply* what has been learned.

Because the second appearance of the target (and the same set of stimuli) does not necessarily require a renewed encoding process, the problem-solving process should start at the point where the relations between the newly paired elements begin. Here participants must inhibit former experience (previous inferred relations) and switch their focus of attention to different characteristics. In other words, the problem solver needs to change his or her frame of reference. A person's susceptibility to interference would lead to difficulties in seeing the target and/or the elements of the pairs from a different perspective, which can be seen as an indicator of a lack of mental flexibility

For Flexible Inference, we created an initial item pool of 135 items. Because each item cannot be administered to every participant, the whole item pool was divided into three subsets; thus each participant tackles 45 items, which are organized into 15 item triplets. For each domain there are five item triplets to work on. It can be assumed that it is harder to find domain-atypical classification rules. Looking for domain-atypical characteristics might not be in accordance with the mindset triggered by the experience on the previous problem. It is expected that the identification of domain-atypical rules will be even harder if the item (as a part of an item triplet) is preceded by an item using the same target and set of stimuli in which domain-typical characteristics were relevant for solution. In other words, the unfamiliarity or novelty effect associated with the domain-atypical item triplet part might be complemented by a transition effect caused by the inhibition costs for previous perspectives on the same set of stimuli employed to infer domain-typical rules, as is usually expected. To test this assumption, we contrasted the effects of different presentation orders of the parts of each item triplet (e.g., atypical–typical–atypical vs. typical–atypical–atypical, etc.).

Because we are interested in finding indicators of the ability to switch frame of reference, the primary goal of this experimental variation of test conditions was to find the particular item-part order that causes the most transition costs. This would allow us to create test conditions that induce the maximal inter-individual variability in coping with the requirement to switch the frame of reference. To test these assumptions regarding potential effects of different intra-item triplet orders, the participants were assigned to one of three item-pool subsets and to one of six intra-item triplet order groups as well, for a total of 18 different experimental groups.

Table 1 gives an overview of the different conditions under which the Flexible Inference test was administered. The markings represent the set of items a given participant deals with if assigned to the condition of intra-item triplet order "typical–atypical–atypical" and item-subset B, which means item triplets with the numbers 5 to 10.

Table 1
Design of Different Testing Conditions for the Test Flexible Inference

Domain	Intra-item Triplet Order	Item Pool Subset
Numerical	domain typical domain atypical1 domain atypical2	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	domain typical domain atypical2 domain atypical1	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	domain atypical1 domain typical domain atypical2	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	domain atypical1 domain atypical2 domain typical	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	domain atypical2 domain typical domain atypical1	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	domain atypical2 domain atypical1 domain typical	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
Verbal	domain typical domain atypical1 domain atypical2	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	…	… … …
	domain atypical2 domain atypical1 domain typical	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
Shape	domain typical domain atypical1 domain atypical2	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)
	…	… … …
	domain atypical2 domain atypical1 domain typical	A (item triplet 1 to 5) B (item triplet 5 to 10) C (item triplet 11 to 15)

When taking the test, participants are presented first with an analogy to observe at their own pace. When the participant chooses, he or she selects the next screen, on which the analogy appears with a prompt and response options. Test item scores were number of correct responses per triplet. Latencies were determined by reaction time to respond to the prompt on the second screen.

Flexible Mapping

An approach similar to the Flexible Inference test was used for another mental flexibility test that focuses on "mapping" as a performance component within the componential subtheory. With the use of analogy problems, we sought to gain information about a person's ability to apply a previously inferred rule across different situations. We utilized the same procedure for item generation. Test instructions were as follows:

> In the following section you will be presented with two shapes (words or numbers). Based on the relationship between the two words, you will be asked to find the best match for the verbal, shape or numerical item from the answer choices.

Sample and practice items were provided at the beginning of the test. Scoring procedures were consistent with the Flexible Inference test.

As in the Flexible Inference test, all items were organized in item triplets. The first part of a given item triplet represents a traditional analogy: A relation between the elements of the analogy stem must be inferred, and a rule based on this relation must be applied to complete the analogy. In traditional analogies, the rule must be mapped to other elements from the same domain (see Figure 5a). In our novel tasks, however, we tried to broaden the mapping distance by introducing domain switches within each of the analogy item triplets. For instance, the relation between two numbers (e.g., 88 and 22) must be inferred (the latter is a fourth of the former) and mapped onto another domain so that the same relation between two words (see Figure 5b), or two shapes (see Figure 5c) will complete the analogy correctly.

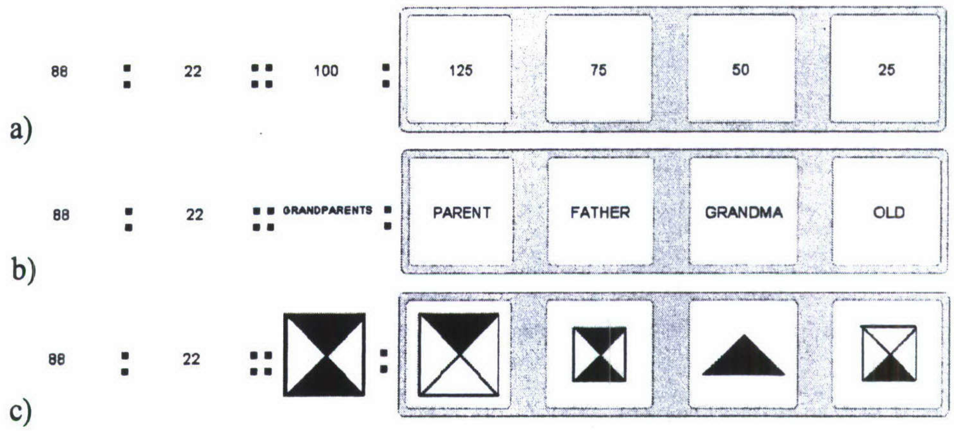

Figure 5. *Example of an analogy task created for the Flexible Mapping test (a–c: part 1 to 3 of the item triplet).*

If the rule is inferred correctly, the participant will recognize that the second term in the analogy stem is one-fourth of the first term. Mapping the rule onto the verbal domain will lead the participant to choose "GRANDMA" because a grandmother is one of four grandparents (see Figure 5b). Mapping onto the shape domain (as required in the third part of the item triplet) should result in choosing the third answer option. Here the single solid triangle represents one-fourth of the shape given (see Figure 5c).

The purpose of this procedure is to obtain an indicator of a person's ability to bridge different mapping distances. Whereas the new classification tasks (Flexible Inference) focus on the ability to infer different relations flexibly, the focus in the analogy test (Flexible Mapping) is on the ability to map rules flexibly.

Similar to Flexible Inference (see classification problems), the item pool for Flexible Mapping also consists of two different classes of items. In Flexible Mapping we have domain-homogeneous items, where no domain switch is required within the given analogy, and domain-heterogeneous items, where the domain of the analogy stem is different from that of the application field. Both classes of items are represented in each item triplet. Because of the wider mapping distance to be bridged in mapping items, these items are expected to be harder (mapping costs). A person's variability in performance (within each item triplet) will be an indicator for the disturbance the domain shift causes individually. Our general expectation is that the specific procedure we have deployed in both tests, Flexible Mapping and Flexible Inference, will cause inter-individual differences in levels of intra-individual variability in performance on items from different classes.

In terms of a componential analysis of the task requirements to solve analogy problems (of the type: A : B :: C : ?) in the Flexible Mapping tests, one needs to:

- *encode* the characteristics of the terms given in the analogy stem (A and B);
- *infer* the relation between the two first terms (A : B);
- *map* the inferred relation between the first two to the third term (C);

- *recognize* a "meta-relation" that relates the first two terms to the third term;
- *apply* the rule to the third term to produce the missing fourth one (D); and finally, under the condition of multiple choice, one needs to
- *justify* the decision about which answer option completes the analogy according to the rule applied.

In the Flexible Mapping test, the analogy stem (A and B) remains the same within each item triplet. Neither the encoding nor the inference process is required in the second or third appearance of the analogy stem (terms A and B). A flexible use of resources should prevent unnecessary problem-solving steps. Since the third element of the analogy in the mapping part of the item triplets comes from another domain than the analogy stem, the mapping process is expected to be more difficult. A more divergent and flexible reasoning process is required to solve domain-heterogeneous as compared with domain-homogeneous (traditional) analogy problems because of the need to map across domains. A representation of the relation to be mapped on an abstract level facilitates the mapping process (e.g., a rule representation such as "22 is four times less than 88" is harder to map across domains than a rule represented as "22 is a fourth of 88"). Difficulties in mapping an already inferred rule onto another situation (introduced by a required domain switch) give evidence for a lack of mental flexibility.

For Flexible Mapping, we also created an initial item pool of 135 items. The same segmentation procedure used in Flexible Inference was applied to this item pool, resulting in three item-pool subsets (A, B, C). To create a test procedure that potentially causes the most inter-individual variability in dealing with domain switches; two different presentation modes were tested. In items presented in "sequential mode," each item triplet part appears on a separate screen. In items presented in "group mode," the preceding part(s) of the given item triplet remains on the screen after it has been answered (without indicating the answer chosen). The underlying assumption here was that the visual availability of previous item triplet parts (same analogy stem in different domain-related contexts) would either facilitate a more abstract representation of the rule inferred, which would be beneficial to complete the second and third part of the item triplet successfully, or would increase domain-switching costs, in case of a high susceptibility toward mindsets induced by the first and domain-homogeneous part of the given item-triplet.

To test these assumptions and to determine the test procedure in the final version of the test, the following design was employed. One group of participants started with three item triplets presented in group mode followed by another three item triplets presented in sequential mode. This rule applied for each domain. The other group started with three item triplets presented in sequential mode followed by another three item triplets in group mode. Because there were only five item triplets for each domain in each item-pool subset, a randomly selected item triplet from one of the other item pool subsets was added to this item-triplet block. Thus, three item triplets were presented in one presentation mode and three item triplets in the other mode. That means, for instance, a participant assigned to item pool subset B starts with item triplets 6 to 8 presented in group mode (if also assigned to the group mode condition). Then item triplets 9 and 10 are presented in sequential mode. To ensure that the number of item triplets presented in sequential mode equaled the number of item triplets presented in group mode, in each domain, an additional item triplet (from item pool subset C, that is, item triplets 11 to 15) is presented in sequential mode to this participant as well (see markings in Table 3).

In the Flexible Mapping test, each participant must evaluate 8 item triplets, six for each domain. Because an additional systematic variation of the intra-item triplet order (homogeneous → heterogeneous x → heterogeneous y vs. homogeneous → heterogeneous y → heterogeneous x) was introduced to check for order effects, the total design for Flexible Mapping ended up with a total of 12 different experimental groups to which the participants were randomly assigned. Table 2 illustrates the administration design of the Flexible Mapping test applied to items in the numerical domain. Analogous designs were applied to items in verbal and figural ("shape") domains.

Table 2
Administration Design for Flexible Mapping in Numerical Domain

Domain	Presentation Mode	Intra-Item Triplet Order	Item-Pool Subset
Numerical	3 item triplets in group mode → 3 item triplets in sequential mode	non-mapping mapping onto verbal mapping onto shape	A (item triplet 1 to 5) + one item triplet out of subset B
			B (item triplet 6 to 10) + one item triplet out of subset C
			C (item triplet 11 to 15) + one item triplet out of subset A
		non-mapping mapping onto shape mapping onto verbal	A (item triplet 1 to 5) + one item triplet out of subset B
			B (item triplet 6 to 10) + one item triplet out of subset C
			C (item triplet 11 to 15) + one item triplet out of subset A
	3 item triplets in sequential mode → 3 item triplets in group mode	non-mapping mapping onto verbal mapping onto shape	A (item triplet 1 to 5) + one item triplet out of subset B
			B (item triplet 6 to 10) + one item triplet out of subset C
			C (item triplet 11 to 15) + one item triplet out of subset A
		non-mapping mapping onto shape mapping onto verbal	A (item triplet 1 to 5) + one item triplet out of subset B
			B (item triplet 6 to 10) + one item triplet out of subset C
			C (item triplet 11 to 15) + one item triplet out of subset A

Counterfactual Analogies

Applying a scheme developed by Sternberg and Gastel (1989a, 1989b), analogy items were developed in which an item stem is preceded by a premise that is either familiar or counterfactual (novel), and either relevant or irrelevant. Items are equally divided among familiar–relevant, familiar–irrelevant, counterfactual–relevant, and counterfactual–irrelevant premise types. Two tests were developed, Counterfactual Analogies Figural and Counterfactual Analogies Verbal. Instructions for the test were as follows:

> For each question below, there are three shapes (words). The first pair of shapes (words) goes together in a certain way. Your task is to choose the shape (word) that goes with the third given shape (word), thus creating a second pair of shapes (words) in the same way that the first pair goes together.
>
> Each question has a "Pretend" statement. You must suppose that this statement is true. Think of the statement, and then decide which shape (word) goes with the third shape (word) in the same way that the first pair of shapes (words) goes together.
>
> Illustrative examples and practice questions were provided for each test version.

The test presentation and scoring was similar to FI/FM. When taking the test, participants are presented first with an analogy to observe at their own pace. When the participant chooses, he or she selects the next screen, on which the analogy appears with a prompt and response options. Sample questions and answers and practice questions are provided at the beginning of the test. Test scores were number of correct responses. Latencies were determined by reaction time to respond to the prompt on the second screen.

Counterfactual Analogies–Figural (CFAF)

For this test, 15 items were developed. Each item is preceded by a premise stating novel, counterfactual statements.

Figure 6 shows an example item in which the number dimension needs to be ignored to find the correct completion of the analogy, which is represented by answer option C.

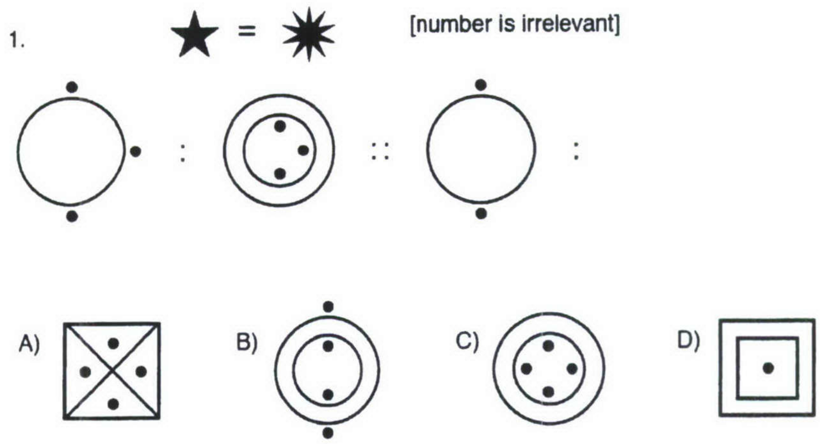

Figure 6. *Example CFAF item.*

The successful integration of novel and counterfactual information into routine ways of reasoning leads in this paradigm to a reduction of the complexity of the analogy, because the counterfactual premise always requires ignoring an otherwise solution-relevant dimension. The successful use of redundancy is expected to distinguish mentally flexible participants from less flexible ones. In this respect, mental flexibility is conceptualized as not necessarily dependent on quantitatively higher levels of mental capacity. Mental flexibility also is expected to manifest itself as the ability to use redundancy (pattern recognition) and to invest limited resources wisely.

Counterfactual Analogies–Verbal (CFAV)

Counterfactual Analogies–Verbal is another test that requires changing assumptions to correctly solve analogy problems. Participants are presented with verbal analogies, each preceded by premises. In some of the items, the premises are counterfactual (e.g., money falls off trees). Participants must solve these analogies as though the counterfactual premises were true (Marr & Sternberg, 1986; Sternberg et al., 1999, 2001). Other premises state familiar things (e.g., milk is liquid). In addition the relevancy of the premise to finding a solution is varied such that in some cases the premise is required to find the correct solution and in other cases it is not. The difference in performance on these two categories of items is expected to be indicative of a person's ability to integrate novel and unexpected information into the problem-solving process, which is considered to be essential for mental flexibility. Scoring, however, aggregates performance on both types of items taken together.

The item pool has been divided into 9 overlapping sets of 32 items. Each subset contains 8 items with familiar but irrelevant premises, 8 items premised with familiar statements relevant to determining the correct completion of the analogy, 8 items that have novel but irrelevant premises, and 8 items in which the premises state novel "facts" that need to be considered as true to correctly complete the analogy (novel/relevant). Figure 7 gives examples for each of the four categories.

		RELEVANCE	
		relevant	irrelevant
NOVELTY	novel	Toothbrushes are made of ice. tool : toolbox :: toothbrush : ? freezer, garage, tool shed, bathroom	People drink gasoline. tree : forest :: water : ? boat, wood, fish, lake
	familiar	Pistols are weapons. dagger : knife :: pistol : ? outlaw, holster, gun, steel	Zebras live in Africa. leopard : spot :: zebra : ? stripe, tail, hoof, mark

Figure 7. *Categories of items in CFAV.*

FlexArt Test

This measure of mental flexibility is designed within the framework of the contextual subtheory. FlexArt was an ideological extension of the other proposed assessments of mental flexibility at the experiential level. However, whereas tests validating the componential and experiential subtheories presented participants with geometrical figures and verbal analogies, a goal of FlexArt was to employ a more complex stimulus—more closely associated with the everyday experience, and consequently, tapping into the practical processing aspect of the theory of successful intelligence.

FlexArt stimuli were reproductions of paintings, drawings, and photographs. The decision to employ two-dimensional art images in the test of mental flexibility had been dictated by the following: a need to move from artificial problem-solving towards reasoning with more natural concepts (external validity objective), and to present a multifaceted stimulus that favors analysis from versatile perspectives (novel situation objective).

FlexArt lays its foundation in the omnipresence of visual images in the form of logos, photographs, illustrations, or reproductions of traditional museum art. Designs rich in detail serve a variety of communication and/or aesthetic functions, which are to be deciphered by their viewers. Interpreting art requires attention to, for example, color, analogy, implicit or explicit messages, etc. Changing criteria for "solving" an image forces viewers to move across the levels of analysis in search of a proper interpretation. Thus, "reading" a visual design might be defined as a type of common problem-solving activity. FlexArt capitalized on Robert L. Solso's idea that much of art has been purposely designed to generate a form of creative tension in the viewer that "cries out for resolution" (2003, p. 237)

A research psychologist and undergraduate student intern selected images and developed FlexArt test items. Items were designed to require respondents to apply the components of successful intelligence—in particular mental flexibility—to classify pictures according to the changing criteria. FlexArt asked questions about the interrelatedness among the art images and the fit of other images within a discovered relationship. No previous experience with art was required for successful completion of the test. It was predicted that FlexArt would correlate with other measures of mental flexibility and creativity and with participants' grade point average.

In each of the 17 items developed for this test, participants are presented with a set of three images of artwork. The task is to complete this set by selecting a fourth image. The participant is requested to give a rating according to the goodness of fit for each of the three answer options provided. In the case where there is no "excellent fit" among the answer options, the participant must describe an image that would represent an excellent fit. In terms of Sternberg's contextual subtheory, the appropriate answer on the latter of these items can be characterized as "shaping," whereas the identification of the excellent fit among the answer options refers to "adaptation." Another category of items within this newly developed test is represented by items where no common theme between the three art images can be inferred. The expected appropriate answer to these items would be to move to the next item. In terms of Sternberg's contextual subtheory, these items require "selection" as an intelligent response.

It is expected that performance in this test will be indicative of a person's ability to adapt to, to select, and to shape the environment in novel situations, which refers to the three kinds of strategies available for achieving success via the application of analytical, creative, and practical abilities, as specified in Sternberg's theory of successful intelligence (Sternberg, 1997). Samples of test questions are displayed in Figures 8–10.

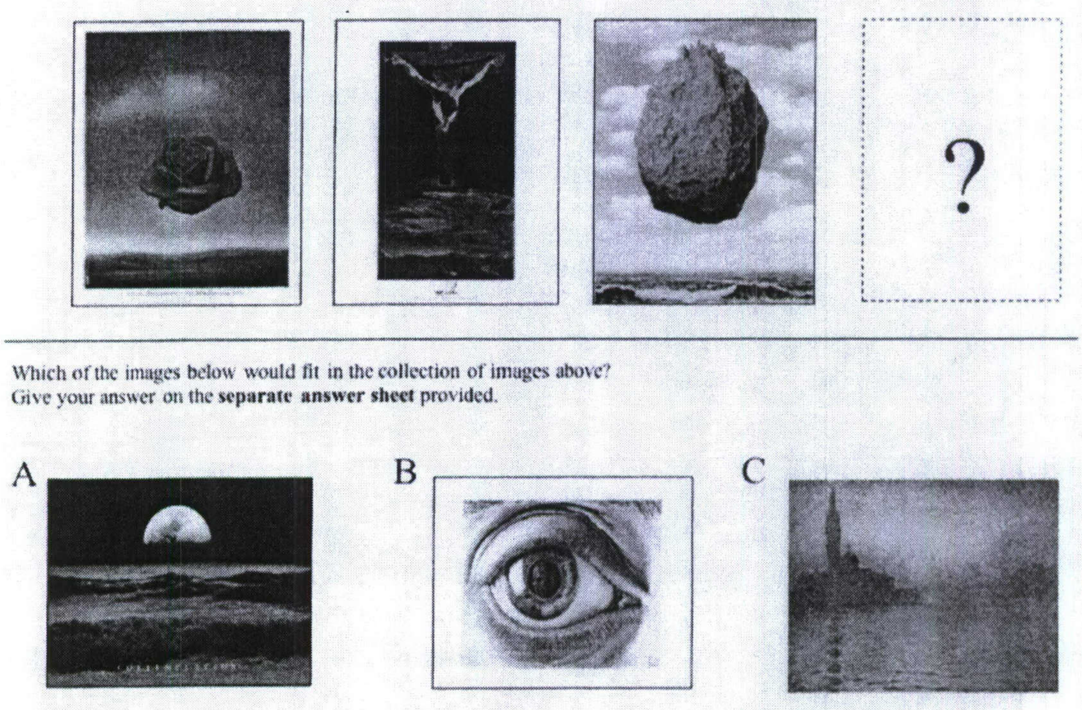

Figure 8. *Example of an "Adapt" item in FlexArt.*

Note. The common theme in the upper set of items could be described as "floating." Answer option A would represent an excellent fit.

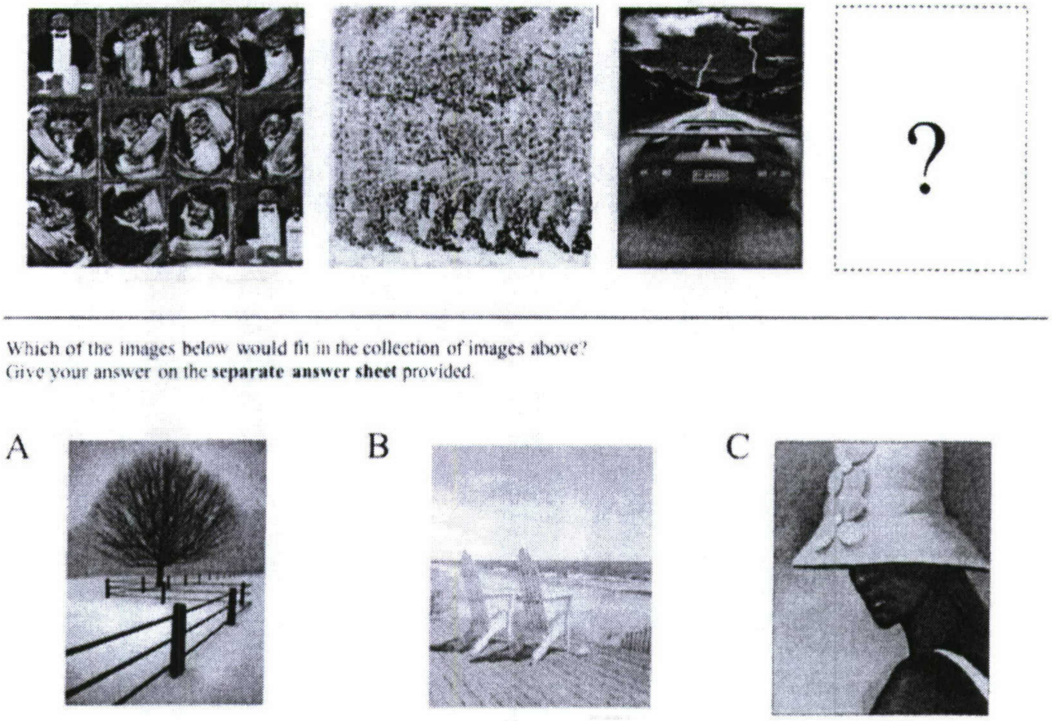

Figure 9. *Example of a "Shape" in FlexArt.*

Note. The common theme in the upper set of items could be described as "movement." Since none of the answer options A to C represent an excellent fit, the test taker is expected to describe an image that would be an excellent fit.

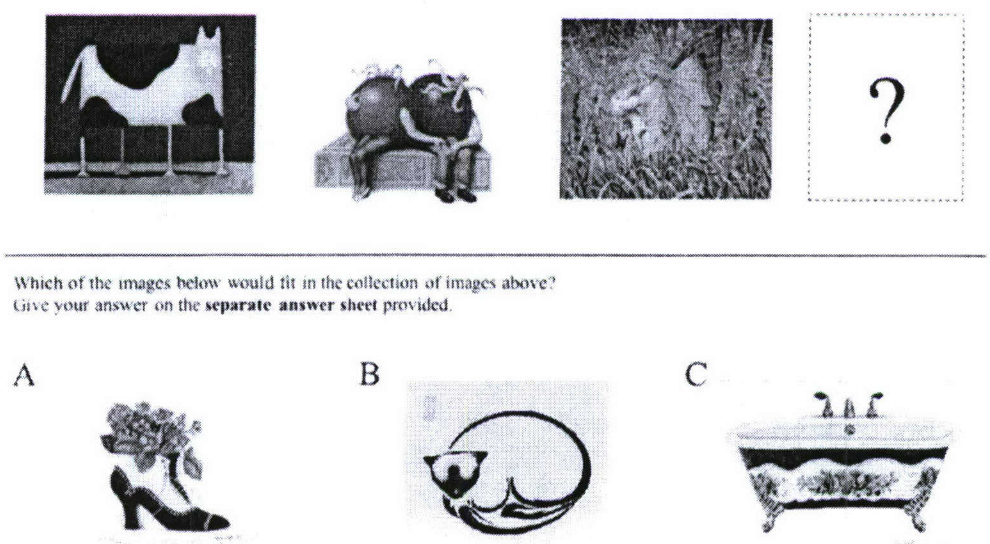

Figure 10. *Example of a "Select" in FlexArt.*

Note. There is no common theme in the upper set of items. The test taker is expected to choose the option "next item" on the answer sheet.

FIELD TESTING

Investigation 1 Formative Evaluation

Purpose

The purpose of this investigation was to field test newly developed tests of mental flexibility. Flexible Inference and Flexible Mapping tests underwent presentation mode, factorial structure, and item analyses. Counterfactual Analogies–Figural, and Counterfactual Analogies–Verbal tests were incomplete at the time of field testing for Investigation 1; full item analyses are reported in Investigations 2 of this manuscript. The FlexArt test was not analyzed because of too few participants.

As a first step, the original item pools for the Flexible Inference and Flexible Mapping tests were screened by three cognitive psychologists and rated regarding classification scheme. In pre-pilot work, the tests were administered to 25 participants to test the reliability of the computer program, the comprehensibility of the instructions, and the screen design and interface. Based on these data and the feedback received, the instruction phase was improved. After revisions of the initial item pool for tests, they were field tested. The primary goals of these investigations were to test the psychometric qualities of the items and to investigate the dependency of the psychometric characteristics of the tests on the different experimental conditions described previously and displayed in Tables 1 and 2. Based on the results obtained, the item pool for each test was reduced and the final procedure for presenting the item triplets was determined.

Method

Participants

A total of 314 underclassman from three universities in the Northeast and one in the Northwest volunteered to participate in an investigation to evaluate newly developed mental flexibility tests. Participants were recruited through fliers and e-mail announcements. Participants were told the purpose of the research was to explore how we "think outside the box" and they were paid $40 for participation in either a single 3 ½ hour testing session or two 1 ½ hour testing sessions. Demographic data were provided by 278 of the participants, of which 69% (n = 193) were female and 31% (n = 85) male. The average age of participants was 19.9 years old.

Reference Measures

Berlin Model of Intelligence Structure (BIS) (Jäger, 1982, 1984). The BIS is a bimodal hierarchical model for describing broad intellectual abilities in the framework of four operational components (processing speed, memory, creativity, and processing capacity) and three content-based components (figural, verbal, numerical). The combination of three content-based and four operational components determines 12 facets of performance.

Eight timed subtests from the BIS4, the most recent version, were administered. Seven content-based subtests (two figural, three numerical, and two verbal) from the creativity operational component (ZF, OJ, ZG, DR, ZR, AM & MA, respectively) and one figural subtest from the processing capacity operation component (BG) were administered. A description of subtests can be seen in Appendix A.

The creativity operation component is defined as fluid, flexible, and original production of ideas requiring the availability of diverse information, wealth of imagination, and ability to see many different sides, variations, reasons, and possibilities in problem-oriented (not purely imaginative) solutions. The processing capacity operational component is defined as the processing of complex information in tasks that are not immediately solvable and that require establishing diverse relations and using exact formal logical reasoning.

The BIS, which is available in German and has been translated into Brazilian Portuguese (Kleine & Jäger, 1987, 1989) and Chilean Spanish (Rosas, 1991), has shown differential as well as predictive validity in other cultures and language environments (Bucik & Neubauer, 1996). Processing speed has been shown to be related to fluid intelligence (Beauducel & Kersting, 2002). The eight selected subtests described above were translated from German into English for purposes of this investigation. Two native German researchers, who were very familiar with the BIS, together translated the selected BIS subtests into English. These translated tests were then reviewed by three native English-speaking research assistants for feedback on meaning and understanding. Translated tests were then modified accordingly.

French Kit of Factor-Referenced Cognitive Tests (F-Kit) Ekstrom, French, Harman, & Dermen, 1976). This test battery is made up of a set of 72 marker tests for 23 cognitive aptitude factors. Two timed subtests (Letter Sets Test–I-1 (rev.); Locations Test–I-2) of the three that make up the Induction factor, one timed subtest (Toothpicks Test–XF-1) of the three that make up the Flexibility (Figural) factor, and one timed subtest (Making Groups–XU-3) of the four that make up the Flexibility of Use factor were administered. The induction factor is defined as reasoning abilities involved in forming and trying out hypotheses that will fit a set of data. The Flexibility (figural) factor is defined as the ability to change set in order to generate new and different solutions to figural problems. The Flexibility of Use factor is defined as the mental set necessary to think of different uses for objects. Adequate reliability and validity has been reported in Ekstrom et al. (1976).

Cognitive Flexibility Scale (Martin & Rubin, 1995). This self-report survey measures three components of cognitive flexibility including: (a) awareness of available options and alternatives; (b) willingness to be flexible and adapt to situations, and (c) self-efficacy in being flexible. The 12-item scale is made up of statements that respondents rate on a 6-point scale, ranging from 1 (strongly disagree) to 6 (strongly agree). A sample item reads, "I can communicate an idea in many different ways." Adequate reliability has been reported. Construct validity has been established in relation to communication competence and confidence, assertiveness, and responsiveness (Martin & Anderson, 1998).

NEO-Personality Inventory Revised (Costa & McCrae, 1992). This personality survey measures five dimensions: neuroticism, extroversion, openness, agreeableness, and conscientiousness. The short form (NEO-FFI) was administered, which contains 60 items that are traditionally rated on a 5-point scale (1 = strongly disagree; 5 = strongly agree). Participants were randomly assigned 2 of the 5 subtests via computer administration. Responses were made on a continuous scale (slider) that ranged from 0 to 100 units to the third decimal point (0 = strongly disagree; 100 = strongly agree). Internal consistency values ranging from .86 to .92 have been reported for the short form. Evidence of adequate content, construct, and criterion-related validity has been reported in Costa & McCrae (1992).

Procedure

Participants took part in either two 1 ½ hour group testing sessions or a single 3 ½ hour group testing session. Sample sizes ranged from 10 to 25 participants per session. Various newly developed mental flexibility and validation measures were administered via paper-and-pencil and computer administration. Data were collected in sessions that varied the order of test administration and the specific validation measures administered. Research designs are illustrated in Tables 3a, 3b and 3c.

Table 3
Administration Designs

Scheme 1				
Session	Tests /Test order	Medium	Content	Purpose
Flexible Mapping Session	GPA-questionnaire	PC, P & P	high school GPA; first- year college GPA	criterion, success from high school to college
	Divergent calculus BIS-DR	P&P	divergent thinking, numerical	reference
	Memory	PC	memory span, all domains, recognition and recall	potential covariate for Flexible Mapping test – sequential mode
	Toothpick test (FKit-TP)	P&P	adaptive flexibility, figural	reference
	Drawings completion BIS-ZF	P&P	divergent thinking, figural	reference
	Locations test (FKit-LC)	P&P	induction, reasoning numerical/figural	reference
	Multiple uses BIS-AM	P&P	divergent thinking, verbal	reference
	Flexible Mapping	PC	analogies, mental flexibility	predictor
Flexible Inference Session	NEO-questionnaire	PC	personality traits (e.g., openness and extraversion)	reference
	Letter set test (French-Kit)	P&P	classification, induction	reference
	Object design BIS-OJ	P&P	divergent thinking, figural domain	reference
	Bongard BIS-BG	P&P	classification, induction	reference
	Number Riddles BIS-ZR	P&P	divergent thinking, numerical	reference
	Making groups (FKit-MG)	P&P	classification, flexibility of use, verbal	reference
	Masselon BIS-MA	P&P	divergent thinking, verbal	reference
	Flexible Inference	PC	mental flexibility	predictor

Note. PC: computerized tests, P&P: paper-and-pencil tests: BIS: Berlin Structure of Intelligence Test.

Table 3

Administration Designs (Continued)

Session	Tests /Test order	Medium	Content	Purpose
Scheme 2				
Session	Tests /Test order	Medium	Content	Purpose
Flexible Inference Session	GPA - Questionnaire	PC, P&P	high school GPA; first- year college GPA	criterion, success from high school to college
	Cog. Flexibility Scale	P&P	personality	reference
	Locations	P&P	induction, reasoning numerical/figural	reference
	ZG-Divergent Equations BIS	P&P	divergent thinking, numerical	reference
	BG-Bongard BIS	P&P	classification, induction	reference
	Insight	P&P	flex	predictor
	CFA verb A (odd sets)	PC	flex, verbal	predictor
	NEO	PC	personality	reference
Flexible Mapping Session	Letter Set	P&P	classification, induction	reference
	ZF-Drawing Completions BIS	P&P	divergent thinking, figural	reference
	AM-Multiple Uses BIS	P&P	divergent thinking, verbal	reference
	CFA fig	P&P	flex, figural	reference
	CFA verb B (even sets)	P&P	flex, verbal	reference
	Flexible Mapping	P&P	flex	predictor
Scheme 3				
Session	Tests /Test order	Medium	Content	Purpose
CFAV Session	GPA - questionnaire	PC P&P	high school GPA; first- year college GPA	criterion, success from high school to college
	NEO	P&P	personality	reference
	Cog. Flexibility Scale	P&P	personality	reference
	Locations (F-Kit)	P&P	induction, reasoning numerical/figural	reference
	ZG-Divergent Equations BIS	P&P	divergent, numerical	reference
	CFA verb	P&P	flex, verbal	predictor
	BG-Bongard BIS	PC	classification, induction	reference
	Letter Set (F-Kit)	PC	personality	reference
Insight & FlexArt Session	ZF-Drawing Completions BIS	P&P	classification, induction	reference
	Insight	P&P	flex	predictor
	AM-Multiple Uses BIS	P&P	divergent thinking, verbal	reference
	Counterfac Analog figural	P&P	flex, figural	predictor
	FlexArt	P&P	flex	predictor
	Letter Set (F-Kit)	P&P	flex	predictor
	ZF-Drawing Completions BIS		divergent thinking, figural	reference

Note. PC: computerized tests, P&P: paper-and-pencil tests: BIS: Berlin Structure of Intelligence Test.

Results

Item Analyses: Test of Effects of Different Presentation Modes

Flexible Inference (FI). Tables 4–9 detail results of item analyses conducted for Flexible Inference items. The first column entails the item identification. Here the letters N, S, and V stand for the domains *numerical*, *shape (figural)*, and *verbal*. The last digit in the item name reflects whether it is an item in which domain typical inferences (.1) or domain atypical inferences (.2, .3, respectively) need to be drawn. Similarly, in Flexible Mapping, the label of domain-homogeneous analogies ends with ".1," and domain-heterogeneous item labels end with either ".2" or ".3." The next four columns in these tables show the distracter probabilities. Column p (item) represents item difficulty. The next four columns show the distracter probability within the subgroup of the 27% lowest performers. The underlined numbers represents the probability of a correct answer within each group. Column d27± contains the discrimination index based on a 27% split. The discrimination efficiency (deffic) represents the ratio between d27± and the maximum discrimination given the difficulty of the item. The next two columns show the point-biserial (rpbis) and biserial (rbis) correlation of the item and the total score. Total scores were summed correct responses across domains.

Items were omitted from the final version of the test according to the following criteria:

1. A distracter was picked by more participants within the 27% best performers than as the correct answer option (p(distracter)27+ > p(correct)27+).
2. The discrimination index d was smaller than .30 in combination with a discrimination efficiency of less than 50% (d27± < .30 AND deffic < .50).
3. The biserial correlation was smaller than .30 (rpbis < .30 AND rbis < .30).
4. One or more of the distracters was never picked (p(distracter)all) = .00.

Flexible Mapping (FM). Tables 10–15 display item analytic results for Flexible Mapping. The same criteria were applied to identify psychometrically problematic items.

Table 4

Flexible Inference, Domain-Typical Inference Item Subset A (1 to 5)

Item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n1.1	.73	.17	.04	.05	.73	.79	.17	.00	.04	.67	.13	.13	.08	.13	.23	.20	.27
n2.1	.03	.77	.10	.10	.77	.00	.92	.00	.08	.00	.67	.21	.13	.25	.60	.29	.40
n3.1	.21	.12	.25	.42	.42	.08	.13	.08	.71	.46	.13	.25	.17	.54	.62	.43	.54
n4.1	.09	.83	.08	.01	.83	.00	1.00	.00	.00	.21	.63	.17	.00	.38	1.00	.34	.49
n5.1	.46	.14	.21	.20	.46	.58	.08	.29	.04	.13	.13	.33	.42	.46	.65	.43	.54
s1.1	.00	.01	.40	.59	.59	.00	.00	.13	.88	.00	.04	.63	.33	.54	.68	.44	.55
s2.1	.26	.02	.00	.72	.72	.13	.00	.00	.88	.38	.04	.00	.58	.29	.54	.18	.24
s3.1	.13	.67	.11	.09	.67	.00	.96	.04	.00	.33	.33	.21	.13	.63	.88	.44	.57
s4.1	.08	.01	.02	.89	.89	.00	.00	.00	1.00	.21	.04	.00	.75	.25	1.00	.29	.48
s5.1	.01	.10	.89	.00	.89	.00	.00	1.00	.00	.04	.25	.71	.00	.29	1.00	.29	.48
v1.1	.04	.04	.89	.02	.89	.00	.00	1.00	.00	.08	.08	.75	.08	.25	1.00	.36	.59
v2.1	.79	.07	.04	.10	.79	.96	.00	.00	.04	.63	.08	.08	.21	.33	.80	.32	.46
v3.1	.63	.03	.05	.28	.28	.50	.00	.00	.50	.63	.00	.17	.21	.29	.41	.26	.34
v4.1	.08	.05	.76	.11	.76	.04	.00	.92	.04	.13	.04	.63	.21	.29	.64	.34	.34
v5.1	.03	.84	.13	.00	.84	.00	.92	.08	.00	.13	.63	.25	.00	.29	.64	.38	.57

Table 5

Flexible Inference, Domain-Atypical Inference Item Subset A (1 to 5)

Item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n1.2	.21	.48	.20	.12	.48	.17	.63	.17	.04	.21	.46	.21	.13	.17	.18	.09	.11
n1.3	.52	.12	.10	.26	.52	.75	.00	.08	.17	.33	.21	.08	.38	.42	.45	.34	.43
n2.2	.39	.15	.32	.14	.39	.63	.21	.13	.04	.21	.08	.33	.38	.42	.50	.30	.38
n2.3	.09	.83	.05	.03	.83	.00	1.00	.00	.00	.17	.71	.13	.00	.29	1.00	.32	.47
n3.2	.09	.58	.10	.24	.58	.04	.83	.04	.08	.13	.29	.17	.42	.54	.62	.40	.50
n3.3	.28	.48	.14	.10	.28	.50	.42	.08	.00	.04	.58	.21	.17	.46	.85	.41	.54
n4.2	.27	.52	.16	.04	.16	.29	.38	.29	.04	.29	.63	.08	.00	.21	.55	.16	.24
n4.3	.18	.37	.21	.24	.24	.13	.38	.17	.33	.21	.38	.33	.08	.25	.60	.18	.25
n5.2	.17	.15	.52	.15	.52	.13	.17	.63	.08	.08	.17	.46	.29	.17	.18	.23	.29
n5.3	.13	.58	.20	.10	.58	.08	.75	.04	.13	.25	.46	.21	.08	.29	.37	.31	.39
s1.2	.02	.96	.00	.02	.96	.04	.96	.00	.00	.00	.96	.00	.04	.00	.00	.06	.13
s1.3	.29	.65	.03	.02	.29	.50	.50	.00	.00	.04	.88	.08	.00	.46	.85	.39	.51
s2.2	.16	.27	.54	.02	.54	.04	.13	.83	.00	.38	.50	.13	.00	.71	.74	.56	.70
s2.3	.38	.11	.10	.41	.41	.13	.08	.04	.75	.54	.00	.17	.29	.46	.48	.31	.40
s3.2	.03	.18	.35	.43	.43	.04	.08	.17	.71	.04	.38	.50	.08	.63	.79	.48	.61
s3.3	.09	.83	.09	.00	.83	.04	.96	.00	.00	.25	.58	.17	.00	.38	.82	.45	.66
s4.2	.13	.45	.16	.26	.45	.08	.71	.08	.13	.21	.17	.21	.42	.54	.62	.48	.60
s4.3	.17	.42	.11	.29	.17	.33	.54	.00	.13	.08	.42	.04	.46	.25	.60	.17	.25
s5.2	.08	.03	.04	.85	.85	.00	.00	.00	1.00	.21	.04	.17	.58	.42	1.00	.48	.73
s5.3	.72	.01	.16	.11	.16	.67	.00	.13	.21	.71	.00	.21	.08	-.08	-.25	-.06	-.09
v1.2	.02	.73	.03	.22	.73	.00	1.00	.00	.00	.04	.33	.08	.54	.67	1.00	.53	.71
v1.3	.30	.14	.12	.43	.30	.54	.13	.00	.33	.13	.13	.17	.58	.42	.63	.41	.53
v2.2	.07	.62	.14	.17	.62	.04	.83	.00	.13	.04	.21	.42	.33	.63	.65	.47	.60
v2.3	.16	.27	.35	.22	.27	.29	.46	.17	.08	.04	.08	.54	.33	.38	.69	.40	.54
v3.2	.11	.27	.52	.10	.11	.29	.17	.50	.04	.00	.33	.54	.13	.29	1.00	.36	.61
v3.3	.10	.05	.33	.52	.33	.08	.04	.58	.29	.13	.08	.08	.71	.50	.75	.42	.54
v4.2	.24	.15	.24	.37	.37	.25	.13	.25	.38	.29	.13	.29	.29	.08	.12	.08	.11
v4.3	.07	.08	.76	.10	.76	.04	.00	.92	.04	.13	.21	.50	.17	.42	.72	.46	.63
v5.2	.90	.02	.03	.04	.90	.96	.00	.00	.04	.75	.04	.13	.08	.21	.71	.33	.57
v5.3	.14	.12	.16	.58	.14	.21	.13	.08	.58	.04	.08	.42	.46	.17	.67	.18	.27

Table 6

Flexible Inference, Domain-Typical Inference Item Subset B (6 to 10)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n6.1	.10	.80	.04	.06	.80	.00	<u>1.00</u>	.00	.00	.29	<u>.46</u>	.17	.08	.54	1.00	.56	.56
n7.1	.83	.00	.14	.02	.83	<u>.96</u>	.00	.04	.00	<u>.54</u>	.00	.38	.08	.42	.83	.53	.53
n8.1	.04	.03	.88	.04	.88	.00	.00	<u>1.00</u>	.00	.08	.13	<u>.67</u>	.13	.33	1.00	.37	.37
n9.1	.29	.04	.13	.53	.53	.21	.00	.08	<u>.71</u>	.33	.13	.17	<u>.38</u>	.33	.36	.25	.32
n10.1	.91	.08	.00	.01	.91	<u>1.00</u>	.00	.00	.00	<u>.75</u>	.25	.00	.00	.25	1.00	.34	.34
s6.1	.02	.02	.00	.96	.96	.00	.00	.00	<u>1.00</u>	.04	.04	.00	<u>.92</u>	.08	1.00	.19	.42
s7.1	.16	.01	.01	.82	.82	.00	.00	.00	<u>1.00</u>	.38	.00	.00	<u>.63</u>	.38	1.00	.49	.71
s8.1	.06	.12	.01	.81	.81	.00	.04	.00	<u>.96</u>	.17	.21	.04	<u>.58</u>	.38	.82	.47	.68
s9.1	.12	.14	.69	.04	.69	.00	.00	<u>1.00</u>	.00	.17	.33	<u>.33</u>	.17	.67	1.00	.58	.75
s10.1	.82	.01	.00	.17	.82	<u>.96</u>	.00	.00	.04	<u>.71</u>	.04	.00	.25	.25	.75	.33	.48
v6.1	.16	.03	.67	.14	.67	.00	.04	<u>.96</u>	.00	.33	.04	<u>.33</u>	.29	.63	.88	.49	.63
v7.1	.72	.17	.08	.03	.72	<u>.92</u>	.04	.04	.00	<u>.50</u>	.29	.08	.13	.42	.72	.41	.54
v8.1	.09	.28	.06	.58	.58	.00	.00	.00	<u>1.00</u>	.17	.33	.08	<u>.42</u>	.58	1.00	.40	.51
v9.1	.00	.00	.97	.03	.97	.00	.00	<u>.96</u>	.04	.00	.00	<u>.96</u>	.04	.00	.00	.10	.24
v10.1	.28	.03	.60	.09	.60	.08	.00	<u>.88</u>	.04	.46	.04	<u>.38</u>	.13	.50	.67	.33	.42

Table 7

Flexible Inference, Domain-Atypical Inference Item Subset B (6 to 10)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n6.2	.21	.22	.11	.46	.21	.33	.25	.04	.38	.08	.08	.21	.63	.25	.60	.20	.29
n6.3	.02	.92	.01	.04	.92	.00	.96	.00	.04	.08	.79	.04	.08	.17	.67	.28	.51
n7.2	.02	.00	.02	.96	.96	.00	.00	.00	1.00	.08	.00	.00	.92	.08	1.00	.19	.42
n7.3	.20	.06	.31	.43	.31	.17	.00	.25	.58	.17	.04	.46	.33	-.21	-.29	-.09	-.12
n8.2	.10	.58	.11	.21	.58	.04	.83	.08	.04	.17	.29	.08	.46	.54	.62	.44	.55
n8.3	.40	.33	.14	.12	.33	.29	.54	.08	.08	.42	.17	.25	.17	.38	.53	.30	.39
n9.2	.88	.04	.04	.03	.88	1.00	.00	.00	.00	.79	.08	.08	.04	.21	1.00	.26	.42
n9.3	.10	.38	.09	.43	.10	.17	.33	.08	.42	.13	.38	.04	.46	.04	.14	.00	-.01
n10.2	.49	.21	.13	.17	.17	.54	.08	.08	.29	.38	.42	.17	.04	.25	.75	.18	.27
n10.3	.59	.17	.19	.06	.59	.67	.13	.04	.17	.33	.17	.46	.04	.33	.33	.24	.31
s6.2	.69	.10	.09	.12	.69	.92	.00	.00	.08	.54	.21	.17	.08	.38	.69	.33	.43
s6.3	.18	.02	.80	.00	.80	.00	.00	1.00	.00	.33	.00	.67	.00	.33	1.00	.37	.53
s7.2	.01	.94	.02	.02	.94	.04	.96	.00	.00	.00	.92	.04	.04	.04	.34	.09	.18
s7.3	.09	.11	.76	.04	.76	.00	.04	.96	.00	.17	.25	.50	.08	.46	.85	.42	.58
s8.2	.03	.09	.77	.11	.77	.00	.00	.96	.04	.13	.21	.38	.29	.58	.87	.54	.74
s8.3	.14	.04	.48	.33	.48	.08	.04	.71	.17	.25	.08	.08	.58	.63	.79	.45	.56
s9.2	.07	.13	.03	.77	.77	.00	.04	.04	.92	.17	.21	.08	.54	.38	.69	.36	.50
s9.3	.66	.08	.18	.09	.08	.54	.04	.33	.08	.63	.13	.21	.04	-.08	-.50	-.10	-.18
s10.2	.16	.10	.03	.71	.03	.08	.17	.04	.71	.33	.08	.00	.58	.04	1.00	.15	.35
s10.3	.10	.22	.40	.28	.40	.17	.13	.58	.13	.04	.08	.33	.54	.25	.27	.22	.28
v6.2	.61	.32	.01	.06	.61	.83	.13	.00	.04	.17	.79	.04	.00	.67	.67	.58	.73
v6.3	.24	.10	.10	.56	.56	.08	.04	.00	.88	.38	.21	.29	.13	.75	.75	.58	.72
v7.2	.32	.21	.40	.07	.40	.00	.17	.79	.04	.63	.21	.04	.13	.75	.90	.59	.75
v7.3	.21	.09	.31	.39	.39	.04	.04	.08	.83	.50	.08	.29	.13	.71	.74	.53	.66
v8.2	.62	.09	.16	.13	.62	.92	.04	.00	.04	.33	.17	.33	.17	.58	.78	.47	.60
v8.3	.29	.12	.36	.23	.23	.21	.04	.33	.42	.33	.21	.42	.04	.38	.82	.36	.49
v9.2	.47	.24	.22	.07	.22	.42	.13	.46	.00	.25	.67	.08	.00	.38	.69	.36	.51
v9.3	.47	.27	.04	.22	.27	.29	.42	.13	.17	.79	.08	.00	.13	.33	.67	.32	.43
v10.2	.29	.47	.14	.10	.47	.08	.83	.08	.00	.42	.13	.21	.25	.71	.74	.53	.67
v10.3	.20	.30	.19	.31	.20	.38	.13	.29	.21	.08	.38	.13	.42	.29	.64	.24	.34

Table 8

Flexible Inference, Domain-Typical Inference Item Subset C (11 to 15)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n11.1	.84	.04	.09	.03	.84	<u>1.00</u>	.00	.00	.00	<u>.68</u>	.08	.16	.08	.32	1.00	.36	.53
n12.1	.18	.12	.09	.62	.62	.00	.04	.04	<u>.92</u>	.32	.16	.20	<u>.32</u>	.60	.79	.51	.51
n13.1	.96	.02	.02	.00	.96	<u>1.00</u>	.00	.00	.00	<u>.88</u>	.08	.04	.00	.12	1.00	.27	.60
n14.1	.43	.19	.11	.28	.19	.28	<u>.48</u>	.08	.16	.44	<u>.04</u>	.24	.28	.44	.85	.45	.65
n15.1	.05	.06	.82	.06	.82	.00	.00	<u>1.00</u>	.00	.12	.08	<u>.72</u>	.08	.28	1.00	.35	.52
s11.1	.94	.04	.00	.02	.94	<u>1.00</u>	.00	.00	.00	<u>.84</u>	.12	.00	.04	.16	1.00	.29	.56
s12.1	.05	.88	.04	.02	.88	.00	<u>1.00</u>	.00	.00	.08	<u>.76</u>	.08	.08	.24	1.00	.32	.52
s13.1	.09	.40	.22	.29	.40	.00	<u>.64</u>	.08	.28	.12	<u>.12</u>	.36	.40	.52	.68	.41	.51
s14.1	.50	.31	.04	.15	.50	<u>.56</u>	.36	.00	.08	<u>.28</u>	.28	.16	.28	.28	.33	.14	.17
s15.1	.19	.06	.06	.68	.68	.08	.00	.00	<u>.92</u>	.44	.08	.12	<u>.36</u>	.56	.78	.51	.67
v11.1	.22	.12	.54	.12	.54	.08	.12	<u>.80</u>	.00	.28	.08	<u>.40</u>	.24	.40	.50	.29	.36
v12.1	.04	.87	.06	.02	.87	.00	<u>1.00</u>	.00	.00	.08	<u>.76</u>	.16	.00	.24	1.00	.33	.53
v13.1	.38	.33	.24	.04	.24	.28	.20	<u>.48</u>	.04	.40	.44	<u>.08</u>	.08	.40	.71	.45	.62
v14.1	.24	.43	.13	.20	.24	<u>.40</u>	.24	.12	.24	<u>.20</u>	.64	.08	.08	.20	.33	.21	.29
v15.1	.09	.04	.01	.86	.86	.04	.00	.00	<u>.96</u>	.12	.16	.04	<u>.68</u>	.28	.78	.21	.32

Table 9

Flexible Inference, Domain-Atypical Inference Item Subset C (11 to 15)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n11.2	.19	.09	.54	.18	.54	.16	.08	.68	.08	.24	.00	.44	.32	.24	.27	.24	.30
n11.3	.95	.03	.00	.02	.95	1.00	.00	.00	.00	.80	.12	.00	.08	.20	1.00	.40	.83
n12.2	.06	.82	.04	.07	.82	.00	1.00	.00	.00	.12	.60	.12	.16	.40	1.00	.36	.52
n12.3	.73	.05	.09	.13	.09	.80	.04	.08	.08	.56	.12	.08	.24	.00	.00	.02	.03
n13.2	.01	.01	.95	.03	.95	.00	.00	1.00	.00	.04	.04	.80	.12	.20	1.00	.28	.58
n13.3	.22	.13	.44	.21	.44	.20	.04	.68	.08	.32	.20	.28	.20	.40	.42	.32	.40
n14.2	.10	.10	.66	.15	.66	.04	.00	.80	.16	.04	.24	.40	.32	.40	.50	.35	.45
n14.3	.37	.27	.23	.13	.27	.32	.32	.16	.20	.40	.32	.20	.08	.00	.00	.13	.18
n15.2	.45	.24	.21	.10	.10	.24	.44	.20	.12	.64	.12	.20	.04	.08	.50	.14	.24
n15.3	.71	.02	.07	.19	.71	.92	.00	.08	.00	.48	.08	.08	.36	.44	.73	.37	.49
s11.2	.15	.04	.72	.09	.72	.00	.04	.96	.00	.40	.08	.24	.28	.72	.90	.60	.80
s11.3	.17	.31	.16	.36	.31	.16	.40	.00	.44	.20	.12	.36	.32	.28	.54	.30	.40
s12.2	.12	.13	.61	.15	.61	.00	.04	.96	.00	.28	.12	.24	.36	.72	.90	.54	.68
s12.3	.31	.61	.01	.07	.61	.08	.88	.00	.04	.72	.12	.00	.16	.76	.76	.61	.77
s13.2	.12	.07	.41	.39	.41	.04	.00	.68	.28	.28	.12	.28	.32	.40	.42	.28	.36
s13.3	.49	.14	.09	.29	.49	.80	.08	.04	.08	.20	.20	.08	.52	.60	.60	.44	.55
s14.2	.34	.00	.04	.62	.62	.08	.00	.00	.92	.60	.00	.08	.32	.60	.79	.49	.62
s14.3	.12	.40	.21	.27	.21	.20	.16	.52	.12	.00	.48	.04	.48	.48	.86	.46	.65
s15.2	.03	.09	.69	.19	.69	.00	.04	.92	.04	.08	.08	.36	.48	.56	.78	.51	.66
s15.3	.23	.05	.65	.06	.65	.40	.00	.52	.08	.08	.12	.72	.08	-.20	-.26	-.21	-.27
v11.2	.31	.50	.06	.13	.31	.64	.24	.04	.08	.16	.68	.04	.12	.48	.60	.42	.54
v11.3	.03	.06	.60	.31	.60	.00	.00	.96	.04	.00	.12	.24	.64	.72	.90	.51	.65
v12.2	.21	.33	.29	.17	.21	.20	.16	.44	.20	.16	.44	.28	.12	.04	.11	.04	.06
v12.3	.23	.32	.31	.14	.32	.16	.44	.32	.08	.36	.16	.32	.16	.28	.47	.35	.45
v13.2	.66	.03	.21	.10	.66	.76	.00	.16	.08	.48	.12	.24	.16	.28	.37	.27	.34
v13.3	.27	.05	.07	.61	.61	.20	.00	.00	.80	.28	.16	.12	.44	.36	.47	.33	.42
v14.2	.11	.83	.05	.01	.83	.08	.92	.00	.00	.24	.64	.08	.04	.28	.64	.34	.50
v14.3	.31	.33	.16	.20	.31	.36	.20	.20	.24	.24	.44	.12	.20	.12	.20	.21	.27
v15.2	.11	.17	.31	.41	.17	.12	.40	.28	.20	.12	.04	.44	.40	.36	.82	.38	.57
v15.3	.29	.29	.30	.13	.30	.20	.12	.60	.08	.40	.32	.20	.08	.40	.50	.33	.44

Table 10

Flexible Mapping, Domain-Homogeneous Analogies, Item Subset A (1 to 5)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n1.1	.29	.02	.66	.03	.66	.04	.00	.96	.00	.63	.04	.29	.04	.67	.89	.59	.76
n2.1	.01	.00	.12	.87	.87	.00	.00	.00	1.00	.04	.00	.25	.71	.29	1.00	.33	.52
n3.1	.19	.01	.62	.19	.62	.04	.00	.88	.08	.33	.00	.33	.33	.54	.68	.47	.59
n4.1	.04	.00	.03	.92	.92	.04	.00	.04	.92	.04	.00	.04	.92	.00	.00	.05	.10
n5.1	.19	.68	.07	.07	.68	.04	.88	.04	.04	.38	.42	.13	.08	.46	.65	.38	.49
s1.1	.03	.85	.03	.08	.85	.00	.92	.00	.08	.00	.79	.08	.13	.13	.43	.09	.14
s2.1	.01	.03	.95	.01	.95	.00	.00	1.00	.00	.00	.08	.88	.04	.13	1.00	.24	.50
s3.1	.33	.46	.20	.01	.20	.21	.54	.25	.00	.42	.38	.21	.00	.04	.09	.00	.01
s4.1	.32	.16	.41	.11	.41	.13	.04	.79	.04	.42	.21	.17	.21	.63	.65	.48	.61
s5.1	.22	.43	.21	.14	.14	.29	.38	.21	.13	.17	.54	.25	.04	.08	.50	.15	.22
v1.1	.09	.24	.60	.07	.60	.04	.08	.88	.00	.21	.38	.29	.13	.58	.70	.48	.60
v2.1	.55	.02	.41	.02	.41	.25	.00	.75	.00	.75	.08	.13	.04	.63	.71	.47	.59
v3.1	.03	.87	.02	.08	.87	.00	.96	.00	.04	.00	.92	.00	.08	.04	.34	.08	.13
v4.1	.08	.90	.01	.01	.90	.00	1.00	.00	.00	.25	.71	.04	.00	.29	1.00	.37	.63
v5.1	.02	.00	.98	.00	.98	.00	.00	1.00	.00	.08	.00	.92	.00	.08	1.00	.21	.58

Table 11

Flexible Mapping, Domain-Heterogeneous Analogies, Item Subset A (1 to 5)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n1.2	.74	.15	.09	.02	.74	.96	.00	.04	.00	.46	.21	.25	.08	.50	.86	.46	.62
n1.3	.32	.63	.04	.01	.63	.33	.63	.04	.00	.38	.58	.04	.00	.04	.05	.04	.05
n2.2	.46	.25	.27	.01	.27	.42	.13	.46	.00	.46	.42	.13	.00	.33	.57	.26	.34
n2.3	.05	.23	.59	.12	.59	.00	.13	.88	.00	.08	.33	.21	.38	.67	.73	.50	.63
n3.2	.16	.02	.80	.01	.80	.04	.04	.92	.00	.38	.00	.58	.04	.33	.67	.33	.47
n3.3	.04	.19	.73	.04	.73	.00	.08	.88	.04	.04	.21	.63	.13	.25	.50	.20	.26
n4.2	.02	.22	.76	.00	.76	.00	.13	.88	.00	.08	.33	.58	.00	.29	.54	.27	.37
n4.3	.34	.33	.01	.32	.33	.29	.42	.00	.29	.25	.21	.04	.50	.21	.33	.17	.23
n5.2	.09	.34	.43	.14	.43	.04	.29	.54	.13	.08	.46	.25	.21	.29	.37	.26	.33
n5.3	.26	.32	.41	.01	.41	.08	.29	.58	.04	.33	.38	.29	.00	.29	.33	.28	.36
s1.2	.18	.11	.69	.02	.69	.21	.04	.75	.00	.21	.21	.54	.04	.21	.29	.16	.22
s1.3	.12	.08	.02	.78	.78	.00	.00	.00	1.00	.13	.17	.08	.63	.38	1.00	.32	.45
s2.2	.43	.01	.51	.05	.51	.04	.00	.92	.04	.75	.00	.17	.08	.75	.82	.59	.73
s2.3	.37	.45	.02	.15	.45	.17	.79	.00	.04	.63	.17	.04	.17	.63	.65	.51	.64
s3.2	.38	.15	.25	.21	.25	.50	.08	.38	.04	.29	.08	.17	.46	.21	.38	.20	.27
s3.3	.51	.08	.16	.25	.51	.63	.08	.13	.17	.38	.04	.13	.46	.25	.25	.27	.33
s4.2	.15	.22	.11	.52	.15	.42	.17	.04	.38	.04	.42	.17	.38	.38	.82	.47	.71
s4.3	.10	.40	.48	.02	.48	.00	.13	.83	.04	.13	.63	.25	.00	.58	.64	.45	.56
s5.2	.43	.26	.21	.10	.43	.50	.38	.08	.04	.21	.25	.38	.17	.29	.41	.09	.11
s5.3	.03	.08	.70	.19	.70	.00	.08	.83	.08	.08	.04	.54	.33	.29	.47	.26	.34
v1.2	.02	.87	.11	.00	.87	.00	.79	.21	.00	.00	.83	.17	.00	-.04	-.11	.02	.03
v1.3	.99	.01	.00	.00	.99	1.00	.00	.00	.00	.96	.04	.00	.00	.04	1.00	.14	.49
v2.2	.11	.47	.12	.30	.30	.21	.13	.13	.54	.08	.58	.17	.17	.38	.53	.32	.42
v2.3	.36	.23	.03	.37	.37	.33	.08	.00	.58	.42	.38	.04	.17	.42	.56	.38	.48
v3.2	.08	.05	.65	.22	.65	.04	.08	.79	.08	.13	.04	.50	.33	.29	.41	.21	.27
v3.3	.09	.12	.79	.00	.79	.04	.00	.96	.00	.08	.17	.75	.00	.21	.71	.20	.28
v4.2	.00	.21	.02	.77	.77	.00	.04	.00	.96	.00	.29	.04	.67	.29	.78	.22	.31
v4.3	.05	.42	.11	.42	.42	.04	.33	.04	.58	.04	.38	.25	.33	.25	.35	.24	.30
v5.2	.03	.90	.02	.04	.90	.00	1.00	.00	.00	.04	.79	.04	.13	.21	1.00	.28	.48
v5.3	.44	.07	.43	.07	.07	.38	.04	.50	.08	.46	.08	.38	.08	.00	.00	.07	.13

Table 12

Flexible Mapping, Domain-Homogeneous Analogies, Item Subset B (6 to 10)

item	p(A)	p(B)	p(C)	p(D)	p (item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n6.1	.04	.93	.00	.02	.93	.00	1.00	.00	.00	.17	.79	.00	.04	.21	1.00	.40	.76
n7.1	.09	.61	.20	.10	.61	.00	.92	.04	.04	.13	.42	.25	.21	.50	.75	.44	.56
n8.1	.13	.17	.09	.61	.13	.21	.08	.04	.67	.08	.21	.13	.58	.13	.43	.18	.28
n9.1	.79	.03	.12	.06	.79	.96	.00	.04	.00	.54	.08	.21	.17	.42	.83	.43	.60
n10.1	.13	.13	.67	.06	.67	.00	.04	.96	.00	.29	.21	.46	.04	.50	.86	.40	.52
s6.1	.61	.36	.01	.02	.36	.50	.50	.00	.00	.71	.25	.00	.04	.25	.33	.21	.26
s7.1	.92	.01	.07	.00	.07	.96	.00	.04	.00	.92	.04	.04	.00	.00	.00	.03	.06
s8.1	.04	.00	.02	.93	.93	.00	.00	.00	1.00	.08	.00	.04	.88	.13	1.00	.12	.23
s9.1	.48	.43	.01	.08	.43	.21	.79	.00	.00	.58	.21	.04	.17	.58	.58	.42	.53
s10.1	.52	.40	.02	.06	.52	.88	.13	.00	.00	.33	.58	.04	.04	.54	.68	.43	.53
v6.1	.02	.76	.11	.10	.76	.00	.92	.04	.04	.04	.58	.21	.17	.33	.67	.34	.47
v7.1	.17	.79	.03	.01	.79	.13	.88	.00	.00	.33	.54	.08	.04	.33	.57	.26	.36
v8.1	.00	.01	.96	.02	.96	.00	.00	1.00	.00	.00	.00	.92	.04	.08	1.00	.22	.48
v9.1	.01	.97	.01	.01	.97	.00	.96	.00	.04	.04	.92	.04	.00	.04	.34	.17	.41
v10.1	.11	.38	.44	.07	.38	.04	.50	.46	.00	.29	.13	.46	.13	.38	.60	.31	.40

Table 13

Flexible Mapping, Domain-Heterogeneous Analogies, Item Subset B (6 to 10)

item	p(A)	p(B)	p(C)	p(D)	p(item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n6.2	.49	.02	.04	.44	.49	.33	.00	.04	.63	.50	.04	.04	.42	-.17	-.20	-.11	-.14
n6.3	.12	.13	.72	.02	.72	.04	.08	.88	.00	.21	.13	.63	.04	.25	.50	.20	.26
n7.2	.31	.01	.34	.34	.34	.13	.04	.58	.25	.54	.00	.00	.46	.58	1.00	.44	.56
n7.3	.07	.80	.02	.11	.80	.00	.83	.04	.13	.25	.63	.00	.13	.21	.38	.20	.28
n8.2	.25	.10	.11	.54	.54	.17	.04	.08	.71	.38	.13	.13	.38	.33	.36	.29	.36
n8.3	.08	.72	.08	.12	.72	.13	.67	.04	.17	.04	.71	.13	.13	-.04	-.07	.01	.01
n9.2	.10	.04	.82	.03	.82	.00	.04	.92	.04	.25	.13	.63	.00	.29	.64	.30	.43
n9.3	.06	.08	.43	.44	.43	.00	.08	.63	.29	.13	.13	.25	.50	.38	.43	.33	.41
n10.2	.38	.37	.17	.08	.37	.17	.63	.21	.00	.46	.25	.21	.08	.38	.43	.36	.45
n10.3	.10	.13	.67	.09	.67	.08	.08	.79	.04	.13	.17	.67	.04	.13	.23	.18	.23
s6.2	.39	.27	.11	.22	.39	.71	.13	.08	.08	.08	.50	.13	.29	.63	.79	.53	.67
s6.3	.25	.04	.71	.00	.71	.08	.00	.92	.00	.46	.13	.42	.00	.50	.75	.40	.53
s7.2	.01	.01	.93	.04	.01	.00	.00	1.00	.00	.00	.04	.83	.13	.00	.00	.03	.12
s7.3	.17	.11	.07	.65	.65	.00	.08	.04	.88	.38	.17	.13	.33	.54	.68	.48	.62
s8.2	.12	.60	.02	.26	.60	.00	.92	.00	.08	.29	.08	.04	.58	.83	.83	.67	.84
s8.3	.01	.76	.02	.20	.76	.00	.92	.00	.08	.04	.42	.08	.46	.50	.75	.45	.61
s9.2	.18	.24	.47	.11	.47	.04	.00	.92	.04	.33	.38	.17	.13	.75	.82	.54	.68
s9.3	.08	.84	.07	.01	.84	.00	.96	.00	.04	.17	.71	.13	.00	.25	.75	.35	.53
s10.2	.27	.61	.07	.06	.07	.00	.88	.04	.08	.50	.38	.08	.04	-.04	-.34	-.09	-.17
s10.3	.21	.55	.02	.21	.21	.04	.67	.00	.29	.50	.33	.04	.13	.17	.31	.22	.31
v6.2	.85	.08	.07	.00	.85	1.00	.00	.00	.00	.71	.17	.13	.00	.29	1.00	.38	.58
v6.3	.01	.96	.03	.00	.96	.00	1.00	.00	.00	.04	.83	.13	.00	.17	1.00	.36	.78
v7.2	.03	.48	.47	.01	.47	.00	.21	.79	.00	.04	.63	.29	.04	.50	.55	.39	.49
v7.3	.07	.70	.01	.22	.70	.00	1.00	.00	.00	.13	.33	.00	.54	.67	1.00	.52	.69
v8.2	.11	.39	.42	.08	.42	.08	.29	.63	.00	.13	.54	.21	.13	.42	.50	.26	.33
v8.3	.88	.03	.06	.03	.88	1.00	.00	.00	.00	.71	.08	.13	.08	.29	1.00	.42	.67
v9.2	.30	.67	.02	.00	.67	.13	.83	.04	.00	.67	.33	.00	.00	.50	.60	.47	.60
v9.3	.08	.08	.04	.80	.80	.00	.00	.00	1.00	.13	.21	.17	.50	.50	1.00	.54	.76
v10.2	.08	.01	.04	.87	.87	.00	.00	.00	1.00	.17	.00	.08	.75	.25	1.00	.34	.53
v10.3	.37	.15	.42	.07	.42	.17	.04	.71	.08	.50	.29	.13	.08	.58	.70	.48	.61

Table 14

Flexible Mapping, Domain-Homogeneous Analogies, Item Subset C (11 to 15)

item	p(A)	p(B)	p(C)	p(D)	p(item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27±	deffic	rpbis	rbis
n11.1	.03	.19	.45	.32	.45	.00	.04	.83	.13	.13	.44	.09	.35	.74	.81	.53	.66
n12.1	.08	.00	.89	.03	.89	.00	.00	1.00	.00	.17	.00	.70	.13	.30	1.00	.47	.77
n13.1	.06	.63	.11	.20	.63	.09	.74	.09	.09	.13	.48	.17	.22	.26	.33	.19	.25
n14.1	.47	.31	.11	.11	.47	.74	.13	.04	.09	.26	.44	.22	.09	.48	.48	.40	.50
n15.1	.19	.61	.15	.05	.61	.00	.96	.04	.00	.39	.39	.13	.09	.57	.87	.51	.64
s11.1	.03	.01	.68	.27	.68	.00	.00	1.00	.00	.13	.04	.30	.52	.70	1.00	.58	.75
s12.1	.23	.08	.56	.14	.56	.09	.04	.87	.00	.26	.17	.22	.35	.65	.71	.53	.66
s13.1	.32	.10	.07	.51	.51	.13	.04	.00	.83	.35	.26	.17	.22	.61	.64	.45	.57
s14.1	.05	.38	.42	.16	.38	.00	.87	.13	.00	.13	.09	.52	.26	.78	.82	.58	.74
s15.1	.40	.40	.19	.01	.19	.30	.30	.39	.00	.44	.48	.09	.00	.30	.64	.25	.35
v11.1	.09	.09	.02	.80	.80	.00	.04	.09	.87	.26	.13	.00	.61	.26	.50	.31	.44
v12.1	.02	.98	.00	.00	.98	.00	1.00	.00	.00	.04	.96	.00	.00	.04	1.00	.13	.36
v13.1	.08	.90	.00	.02	.90	.04	.96	.00	.00	.13	.83	.00	.04	.13	.60	.30	.51
v14.1	.03	.22	.65	.10	.65	.00	.09	.83	.09	.13	.30	.39	.17	.44	.56	.39	.50
v15.1	.14	.06	.13	.68	.68	.00	.00	.00	1.00	.39	.09	.22	.30	.70	1.00	.57	.74

Table 15

Flexible Mapping, Domain-Heterogeneous Analogies, Item Subset C (10 to 15)

item	p(A)	p(B)	p(C)	p(D)	p(item)	p(A) 27+	p(B) 27+	p(C) 27+	p(D) 27+	p(A) 27–	p(B) 27–	p(C) 27–	p(D) 27–	d27 ±	deffic	rpbis	rbis
n11.2	.07	.64	.01	.28	.64	.00	.83	.00	.17	.17	.44	.04	.35	.39	.53	.40	.51
n11.3	.11	.80	.09	.00	.80	.09	.83	.09	.00	.22	.74	.04	.00	.09	.20	.06	.09
n12.2	.00	.34	.01	.65	.65	.00	.04	.00	.96	.00	.70	.04	.26	.70	.89	.45	.57
n12.3	.01	.93	.03	.02	.93	.00	1.00	.00	.00	.04	.78	.09	.09	.22	1.00	.39	.74
n13.2	.14	.14	.59	.14	.59	.00	.09	.91	.00	.30	.17	.22	.30	.70	.80	.54	.68
n13.3	.06	.23	.58	.14	.58	.13	.13	.70	.04	.09	.17	.52	.22	.17	.22	.15	.19
n14.2	.23	.26	.19	.32	.32	.17	.17	.09	.57	.30	.35	.17	.17	.39	.53	.25	.33
n14.3	.45	.16	.15	.24	.45	.87	.04	.09	.00	.13	.30	.13	.44	.74	.74	.51	.63
n15.2	.13	.09	.75	.03	.75	.04	.04	.91	.00	.30	.13	.57	.00	.35	.67	.44	.59
n15.3	.05	.72	.17	.07	.72	.00	.91	.04	.04	.13	.39	.44	.04	.52	.75	.54	.71
s11.2	.08	.06	.05	.82	.82	.00	.00	.04	.96	.17	.09	.04	.70	.26	.75	.24	.36
s11.3	.74	.15	.07	.05	.74	.91	.09	.00	.00	.48	.30	.13	.09	.44	.71	.38	.51
s12.2	.65	.10	.16	.09	.16	.39	.09	.39	.13	.78	.17	.00	.04	.39	1.00	.37	.56
s12.3	.08	.82	.03	.07	.82	.04	.78	.04	.13	.13	.70	.04	.13	.09	.17	.08	.11
s13.2	.24	.09	.27	.40	.27	.04	.00	.35	.61	.39	.13	.17	.30	.17	.33	.15	.20
s13.3	.49	.32	.06	.14	.32	.30	.61	.00	.09	.74	.09	.04	.13	.52	.75	.44	.57
s14.2	.73	.08	.14	.06	.73	.87	.04	.04	.04	.48	.13	.26	.13	.39	.60	.43	.58
s14.3	.08	.15	.08	.69	.69	.00	.00	.00	1.00	.13	.22	.30	.35	.65	1.00	.55	.72
s15.2	.42	.11	.22	.25	.42	.61	.04	.13	.22	.13	.26	.22	.39	.48	.65	.41	.51
s15.3	.18	.58	.19	.05	.58	.09	.70	.17	.04	.26	.52	.22	.00	.17	.22	.26	.33
v11.2	.02	.93	.05	.00	.93	.00	1.00	.00	.00	.04	.87	.09	.00	.13	1.00	.21	.40
v11.3	.14	.66	.17	.03	.66	.04	.91	.00	.04	.39	.26	.26	.09	.65	.79	.53	.68
v12.2	.93	.05	.00	.02	.93	.96	.04	.00	.00	.83	.09	.00	.09	.13	.60	.14	.27
v12.3	.03	.01	.02	.93	.93	.00	.00	.00	1.00	.13	.04	.04	.78	.22	1.00	.37	.70
v13.2	.10	.31	.17	.42	.31	.00	.52	.22	.26	.13	.09	.17	.61	.44	.71	.44	.58
v13.3	.48	.08	.15	.30	.48	.74	.00	.09	.17	.35	.17	.30	.17	.39	.43	.25	.31
v14.2	.00	.17	.18	.65	.65	.00	.00	.04	.96	.00	.39	.39	.22	.74	.89	.54	.70
v14.3	.19	.74	.05	.02	.74	.00	1.00	.00	.00	.48	.39	.13	.00	.61	1.00	.55	.74
v15.2	.90	.01	.01	.08	.90	1.00	.00	.00	.00	.74	.04	.04	.17	.26	1.00	.40	.68
v15.3	.05	.26	.02	.67	.67	.00	.00	.00	1.00	.04	.43	.09	.44	.57	1.00	.52	.67

Test of Effects of Different Presentation Modes

Flexible Inference. To test whether the variation of domain-typical versus domain-atypical inferences within each item triplet causes costs in performance and/or response latencies, a Multivariate Analysis of Variance (MANOVA) was conducted. The dependent variables were aggregated number of correct responses (performance) and response times (latencies). The within-subject factor "typicality of inferences" has two factor levels (domain typical versus domain atypical) and the between-subject factor "intra-item triplet order" has 6 factor levels (all combinations of orders of three item triplet parts). As expected, "typicality" causes significant effects (on performance: $F(1,270) = 492.68$, $p < .001$, $eta^2 = .65$; on latencies: $F(1,270) = 186.41$, $p < .001$, $eta^2 = .41$). Item triplet parts that require the inference of domain-atypical rules are more difficult, and more time is needed to solve domain-atypical inference items than domain-typical inference items. The intra-item triplet order, however, causes no statistically significant effect on either test performance ($F(5,270) = 1.52, p > .10$, $eta^2 = .03$) or response latencies ($F(5,270) = 1.68, p > .10$, $eta^2 = .03$).

Two main conclusions can be drawn from these results. First, they give evidence that the developed extension of the item design for classification tasks causes systematic inter-individual variability. The question of whether this variability is systematically indicative to mental flexibility needs to be addressed in validity-oriented research that are reported later. Second, all items, independent from their presentation order, can be combined in later item analyses.

Flexible Mapping. In Flexible Mapping, two different presentation modes were employed: parts of item triplets were presented either individually on a separate screen or together on one single screen. Because each participant was presented with both modes, counterbalanced across individuals, the effect of a within-subject factor on performance and latency can be analyzed. Since in Flexible Mapping all item triplets start with a domain-homogeneous item, the intra-item triplet order varies only on two factor levels. Most relevant for the validity of the item design (introduction of domain-heterogeneous analogies within each item triplet) was the effect of "homogeneity" of the analogy on performance and latency.

A MANOVA was conducted to test whether the variation of domain-homogeneous versus domain-heterogeneous inferences within each item triplet causes costs in performance and/or response latencies. A significant main effect of homogeneity on performance ($F(1,266) = 10.84$, $p = .001$, eta2 = .04) and on response latencies ($F(1,266) = 301.85$, $p < .001$, eta2 = .53) was found. However, whereas performance declines on domain-heterogeneous items, the response latencies are shorter. This disparity can be explained by the fact that no encoding and inference processes were necessary in the second or third appearance of the analogy stem in the domain-heterogeneous part of each item triplet.

The presentation mode (grouped vs. sequential) caused no statistically significant effect on performance ($F(1,266) = .10, p = .76$, $eta^2 = .00$) but a small effect on latency ($F(1,266) = 5.13, p = .02$, $eta^2 = .019$), indicating that item triplets presented sequentially were answered slightly slower than group-presented item triplets. The variation of the order of the domain-heterogeneous analogies within each item triplet did not cause any effect, either on performance ($F(1,266) = .41, p = .52$, $eta^2 = .002$) or on latency ($F(1,266) = .08, p = .78$, $eta^2 = .00$).

These results indicate that the introduction of domain-heterogeneous analogies caused a systematic increase of difficulty to the items, although the analogy stem remains the same from their domain-homogeneous counterparts. The response latencies in domain-heterogeneous item triplet parts are shorter since no re-encoding, or re-inference of the relation between the elements of the analogy stem was necessary.

That no other effect was found made it possible for all items, independent of their item triplet part orders or their presentation mode, to be combined in the item analyses.

Factorial Structure of the Mental Flexibility Tests

Flexible Inference. A principal-component analysis, specifying initial Eigenvalue greater than 1 with Varimax rotation, was conducted to explore the factorial structure of the item pool of the Flexible Inference test. Therefore, all items in which the inference of domain-typical relations is required are aggregated to a performance score (FI domain 1) for each domain. All items in which the inference of domain-atypical relations is required are aggregated to another performance score (FI domain 2/3), expecting that the latter reflects a separable latent factor.

Table 16
Flexible Inference, PCA With Varimax Rotation, Components Matrix

	1	2
FI num 1	.00	**.78**
FI fig 1	.27	**.65**
FI verb 1	.11	**.72**
FI num 2/3	**.68**	.10
FI fig 2/3	**.79**	.15
FI verb 2/3	**.73**	.13

Note. Bold values are included in component.

The analysis results in a two-factor solution with a clear distinction between domain-typical and domain-atypical inference of relations in classification items, displayed in Table 3. These two factors explained 56% of the variance, with the first factor explaining 29% and the second 27% of variance. Indicators were included when loading was greater than .60 (Stevens, 1996).

Based on this factor solution, individual factor scores were calculated for some preliminary analyses of association with reference tests.

Flexible Mapping. An analogous analysis was conducted for Flexible Mapping.

Table 17
Flexible Mapping, PCA With Varimax Rotation, Components Matrix

	1	2
FM num 1	.15	**.82**
FM fig 1	*.42*	*.41*
FM verb 1	.13	**.75**
FM num 2/3	**.68**	.28
FM fig 2/3	**.77**	.22
FM verb 2/3	**.87**	.00

Note. Bold values are included in component. Italicized values are double-loaded.

If forced to a two-factor solution (Eigenvalue of the second component: .97), we also obtain a similar factorial structure as compared with that observed in Flexible Inference. As can be seen in Table 5, except for the performance aggregate for domain-homogeneous analogies using figural stimuli, a clear distinction between the two hypothesized latent factors can be seen. These two factors explained 60% of the variance, with the first factor explaining 34% and the second 26% of variance. Indicators were included when loading was greater than .60 (Stevens, 1996). One domain homogeneous indicator was double-loaded.

Based on this factor solution, individual factor scores were calculated for some preliminary analyses of association with reference tests.

Preliminary Construct Validation

In a first step, the factorial structure of the reference tests was explored, as with the flexibility tests. Two components that together explained 55% of the variance were found. The first component explained 28% and the second 27% of the variance.

Table 18
Reference Tests, PCA With Varimax Rotation, Components Matrix

	1	2
Letter Sets KIT	**.70**	.00
Locations KIT	**.67**	.25
Bongrad BIS	**.77**	-.17
Diverg Calculus BIS	.35	**.62**
Multiple Uses BIS	.11	**.77**
Drawing Completion BIS	-.16	**.75**

Note. Bold values are included in component.

Component 1 contains tests for fluid intelligence, and component 2 reflects performance in divergent-thinking tasks. Based on this result, individual factor scores for convergent abilities and for divergent abilities were calculated.

Next, a second-order factor analysis was conducted combining the two factor scores from Flexible Mapping, the two from Flexible Inference, and the two from the reference tests. The result was a second-order two factor structure that explained 50% of variance with each component explaining 25%.

Table 19
Second-Order PCA With Varimax Rotation, Component Matrix

	1	2
FI factor score atypical	**.71**	.28
FI factor score typical	-.17	**.63**
FM factor score heterogeneous	**.75**	.00
FM factor score homogeneous	.15	**.77**
Divergent factor score	**.51**	.00
Convergent factor score	.35	**.65**

Component 1 can be interpreted as a flexibility component, whereas component 2 reflects the fluid abilities measured by traditional approaches.

Summary

As a result of the foregoing item analyses, Flexible Inference was reduced to 18 items made up of 6-item triplets in figural, verbal, and numerical content domains (54 item parts), and Flexible Mapping was reduced to 21 items made up of 7-item triplets in each of the figural, verbal, and numerical domains (63 item parts). Counterfactual Verbal was reduced to 48 items equally divided among familiar relevant, familiar irrelevant, counterfactual relevant, and counterfactual irrelevant premise types.

Investigation 2: Summative Evaluation

Purpose

The purpose of this research was to examine the construct and criterion-related validity of the newly developed tests of Flexible Inference (FI), Flexible Mapping (FM), Counterfactual Analogies–Figural (CFAF), Counterfactual Analogies–Verbal (CFAV), and Insight. The primary objectives included: (1) internal analysis of CFAF, CFAV, and Insight tests; (2) assessment of construct and criterion-related validity of each new mental flexibility test by comparisons with tests of cognitive ability, personality, and pattern recognition; (3) assessment of the validity of the full test battery; and (4) partial construct validation of the theory of successful intelligence.

Method

Participants

A total of 476 college students from five private undergraduate institutions in the Northeast volunteered to participate in the investigation. Participants were recruited through fliers and e-mail announcements. They were told the purpose of the research was to explore how we "think outside the box" and they were paid $40 for their participation in a single 3- to 3.5-hour testing session.

A total of 462 participants completed the demographic survey. The average age of participants was 19.4 years of age with a range of 17 to 27 years and a standard deviation of 1.3 years. Of these participants, 70.6% were female (n = 326) and 29.4% were male (n = 136). The vast majority of participants were native English speaking (95.4%, n = 443); 4.1% (n = 19) of participants were not native English speaking. In terms of ethnic background, 3.5% (n = 16) were African American, 4.3% (n = 20) were Asian American, 6.5% were Hispanic American (n = 30), and 85.7% were European American/Other (n = 396). The average number of semesters of college completed by participants was 2.8 with a range of 0 to 8 and a standard deviation of 2.1 semesters.

Measures – Mental Flexibility

Mental Flexibility. Based on the foregoing formative analysis, a reduced set of items for each mental flexibility test was used in this investigation.

Flexible Inference. This new test of mental flexibility is computer-administered and made up of classification problems designed to assess the ability to infer relations flexibly. It contains 18 items: 6-item triplets in figural, verbal, and numerical content domains (54 item parts). Each item contains an item prompt and a set of four domain-consistent response pair options, one of which must be linked to the prompt by inferring common properties. Each item triplet contains the same stimuli. The three parts of an item differ in the arrangement of elements in response pairs such that a common property must be inferred to link the prompt to the correct pair. To solve an item part, previously inferred relations must be inhibited and new ones identified. The inferred relation that links a prompt to the correct response pair option is classified as *domain typical* or *domain atypical*. *Domain-typical* relations are based on properties that are dominant and might typically be considered in the domain of reference. *Domain-atypical* relations are based on properties that might be secondary and would less often be considered in the domain of reference. Each item triplet is made up of one part that requires a domain-typical inference to identify the correct match and two parts that require domain-atypical inferences. FI accuracy scores are calculated by taking the mean of correct domain-typical responses (part 1) and domain-atypical responses (mean of parts 2 & 3). The rationale for aggregation of domain-typical and domain-atypical items rests in the expectation that performance on the test-as-a-whole captures the ability to respond correctly when item types are presented alternatively; in other words, when the respondent is required to switch his or her thinking from domain-typical to domain-atypical in the same testing session. Response latencies are calculated similarly. The coefficient alpha estimate of reliability for this testing session was .82.

Flexible Mapping. This new test of mental flexibility is computer-administered and made up of analogy problems designed to assess the ability to map inferred rules across content domains. It contains 21 items: 7-item triplets in each of the figural, verbal, and numerical domains (63 item parts). Each item triplet is made up of an analogy that varies in terms of the content domain to which it must be mapped. For each triplet, one part must be mapped to the same content domain (*domain homogeneous*), and two that must be mapped to different content domains (*domain heterogeneous*). FM accuracy scores are calculated by taking the mean of correct domain-homogeneous responses (part 1) and correct domain-heterogeneous responses (mean of parts 2 & 3). Response latencies are calculated similarly. Consistent with the rationale for aggregation FI scores, performance on the test-as-a-whole is expected capture the capacity to correctly respond to items when the presentation is alternated between domain homogeneous and domain heterogeneous item types. The coefficient alpha estimate of reliability for this testing session was .83.

Counterfactual Analogies (CFA). This is a new set made up of two subtests of mental flexibility (verbal and figural versions) containing counterfactual (novel) and familiar analogy problems drawn from Sternberg (1987) and designed to assess the ability to cope with relative novelty in the verbal and figural domains. Tests contain a mix of items, some requiring reasoning based on facts and others requiring reasoning based on counterfactual (novel) premises.

<u>*CFA–Verbal*</u> is a computer-administered test that contains 48 verbal analogy items with four response options. Applying a scheme developed by Sternberg and Gastel (1989a, 1989b), all items are preceded by a premise that is either familiar or counterfactual (novel), and either relevant or irrelevant. Items are equally divided among familiar relevant, familiar irrelevant, counterfactual relevant, and counterfactual irrelevant premise types. Participants are first presented with the premise and given as long as they wish to read it. They then press a button, which results in the disappearance of the premise and the immediate appearance of the analogy item. Accuracy scores for CFA–Verbal are calculated by summing correct responses. The coefficient alpha estimate of reliability for this testing session was .76.

<u>*CFA–Figural*</u> is a computer-administered test that contains 30 figural analogy items with four response options. Applying a partially modified scheme of one developed by Sternberg and Gastel (1989), all items are preceded by a premise that is either familiar or counterfactual (novel). Items were equally divided among these two premise types. Participants are first presented with the premise and analogy stem, and are given as long as they wish to view it. They then press a button, which results in the disappearance of the premise and the immediate appearance of the analogy item. Accuracy scores for CFA–Figural are calculated by summing correct responses. The coefficient alpha estimate of reliability for this testing session was .95.

Insight Test. This new test of mental flexibility is a paper-and-pencil administered test of coping with novelty through insight. It contains nine insight problems drawn from the literature (Fixx, 1972; Metcalfe, 1986a; Seifert & Patalano, 1991; Sternberg & Davidson, 1982; Weisberg, 1988) that represent a mix of verbal, figural, and numerical problem types. Participants are asked to provide open-ended responses to insight problems and are given as long as they wish to complete the test. A sample problem reads as follows, "A bottle of wine costs $10. The wine was worth $9 more than the bottle. How much was the wine worth?" (Sternberg & Davidson, 1982). Responses are dichotomously scored as correct or incorrect by human raters using a scoring rubric. One human rater scored each test. Ambiguous responses were discussed and scored by consensus. The coefficient alpha estimate of reliability of the insight test for this session was .61.

Measures – Cognitive Ability

Berlin Model of Intelligence Structure (BIS) (Jäger, 1982, 1984). Four timed subtests from the BIS4, the most recent version, were administered via paper and pencil. Three of the subtests were content-based (figural, numerical, and verbal) from the creativity operational component (ZF, ZG, and AM), and one was a figural subtest from the processing capacity operation component (BG). These four selected subtests were translated from German into English for the purposes of this research. Two raters scored each of the three creative component subtests (ZF, ZG, and AM). Interrater reliability ranged from r (154) = .85 to r (154) = .99. Summary scores were calculated for each of the four subtests, ZF, ZG, AM, and BG. A BIS creativity mean score was calculated from ZF, ZG, and AM summary scores, and a BIS processing score that reflected the BG summary score. Adequate reliability and validity have been reported by Jager (1982, 1984).

French Kit of Factor-Referenced Cognitive Tests (F-Kit: Ekstrom, French, Harman, & Dermen 1976). This test battery is made up of a set of 72 marker tests for 23 cognitive aptitude factors. Two subtests (Letter Sets Test–I-1 (rev.), Locations Test–I-2) of three that comprise the induction factor (convergent) were administered. The induction factor is defined as reasoning abilities involved in forming and trying out hypotheses that will fit a set of data. Letter Sets Test–I-1 (rev.) is a 15-item timed test, in which 5 sets of 4 letters each are presented. The task is to find the rule that relates four of the sets to each other and mark the one that does not fit the rule. Locations Test–I-2 is a 14-item timed test in which 5 rows of places and gaps are given. In each of the first 4 rows, 1 place in each row is marked according to a rule. The task is to discover the rule and to mark 1 of the 5 numbered places in the fifth row. An induction factor score was calculated by aggregating scores on the Letter Sets and Locations Tests. Adequate reliability has been reported. The Letter Sets Test and Locations Test are well-validated measures of fluid intelligence (Ekstrom et al, 1976).

Measures – Pattern Recognition

Soluble/Insoluble Analogy Test. This is a 30-item multiple-choice figural analogy test developed for this investigation. It is designed to measure pattern recognition by comparing response accuracy on soluble versus insoluble items. The American Council on Education Psychological Examination for College Freshman, 1949 edition, (Thurstone, 1925, 1926; Thurstone & Thurstone, 1949), a test of general scholastic aptitude, was adapted to create this measure. Fifteen items were selected at random from the original test and correct response options modified to be incorrect. An "insoluble" response option was added to the response option set for each of the 30 (soluble and insoluble) items. After viewing a figural analogy stem, the respondent chooses 1 of 5 possible figural solutions or the insoluble answer option. Scores are obtained by calculating sensitivity and bias indices according to signal detection theory procedures (Snodgrass & Corwin, 1988) detailed in Appendix B. Cronbach's coefficient alpha estimate of reliability for this test administration was .81. In addition, evidence of construct validity is suggested by a positive correlation between SI sensitivity scores with Group Embedded Figures Test scores (r (183) = .34, p = .00) also administered in this investigation.

Group Embedded Figures Test (GEFT) (Witkin, Oltman, Raskin, & Karp, 1971, 2002). This test is an adaptation of the Embedded Figures Test (EFT) (Witkin, 1950; Witkin, Dyk, Faterson, Goodenough, & Karp, 1962) modified for group administration. It measures competence in perceptual field independence, which has been associated with a global-versus-analytical dimension of cognitive functioning. The participant's task is to locate a previously seen simple figure within a larger complex figure organized to embed the simple figure. Participants are presented with 25 complex figures and must locate a simple figure printed on a separate sheet of paper in a 20-minute timed session. Adequate reliability and validity have been reported by Witkin et al. (1971).

Revised Minnesota Paper Form Board Test (Likert & Quasha, 1970; Paterson, Elliott, Anderson, Toops, & Heidbreder, 1930). This test measures the capacity to visualize and manipulate objects in space. It is a 20-minute speeded test consisting of 64 two-dimensional diagrams cut into separate parts. For each diagram, there are 5 figures with lines indicating the different shapes out of which they are made. From these, the participant chooses the one figure that is composed of the exact parts shown in the original diagram. Series AA was administered. Adequate reliability and validity have been reported by Likert and Quasha (1970).

Minnesota Clerical Test (Andrew & Paterson, 1959). This is a test of perceptual speed and accuracy in recognizing name and number sequence pairs. The first part of the test consists of names that contain 7 to 17 letters, and the second part, number sequences ranging from 3 through 12 digits. Each part contains 200 items consisting of 100 identical pairs and 100 dissimilar pairs. The participant is asked to check the identical pairs in each part. Separate time limits are used for the two parts. The total testing time is 15 minutes. A single score was calculated by taking the mean of scores on test parts. Adequate reliability and validity have been reported by Andrew and Paterson (1959).

Measures – Personality

Cognitive Flexibility Scale (CFS) (Martin & Rubin, 1995). This self-report survey is designed to measure three components of cognitive flexibility including: (a) awareness of available options and alternatives; (b) willingness to be flexible and adapt to situations, and (c) self-efficacy in being flexible. The 12-item scale is made up of statements that respondents rate on a 6-point scale, ranging from 1 (strongly disagree) to 6 (strongly agree). A sample item reads, "I can communicate an idea in many different ways." Adequate reliability has been reported. Construct validity has been established in relation to communication competence and confidence, assertiveness, and responsiveness (Martin & Anderson, 1998).

NEO-Personality Inventory Revised (Costa & McCrae, 1992). This personality survey measures five dimensions including: neuroticism, extroversion, openness, agreeableness, and conscientiousness. The short form (NEO-FFI) was administered, which contains 60 items that are traditionally rated on a 5-point scale (1 = strongly disagree; 5 = strongly agree). Participants were randomly assigned 2 of the 5 subtests via computer administration. Responses were made on a continuous scale (slider) that ranged from 0 to 100 units to the third decimal point (0 = strongly disagree; 100 = strongly agree). Internal consistency values ranging from .86 to .92 have been reported for the short form. Evidence of adequate content, construct, and criterion-related validity has been reported by Costa and McCrae (1992).

Measures – Criterion

College GPA. Participants were asked to report college GPA to date and maximum GPA. Because maximum GPA scores varied from 4.0 to 5.0 depending on the school, scores were calculated by the ratio of GPA-to-date divided by maximum GPA indicated.

Creative awards. Participants were asked to respond to the following question, "Have you received an award or formal recognition for unique, innovative, or creative work?" on a 4-point scale (0 = never; 3 = more than two occasions).

Self-reported flexible thinking. Participants were asked to respond to the following question, "Compared to most people, how well do you 'think on your feet' when faced with an unusual situation or problem," on a 5-point scale (1= worse than most people; 5 = much better than most people).

Self-reported flexible behavior. Participants were asked to respond to the following question, "Compared to most people, how well do you deal with entirely novel situations or problems?" on a 5-point scale (1 = much worse than most people; 5 = exceptional). A summary self-report flexible performance score was calculated.

Procedure

Participants took part in a single, 3- to 3.5-hour group session. Newly developed mental flexibility tests were administered in all sessions and selected validation measures were administered in the form of three sub-investigations, as detailed in Table 20.

Table 20
Sub-Investigation Test Administration Schedule

Sub-Investigation 1			
Mode	**Test**	**Type**	**Domain**
Paper & Pencil	Cog Flex Scale (CFS)	Personality	
	French Kit: Location (1,2) (KIT F)	Convergent	Figural
	French Kit: Letter set (KIT V)	Convergent	Verbal
	Insight	Mental Flex	Verb/Num/Fig
	Berlin Intelligence: BG (BIS- F) (BIS Process)	Convergent	Figural
	Berlin Intelligence: ZG (BIS-N) (BIS Creativity	Divergent	Numerical
	Berlin Intelligence: AM (BIS-V) (BIS Creativity	Divergent	Verbal
	Berlin Intelligence: ZF (BIS-F) (BIS Creativity	Divergent	Figural
Online	Flexible Inference (FI)	Mental Flex	Verb/Num/Fig
	Counterfactual Analogy–Verbal (CFA-V)	Mental Flex	Verbal
	Flexible Mapping (FM)	Mental Flex	Verb/Num/Fig
	Counterfactual Analogy–Figural (CFA-F)	Mental Flex	Figural
	Demographic Survey	Criterion	
Sub-Investigation 2			
Mode	**Test**	**Type**	**Domain**
Paper & Pencil	Group Embedded Figures Test (GEFT)	Field Independence	
	Minn. Paper Form Board Test (PFBT)	Spatial Ability	Nonverbal
	Minn. Clerical Test (MC)	Perceptual Speed	Verbal
	Insight	Mental Flex	Verb/Num/Fig
Online	Flexible Inference (FI)	Mental Flex	Verb/Num/Fig
	Counter Factual Analogy–Verbal (CFA-V)	Mental Flex	Verbal
	Flexible Mapping (FM)	Mental Flex	Verb/Num/Fig
	Counter Factual Analogy–Figural (CFA-F)	Mental Flex	Figural
	Demographic Survey	Criterion	
Sub-Investigation 3			
Mode	**Test**	**Type**	**Domain**
Paper & Pencil	Group Embedded Figures Test (GEFT)	Field Independence	
	Insight	Mental Flex	Verb/Num/Fig
Online	NEO Subscales (NEO)	Personality	
	Flexible Inference (FI)	Mental Flex	Verb/Num/Fig
	Counterfactual Analogy–Verbal (CFA-V)	Mental Flex	Verbal
	Flexible Mapping (FM)	Mental Flex	Verb/Num/Fig
	Counterfactual Analogy–Figural (CFA-F)	Mental Flex	Figural
	Soluble/Insoluble Analogies–SDT (SI)	Pattern Recognition	Figural
	Demographic Survey	Criterion	

After participants completed informed consent forms, a series of timed paper-and-pencil tests were administered. Following the paper-and-pencil test administration, participants took a 10-minute break, during which snacks were provided. After the break, experimenters reviewed online test procedures for self-administration of computer-administered tests. Participants were encouraged to take a brief break after completing half of the online tests to reduce the potential affect of fatigue on performance. Participants completed the online tests at their own pace. Upon completion of the testing session, participants were paid and given a written debriefing handout.

Results

Overview

Descriptive statistics and intercorrelations among all measures are presented first. Next, each mental flexibility test is assessed separately, and the test battery evaluated as a whole. Mental flexibility tests that did not undergo a full internal analysis in Investigation 1 (Counterfactual Analogies–Figural, Counterfactual Analogies–Verbal, and Insight) were examined and revised accordingly. In addition to item analyses, the internal conceptual structure of tests was explored using factor analyses and comparisons of subscale means. To assess construct validity, all new mental flexibility tests underwent correlation and regression analyses with reference tests of cognitive ability, personality, and pattern recognition. To assess criterion-related validity, all tests underwent correlation analyses with criterion measures.

With regard to external validation, small to moderate positive correlations are expected between all of the mental flexibility tests and fluid intelligence tests (KIT Induction; BIS Creative; BIS Processing) and criterion measures (college GPA, self-report flexible performance, creative awards). Small positive correlations are expected between all mental flexibility tests and NEO-openness and CPS.

Incremental and discriminant validity of the full test battery was assessed using factor and regression analyses. The latent structure of the test battery as predicted by the theory of successful intelligence was assessed by testing a structural equation model. Finally, results related to the role of pattern recognition measures are summarized.

Descriptive Statistics and Correlation Analyses

Descriptive statistics and correlation analyses of mental flexibility tests, validation tests, and criterion measures are displayed in Table 21. There were no significant gender differences found in mental flexibility test scores, with the exception of the insight test, in which the mean difference in scores for males was significantly higher than scores for females (t (458) = 2.16, p = .03, two-tailed). There were no significant gender differences found in validation tests. With regard to criterion measures, mean GPA was significantly higher for females than for males (t (408) = -2.61, p = .00, two-tailed). In contrast, males scored higher than females on SR flexible thinking (t (460) = 5.65, p = .00, two-tailed) and SR flexible behavior (t (219, equal variances not assumed) = 3.21, p = .00, two-tailed), but not creative awards.

Intercorrelations among full-test scores of the measures are presented in Table 22. Subtest intercorrelations are examined in individual test analyses that follow.

As expected, there were strong correlations among newly developed mental flexibility tests ranging from .54 to .98, suggestive of convergent validity. In regard the relations to measures of cognitive ability, there were positive, significant correlations between mental flexibility test scores and cognitive ability test scores that were moderate but not high, which were expected and support discriminant validity. Correlations with KIT Fluency scores were strongest and ranged from .42 to .46; correlations with BIS creativity scores ranged from .28 and .32. With BIS processing scores, the range was .21 to .28. The correlation between mental flexibility test scores and measures of pattern recognition were slightly higher than cognitive ability test scores and ranged between .25 to .51.

Trends in the patterns of correlations between the mental flexibility scores and NEO personality subscale scores were consistent with the literature on creative personality type (Barron & Harrington, 1981), with low positive correlations with NEO-Openness and low negative correlations with extraversion.

In regard to criterion measures, mental flexibility test scores were positively correlated with self-reported flexible performance scores ranging from .18 to .23 and also positively correlated with college GPA scores with a range of .20 to .29, which suggests modest criterion validity.

Table 21
Descriptive Statistics for all Measures

	N	Mean	SD	Skew	SE	Kurtosis	SE
			Mental Flexibility				
FI	452	0.56	0.15	0.05	0.11	-0.38	0.23
FM	452	0.56	0.15	0.05	0.11	-0.48	0.23
CFA-V	465	30.03	5.98	-0.26	0.11	0.45	0.23
CFA-F	450	13.46	4.47	0.27	0.11	-0.21	0.23
Insight	470	2.71	1.89	0.74	0.11	0.42	0.23
			Cognitive Ability				
BIS Processing	153	2.00	1.16	0.46	0.20	-0.15	0.39
BIS-N	154	9.98	4.11	0.41	0.20	-0.17	0.39
BIS-V	154	4.30	1.75	0.09	0.20	0.25	0.39
BIS-F	154	5.14	1.39	-0.04	0.20	0.31	0.39
KIT-letter	154	23.00	3.64	-1.31	0.20	2.93	0.39
KIT-location	152	13.44	5.19	-0.12	0.20	-0.42	0.39
			Pattern Recognition				
SI - sensitivity	232	0.04	0.32	-0.39	0.16	-0.27	0.32
SI - bias	232	0.55	0.11	-0.07	0.16	0.84	0.32
GEFT	318	12.82	4.53	-0.69	0.14	-0.29	0.27
PFBT	151	43.56	10.51	-0.53	0.20	0.36	0.39
MC-name	144	124.2	25.00	-0.05	0.20	-0.39	0.40
MC-number	144	121.17	25.31	0.31	0.20	0.20	0.40
			Personality				
NEO-openness	116	61.01	15.91	0.08	0.22	-0.72	0.45
NEO-conscientiousness	127	68.97	15.28	-0.27	0.21	-0.71	0.43
NEO-extraversion	114	69.18	13.88	-0.49	0.23	0.22	0.45
NEO-agreeableness	123	63.91	13.10	-1.08	0.22	3.78	0.43
NEO-neuroticism	124	50.11	16.77	0.13	0.22	0.11	0.43
CFS	154	58.75	4.95	-0.69	0.19	1.12	0.39
			Criterion				
GPA	409	0.82	0.10	-0.61	0.12	0.14	0.24
SR-flexible thinking	461	2.96	0.98	0.27	0.11	-0.71	0.23
SR-flexible behavior	461	3.27	0.68	0.10	0.11	1.06	0.23
SR Flexible performance	461	6.22	1.42	0.34	0.11	-0.14	0.23
Creative award	461	1.39	1.16	0.25	0.11	-1.39	0.23

Table 22

Intercorrelations Among Measures

	FI	FM	CFA	Insight	BIS Process	BIS Creative	KIT Fluid	SI	GEFT	PFBT	MC	NEO O	NEO C	NEO E	NEO A	NEO N	CFS	GPA	SR Flex Perf
FI	1.0																		
FM	.98**	1.0																	
CFA	.73**	.75**	1.0																
Insight	.54**	.55**	.58**	1.0															
BIS Process	.21*	.22**	.26**	.28**	1.0														
BIS Creative	.30**	.31**	.28**	.32**	.19*	1.0													
KIT Fluid	.46**	.46**	.42**	.42**	.37**	.26**	1.0												
SI	.45**	.47**	.51**	.31**	.35*	.15	.37**	1.0											
GEFT	.47**	.47**	.50**	.37**	–	–	–	.34**	1.0										
PFBT	.33**	.37**	.39**	.25**	–	–	–	.23	.51**	1.0									
MC	.07	.07	.14	.12	–	–	–	.00	.14	.30*	1.0								
NEO-O	.17	.18	.20*	.17	.244	.30*	-.02	.24*	.18	–	–	1.0							
NEO-C	-.04	-.02	-.02	-.22*	-.070	.07	-.03	.11	-.24*	–	–	-.02	1.0						
NEO-E	-.26**	-.28**	-.31**	-.21*	-.353*	.20	-.11	-.12	-.26*	–	–	-.18	.47**	1.0					
NEO-A	.12	.11	.18	.10	-.228	.03	.05	.05	.02	–	–	.24	.31	-.03	1.0				
NEO-N	-.08	-.08	-.29*	-.14	.033	.05	.00	-.03	-.15	–	–	-.06	-.29	-.16	-.47**	1.0			
CFS	-.07	-.06	-.03	-.04	-.016	.05	-.11	-.14	–	–	–	.20	.23	.50	.07	-.15	1.0		
GPA	.27	.28**	.22**	.20**	.079	.08	.12	.12	.22**	.16	.15	.22*	.10	-.09	.05	-.02	.03	1.0	
SR Flex Perf	.18**	.17*	.19*	.23**	-.032	.19*	.12	.13	.19**	-.03	.06	.23*	.06	.22*	-.22	-.15	.29**	.13**	1.0
Creative award	.03	.05	.04	.09*	-.157	.07	-.06	.07	.03	-.12	-.00	.28**	.05	.16	-.03	.09	.12	.09	.19*

Note. N per cell varies between 113 and 452; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed)

Individual Test Analyses

Flexible Inference. This classification test of mental flexibility contains 18 items: 6-item triplets in figural, verbal, and numerical content domains (54 item parts). Each item triplet is made up of one part that requires a domain-typical inference to identify the correct match and two parts that require domain-atypical inferences. FI accuracy and latency scores are designed to measure the performance components of mental flexibility. Internal test analyses are reported in the results of Investigation 1. With regard to external validation, small to moderate positive correlations with KIT Induction, BIS Creativity, and BIS Processing, and criterion measures (GPA, SR-Flexible performance and Creative Awards) are expected. Given that the test is based on a componential level of analysis, very low correlations with NEO-Openness and CPS tests of personality are expected.

To explore construct validity, FI accuracy and latency scores were correlated with scores on cognitive ability (Table 23) and personality measures (Table 24). As can be seen in Table 23, FI accuracy scores correlated positively with BIS numerical subtest scores (ZG) (r (150) = .33, *p* = .00), BIS verbal subtest scores (AM) (*r* (150) = .20, *p* = .02), BIS processing capacity (BG) scores (*r* (149) = .21, *p* = .01), and KIT induction factors test scores (Letter: *r* (150) = .43, *p* = .00; Location: *r* (148) = .36, *p* = .00). However, contrary to expectations, FI accuracy scores did not correlate with BIS (BF) figural subtest scores.

Table 23
FI: Correlations With Cognitive Ability Measures

	FI Accuracy	FI Latency	BIS zg num	BIS am verbal	BIS zf figural	BIS bg process	KIT letter	KIT location
FI Accuracy	1.00							
FI Latency	.53**	1.00						
BIS zg num	.33**	.08	1.00					
BIS am verbal	.20*	.10	.05	1.00				
BIS zf figural	.10	.05	.16*	.22**	1.00			
BIS bg process	.21**	-.05	.15	.08	.12	1.00		
KIT letter	.43**	.15	.29**	.17*	.06	.29**	1.00	
KIT location	.36**	-.03	.29**	.07	.08	.34**	.46**	1.00

Note. N per cell varies between 148 and 150; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed)

The correlation between FI accuracy and NEO Openness scores approached significance (*r* (114) = .17, *p* = .07). FI accuracy scores were negatively correlated with NEO Extraversion scores (*r* (113) = -.26, *p* = .01). FI Latency scores were positively correlated with NEO Agreeableness scores (*r* (123) = .32, *p* = .00).

Table 24
FI: Correlations With Personality Measures

	FI Acc	FI Lat	CFS	NEO-Open	NEO-Consc	NEO-Extra	NEO-Agree	NEO-Neuro
FI Accuracy	1.00							
FI Latency	0.53**	1.00						
CFS	-0.07	0.04	1.00					
Openness	.17	.06	.20	1.00				
Conscientious	-.04	.14	.23	-.02	1.00			
Extraversion	.26**	-.07	.50**	-.18	.47**	1.00		
Agreeable	.12	.32**	.07	.24	.31	-.03	1.00	
Neuroticism	-.08	.16	-.15	-.06	-.29	-.16	-.47**	1.00

Note. *N* per cell varies between 114 and 150; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed). Participants were randomly assigned 2 of the 5 NEO personality subtests, resulting in a low *n* per cell.

To examine the cognitive processes that contribute to performance on the FI, FI accuracy scores were regressed on BIS creativity component scores, BIS processing scores and KIT induction factor scores entered together. The regression was significant and the predictors explained 25% of total variance (*F change* (3, 143) = 15.992, $p = .00$). Significant predictors in the model were KIT induction factor scores ($\beta = 0.7$, $t = 5.158$, $p = .00$), and BIS creative component ($\beta = .048$, $t = 2.825$, $p = .00$). BIS processing was not a significant contributor to the regression.

To explore criterion-related validity, FI accuracy scores were correlated with college GPA, self-report flexible performance (sum of self-report thinking and self-report behavior) and creative award. As can be seen in Table 25, FI accuracy scores correlated positively with college GPA (r (382) = .27, $p = .00$), Self-Report Flexible Performance (r (430) = .18, $p = .00$), but not with creative awards.

Table 25
FI: Correlations With Criterion Measures

	FI Accuracy	College GPA	Self-Report Flex Performance	Creative Award
FI Accuracy	1.00			
College GPA	0.27**	1.00		
Self-Report Flexible Performance	0.18**	0.13**	1.00	
Creative Award	0.03	0.09	0.19**	1.00

Note. *N* per cell varies between 382 and 430; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

In sum, the FI showed small to moderate correlations with tests of divergent and convergent fluid intelligence, moderate correlations with tests of pattern recognition, and small correlation with the criterion measures of college GPA and self-report flexible performance.

Flexible Mapping (FM). This analogy test of mental flexibility contains 21 items: 7-item triplets in figural, verbal, and numerical content domains (63 item parts). Each item triplet is made up of one part that requires a domain-homogeneous classification to identify the correct match and two parts that require domain-heterogeneous classification. FM accuracy and latency scores measure mental flexibility at the level of performance components. Internal test analyses are reported in the results of Investigation 1. With regard to external validation, small to moderate correlations with tests of divergent and convergent abilities and criterion measures are expected. Given the test is based on a componential level of analysis, only small correlations with tests of personality are expected.

To explore construct validity, FM accuracy and latency scores were correlated with scores on cognitive ability (Table 26) and personality measures (Table 27). As can be seen in Table 27, FM accuracy scores correlated positively with BIS numerical subtest scores (ZG) (r (150) = .30, p = .00), BIS verbal subtest scores (AM) (r (150) =.22, p =.02), BIS processing capacity (BG) scores (r (149) = .22, p = .01), and French Kit induction factors test scores (Letter: r (150) = .44, p =.00; Location: r (148) = .34, p = .00), as expected. However, contrary to expectations, FM accuracy scores did not correlate with BIS figural subtest scores and FM latency scores did not correlate with scores on cognitive ability tests.

Table 26
FM: Correlations With Cognitive Ability Measures

	FM Accuracy	FM Latency	BIS zg Num	BIS am Verbal	BIS zf Figural	BIS bg Process	KIT I Letter	KIT I Location
FM Accuracy	1.00							
FM Latency	.54**	1.00						
BIS zg Num	.30**	.08	1.00					
BIS am Verbal	.22**	.08	.05	1.00				
BIS zf Figural	.11	.05	.16*	.22**	1.00			
BIS bg Processing	.22**	-.06	.15	.08	.12	1.00		
KIT I Letter	.44**	.15	.29**	.17*	.06	.29**	1.00	
KIT I Location	.34**	-.03	.29**	.07	.08	.34**	.46**	1.00

Note. N per cell varies between 148 and 150; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

The correlation between FM accuracy scores and NEO Openness approached significance (r (114) = .18, p = .06). FM accuracy scores correlated negatively with NEO Extraversion scores (r (113) = -.28, p = .00). FM Latency scores correlated positively with NEO Agreeableness scores (r (123) = .32, p = .00).

Table 27
FM: Correlations With Personality Measures

	FM Acc	FM Lat	CFS	Open	Consc	Extra	Agree	Neuro
FM Accuracy	1.00							
FM Latency	.54**	1.00						
CFS	-.06	.04	1.00					
Openness	.18	.02	.20	1.00				
Conscientious	-.02	.13	.23	-.02	1.00			
Extraversion	-.28**	-.10	.50**	-.18	.47**	1.00		
Agreeableness	.11	.32**	.07	.24	.31	-.03	1.00	
Neuroticism	-.08	-.15	-.15	-.06	-.29	-.16	-.47**	1.00

Note. N per cell varies between 114 and 150; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed). Participants were randomly assigned 2 of the 5 NEO personality subtests, resulting in a low *n* per cell.

To examine the cognitive processes that contribute to performance on the FM, FM accuracy scores were regressed on BIS creativity component scores, BIS processing scores, and KIT induction factor scores entered together. The regression was significant and the predictors explained 25% of total variance (*F change* (3, 143) = 15.94, $p = .00$). Significant predictors in the model were KIT induction factor scores ($\beta = 0.7$, $t = 5.09$, $p = .00$) and BIS creative component ($\beta = 0.05$, $t = 2.85$, $p = .00$). BIS processing was not a significant contributor to the regression.

To explore criterion-related validity, FM accuracy scores were correlated with college GPA, self-report flexible performance, and creative award. As can be seen in Table 28, FM accuracy scores correlated positively with college GPA (r (382) = .28, p =.00), self-report flexible performance (r (430) = .18, p = .00), but not with creative awards.

Table 28
FM: Correlations With Criterion Measures

	FM Accuracy	College GPA	Self-Report Flex Perf	Creative Award
FM Accuracy	1.00			
College GPA	.28**	1.00		
SR Flex Performance	.18**	.13*	1.00	
Creative Award	.05	.09	.19**	1.00

Note. N per cell varies between 382 and 430; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

In sum, the FM showed a similar pattern of association with external measures as the FI. There were small to moderate correlations found with tests of divergent and convergent fluid intelligence and moderate correlations with tests of pattern recognition. Small correlations were found with criterion measures of college GPA and self-report flexible performance.

Counterfactual Analogies. This test is made up of two versions that differ in domain (figural and verbal). Because the verbal version of the test has an additional item type (relevant-irrelevant) the tests will be analyzed and validated separately.

Figural (CFAF). This analogy test of mental flexibility is made up of figural items preceded by a premise that is either familiar or counterfactual (novel) in a random order of presentation. It is predicted that the capacity to shift from familiar to novel premises requires mental flexibility, which is measured by accuracy and latency scores.

Internal Test Analyses. It is expected that items with novel premises will be more difficult to solve correctly and require more time to process. In addition, it is expected that the test is made up of two latent dimensions that reflect the difference between familiar and novel processing demands.

Table 29 presents the results of classical item analyses of the CFAF, which includes item difficulty estimates and discrimination indices. Item discrimination estimates are computed by examining the relative test performance between examinees whose total score fell in the upper 27% of the examinee group and those whose total scores fell in the lower 27% of the examinee group (Crocker & Algina, 1986).

Table 29

CFAF: Difficulty Estimates and Discrimination Indices

Item	Item Difficulty	Discrimination Index
CFAF01*	.32	.21
CFAF02	.76	.38
CFAF03	.70	.40
CFAF04*	.25	.21
CFAF05*	*.07*	*-.15*
CFAF06	.80	.38
CFAF07	.38	.14
CFAF08*	.55	.51
CFAF09	.67	.54
CFAF10	.69	.55
CFAF11	.55	.51
CFAF12*	.38	.36
CFAF13*	.22	.26
CFAF14	.61	.61
CFAF15	.61	.33
CFAF16*	.24	.08
CFAF17	.74	.46
CFAF18*	*.12*	.13
CFAF19*	.30	.43
CFAF20	.61	.48
CFAF21*	.20	.09
CFAF22	.46	.72
CFAF23*	.37	.38
CFAF24*	.76	.42
CFAF25*	.38	.23
CFAF26	.55	.45
CFAF27	.30	.26
CFAF28*	.45	.30
CFAF29*	.43	.27
CFAF30	.19	.39

Note. *Indicates item with a novel premise. Italicized values fall below the level expected for guessing.

As can be seen in Table 29, item difficulty for all but a few items with novel premises (see items with an asterisk in Table 31) fell below $p = .5$. Difficulties for items 5 and 18 are well below p-values that would be expected from guessing among 4 response options, which would be $1/m$ ($p = .25$), with m being the number of response options. Difficulties for most of the items with familiar premises were considerably higher, with 11 of the 15 items with familiar premises falling above p = .5. Models for calculating ideal difficulty values for optimal score distribution and maximizing total score reliability that adjust for guessing suggest difficulty values of .75 (.5 + 1/m) (Suen.1990), .74 (Lord, 1952), and .62 (.50 + .50/m) (Crocker & Algina, 1986).

Discrimination indices were low to moderate for all items. The index for item 5 was negative, suggesting it does not discriminate between high and low total scores. Accordingly, Item 5 was removed from the scale in subsequent analyses due to the low difficulty (p = .07) and negative discrimination index (-.15). In addition, Item 18 was removed from the scale in subsequent analyses because both difficulty and discrimination values were very low. Both items contained novel premises.

A summary of full-scale and subscale (familiar and novel premise) difficulty and discrimination estimates and Cronbach's alpha internal consistency estimates is presented in Table 30. As can be seen in the table, the internal consistency estimate for the novel subscale was low (α = .45). When two items were removed from the novel subscale according to poor difficulty and discrimination estimates noted above, the internal consistency estimate did not improve (α = .43). Low reliabilities on the novel subscale may suggest that multiple strategies may be used when processing analogies with novel versus familiar premises. Alternatively, it may be the results of error in measurement consistent with low difficulty and discrimination estimates.

Table 30
CFAF: Summary of Difficulty, Discrimination, and Internal Consistency Estimates

Scale	*N*	Median difficulty	Median discrimination	Alpha
CFAF Novel	15	.31 (.07 to .76)	.27 (-.15 to .42)	.45
CFAF Familiar	15	.61 (.19 to .80)	.40 (.14 to .72)	.70
Full scale	30	.38 (.07 to .80)	.38 (-.15 to .72)	.72

Results of a principal-components analysis with two-factor extraction and Varimax rotation on the revised 27-item CFAF scale suggest two components that account for 19.73% of the variance and roughly conform to the conceptual structure of the test. As shown in Table 31, a majority of items (12 out of 15) with familiar premises loaded on the first factor, which accounted for 12.7 % of the variance. Five (5) out of 12 items with novel premises loaded on the second factor, which accounted for 7.2% of the variance. These results suggest that items with novel premises may be more dimensionally complex than items with familiar premises. It also could mean that some items do not belong to a common dimension.

Table 31

CFAF: Principal-Component Factor Analysis With Two Factors Imposed and Varimax Rotation Presented by Item Type

Familiar Items	FACTOR 1	FACTOR 2
CFAF 2	**.40**	.08
CFAF 3	**.40**	.04
CFAF 6	**.52**	-.16
CFAF 7	.04	.07
CFAF 9	**.47**	.12
CFAF 10	**.60**	-.05
CFAF 11	**.44**	.14
CFAF 14	**.54**	.09
CFAF 15	**.32**	-.01
CFAF 17	**.51**	.02
CFAF 20	.36	.23
CFAF 22	**.60**	.19
CFAF 26	**.36**	.16
CFAF 27	.20	.09
CFAF 30	.35	.47
Novel Items		
CFAF 1	.10	**.40**
CFAF 4	-.07	**.67**
CFAF 8	**.36***	.14
CFAF 12	.16	**.47**
CFAF 13	.12	**.60**
CFAF 16	-.15	**.43**
CFAF 19	.39	.21
CFAF 21	-.09	.26
CFAF 23	.23	.22
CFAF 24	**.44***	.00
CFAF 25	.11	.16
CFAF 28	.28	-.07
CFAF 29	.20	-.10

*Loadings that are not consistent with conceptualized item structure.

Note. Factor loadings are in bold if they are above .30 and not double-loaded by more than one half (Stevens, 1996). Both items 5 and 18 were omitted.

The mean total score for the familiar premise subscale (mean (452) = 8.61; sd = 3.06) was significantly greater than was the mean total score on the novel premise subscale (mean (452) = 4.84; sd = 2.15), as predicted. A two-tailed, paired-sample test of mean differences revealed a significant difference, t (1,451) = 28.312, p = .000, suggesting the items with novel premises are more difficult to solve correctly.

Examination of reaction time data also confirms the expectation that items with novel premises may be more cognitively demanding. The mean latency for novel premise subscale items (mean = 12.61; sd = 6.29) was significantly greater than was the mean latency for familiar premise items (mean = 11.45; sd = 5.11), as shown by a two-tailed, paired-sample test of mean differences, $t\ (1{,}451) = -8.854$, $p = .000$.

In sum, the CFAF is made up of two dimensions. The majority (11 of 15 items), with familiar premises, fell in one dimension. However, fewer than half (4 out of 13) of the items with familiar premises fell in the second dimension. Items in the novel dimension may be more difficult to process as evidenced by mean differences in accuracy scores and greater latency scores.

<u>External Validation Analyses</u>. To explore construct validity, CFAF novel and familiar subscales were combined to capture respondents' ability to shift between novel and familiar question types in a single test session. Accordingly, CFAF accuracy scores were correlated with scores on cognitive ability (Table 32) and personality measures (Table 33). With regard to cognitive ability, it was expected that there would be small to moderate correlations between the CFAF and BIS creative component subtests (Figural–ZF Verbal–AM, and Numerical–ZG) and processing capacity (BG) scores. Similarly, small to moderate positive correlations were expected between the CFAF and French Kit induction factor scores (Letter and Location). With regard to personality measures, small correlations were expected between CFAF and NEO Openness and CFS scores.

As expected, small correlations were found between CFAF accuracy scores and BIS numerical subtest scores (ZG) ($r\ (146) = .19$, $p = .02$), and the BIS verbal subtest scores (AM) ($r\ (146) = .20$, $p = .01$), but not BIS figural subtest scores (ZF) ($r\ (146) = .08$, *ns*). In addition, a small correlation was found between CFAF and BIS processing capacity (BG) scores ($r\ (145) = .22$, $p = .01$). Consistent with expectations, moderate, positive correlations were found between CFAF accuracy scores and French Kit letter sets scores ($r\ (146) = .41$, $p = .00$) and French Kit location scores ($r\ (151) = .32$, $p = .00$).

Table 32
CFAF: Correlations With Cognitive Ability Correlates

	CFAF Accuracy	CFAF Latency	BIS zg	BIS am	BIS zf	BIS bg	KIT I Letter
CFAF Accuracy							
CFAF Latency	.71**						
BIS zg (Numerical)	.19*	.04					
BIS am (Verbal)	.20*	.13	.05				
BIS zf (Figural)	.08	-.02	.16*	.22			
BIS bg (Process)	.22**	.07	.15	.08	.12		
KIT Letter	.41**	.25**	.22*	.11	-.07	.24**	
KIT Location	.33**	.13	.12	.32*	.03	.26	.07

Note. N per cell ranged from 144 to 146; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

Small correlations with personality measures were also found. CFAF accuracy scores were negatively correlated with NEO Extraversion scores (r (109) = -.25, p = .01) and NEO Neuroticism scores (r (115) = -.22, p = .02). CFAF latency scores were positively correlated with NEO Openness scores (r (106) = .23, p = .02), consistent with expectations. However, CFAF accuracy scores did not correlate with NEO Openness or CFS scores.

Table 33
CFAF: Correlations With Personality Measures

	CFAF Accuracy	CFAF Latency	CFS	Open	Consc	Extra	Agree	Neuro
CFAF Accuracy	1.00							
CFAF Latency	.71**	1.00						
CFS	-.03	.03	1.00					
Openness	.13	.23*	.20	1.00				
Conscientious	.03	.06	.23	-.02	1.00			
Extraversion	-.25*	-.15	.50**	-.18	.47**	1.00		
Agreeableness	.10	.09	.07	.24	.31	-.03	1.00	
Neuroticism	-.22*	-.08	-.15	-.06	-.29	-.16	-.47**	1.00

Note. N per cell varies from 109 to 145. Participants were randomly assigned 2 of the 5 NEO personality subtests, resulting in a lower *n* per cell; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

To examine the cognitive processes that contribute to performance on the CFAF, CFAF scores were regressed on BIS creativity component scores, BIS processing scores, and KIT induction factor scores entered together. The regression was significant and the predictors explained 21% of total variance in CFAF scores (*F change* (3, 139) = 12.59, p = .00). Significant predictors in the model were KIT induction factor scores ($\beta = 1.99$, $t = 4.754$, $p = .00$) and BIS creative component ($\beta = 1.03$, $t = 1.946$, $p = .05$). BIS processing was not a significant contributor to the regression.

To explore criterion-related validity, CFAF accuracy scores were correlated with college GPA, self-report flexible performance, and creative award. As can be seen in Table 34, CFAF accuracy scores correlated positively with college GPA (r (397) = .15, p = .00) and self-report flexible performance (r (445) = .14, p = 0.00), but not with creative awards (r (445) = -0.02, ns).

Table 34
CFAF: Correlations With Criterion Measures

	CFAF Accuracy	College GPA	SR Flex Performance	Creative Award
CFAF Accuracy	1.00			
College GPA	.15**	1.00		
SR Flex Performance	.14*	.13**	1.00	
Creative Award	-.02	.10*	.19**	1.00

Note. N per cell ranged from 397 to 445; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

In sum, the CFAF showed small to moderate correlations with measures of divergent and convergent fluid intelligence. Correlations with criterion measures of college GPA and self-report flexible performance were small, ranging from .13 to .15.

Verbal (CFAV). This analogy test of mental flexibility is made up of 48 verbal items preceded by a premise that is either familiar or counterfactual (novel) and relevant or irrelevant to finding a solution. These four item types (Novel–Relevant, Novel–Irrelevant, Familiar–Relevant, and Familiar–Irrelevant) are presented to the test taker in random order.

Internal Test Analyses. It is predicted that the capacity to shift from familiar to novel premises is an aspect of mental flexibility that can be measured by CFAV accuracy and latency scores. It is also expected that the capacity to assess the relevance of information presented is a processing requirement for mental flexibility. Accordingly, items with novel-relevant premises are expected to be more difficult to solve correctly and require more time to process than are items with familiar premises. In addition, items with novel–irrelevant premises are expected to be processed differently than items with novel–relevant premises, of which the latter should be more difficult to solve correctly and require more processing time. Finally, it is expected that the test is made up of two latent dimensions that reflect the difference between familiar and novel processing demands.

Table 35 presents the results of classical item analyses of the CFAV. Item difficulties are computed in the same manner as CFAF analyses detailed above.

Table 35

CFAV: Difficulty Estimates and Discrimination Indices Presented by Item Type

Familiar-Relevant	Difficulty	Discrimination
CFAV 2	.55	.41
CFAV 7	.76	.17
CFAV 8	.83	.26
CFAV 10	.96	.10
CFAV 11	.92	.16
CFAV 24	.93	.20
CFAV 29	.82	.27
CFAV 31	.75	.34
CFAV 40	.75	.21
CFAV 41	.40	.60
CFAV 42	.95	.13
CFAV 45	.72	.35
Familiar-Irrelevant		
CFAV 1	.83	.26
CFAV 4	.95	.12
CFAV 5	.61	.18
CFAV 15	.83	.32
CFAV 19	.83	.26
CFAV 21	.73	.34
CFAV 23	.83	.17
CFAV 27	.82	.39
CFAV 30	.83	.29
CFAV 32	.83	.31
CFAV 36	.47	.19
CFAV 43	.45	.28
Novel-Relevant		
CFAV 3	.28	.35
CFAV 9	.42	.37
CFAV 12	.44	.43
CFAV 17	.50	.33
CFAV 18	.30	.37
CFAV 20	.56	.30
CFAV 22	.36	.54
CFAV 26	.56	.43
CFAV 37	.70	.26
CFAV 39	.50	.50
CFAV 46	.34	.23
CFAV 47	.44	.43

Novel-Irrelevant		
CFAV 6	.67	.13
CFAV 13	.60	.50
CFAV 14	.55	.17
CFAV 16	.28	.12
CFAV 25	.82	.28
CFAV 28	.66	-.09*
CFAV 33	.21	.34
CFAV 34	.82	.33
CFAV 35	.93	.12
CFAV 38	.78	.36
CFAV 44	.63	.45
CFAV 48	.60	-.07*

Note. *Negative discrimination indices.

As can be seen in Table 35, item difficulties all fall above guessing (p = .25) across subscales. Discrimination indices were low to moderate across subscales. Items 28 and 48 had negative discrimination indices, suggesting they do not discriminate between high and low total scores. Accordingly items 28 and 48 were removed from the scale in subsequent analyses.

A summary of full-scale and subscale (familiar and novel premise) difficulty and discrimination estimates and Cronbach's alpha internal consistency estimates is presented in Table 36. Consistent with findings for CFAF, the internal consistency estimate for the novel subscale was lower ($\alpha = .63$) than was the estimate for the familiar subscale ($\alpha = .73$). When items 28 and 48 were removed, the internal consistency estimate of the novel irrelevant subscale went down to $\alpha = .50$

Table 36
CFAV: Summary of Difficulty, Discrimination, and Internal Consistency Estimates

Scale	*N*	Median difficulty	Median discrimination	Alpha
CFAV Novel	24	.55 (.20 to .93)	.34 (-.09 to .54)	.63
CFAV Novel Relevant	12	.44 (.28 to .70)	.37 (.23 to .54)	.84
CFAV Novel Irrelevant	12	.64 (.21 to .93)	.22 (-.09 to .49)	.54
CFAF Familiar	24	.82 (.40 to .96)	.26 (.10 to .60)	.73
CFAF Familiar Relevant	12	.82 (.40 to .96)	.22 (.10 to .60)	.59
CFAV Familiar Irrelevant	12	.82 (.45. to .95)	.27 (.12 to .39)	.55
Full scale	48	.64 (.20 to .96)	.28 (1.09 to .60)	.76

Results of a principal-components analysis with two-factor extraction and Varimax rotation on the revised 46-item CFAV scale suggest two components that account for 22.14% of the variance and roughly conform to the conceptual structure of the test. As shown in Table 37, a majority of items (10 out of 22) with *novel–relevant* premises, loaded on the first factor, which accounted for 11.38 % of the variance. Sixteen out of 24 items with familiar premises (relevant and irrelevant) and 5 with novel premises (4 irrelevant, 1 relevant) loaded on the second factor, which accounted for 10.75% of the variance. This pattern of results suggests that items with *novel–relevant* premises are dimensionally distinct from items with *novel–irrelevant* and familiar premises (relevant and irrelevant).

Table 37
CFAV: Principal-Component Factor Analysis, Two-Factors, Varimax Rotation

Rotated	Factor 1	Factor 2
Familiar		
CFAV1*	-.03	**.33**
CFAV2	.20	.32
CFAV4*	.09	**.39**
CFAV5*	-.08	.21
CFAV7	-.20	.35
CFAV8	.23	.27
CFAV10	.12	**.36**
CFAV11	.13	**.38**
CFAV15*	.02	**.52**
CFAV19*	-.02	**.41**
CFAV21*	-.02	**.43**
CFAV23*	.18	.33
CFAV24	-.01	**.59**
CFAV27*	.23	**.46**
CFAV29	.05	**.39**
CFAV30*	-.08	**.43**
CFAV31	.20	.33
CFAV32*	-.06	**.55**
CFAV36*	.14	.06
CFAV40	-.14	**.32**
CFAV41	.32	.38
CFAV42	.10	**.38**
CFAV43*	-.12	.26
CFAV45	.03	**.40**

Novel		
CFAV3	**.42**	.09
CFAV6	-.13	.15
CFAV9	**.40**	.08
CFAV12	**.74**	-.08
CFAV13*	.07	***.47***
CFAV14*	-.34	.42
CFAV16*	-.14	.15
CFAV17	**.55**	-.06
CFAV18	**.71**	-.06
CFAV20	**.61**	-.11
CFAV22	**.60**	.12
CFAV25*	-.06	***.50***
CFAV26	**.78**	-.13
CFAV33*	.25	.18
CFAV34*	.05	***.39***
CFAV35*	-.13	**.38**
CFAV37	.22	.14
CFAV38*	.12	***.34***
CFAV39	**.80**	-.04
CFAV44*	.09	***.37***
CFAV46	**.51**	-.16
CFAV47	**.71**	-.01

Note. *Items with irrelevant premises; Factor loadings are in bold if they are above .30 and not double-loaded by more than half (Stevens, 1996). Italicized loading is not consistent with conceptualized item structure.

As expected, the mean total score on items with *familiar–relevant* premises was significantly greater than was the mean total score for items with *novel–relevant* premises. A two-tailed, paired-sample test of mean differences revealed a significant difference, t (1,465) = -22.50, p = .000, suggesting the items with relevant novel premises are more difficult to solve correctly. In addition, the correlation between novel– relevant scores and familiar–relevant scores was low (r =. 14, p = .00), suggesting different modes of processing. A two-tailed, paired-sample test of mean differences between *familiar–irrelevant* and *novel–irrelevant* items showed these items types are also significantly different, t (1, 465) = -22.50, p = .000, suggesting again that novel types are more difficult to solve than are familiar types. However, unlike the comparison between *familiar–relevant* and *novel–relevant* item scores, these scores were moderately correlated (r = .58, p = .00), suggesting a similar mode of cognitive processing.

To further explore the differences between relevant and irrelevant premised items, a two-tailed, paired-sample test of mean differences was conducted between mean total scores on *novel–relevant* and *novel–irrelevant* items and between mean total scores on *familiar–relevant* and *familiar–irrelevant* items. Items with *novel–irrelevant* premises were easier to solve than were items with *novel–relevant* premises (t (1, 465) = -4.75, p = .000). In contrast, *familiar–relevant* premises were easier to solve than items with *familiar–irrelevant* premises (t (1, 465) = 3.933, p = .000). Moreover, scores on novel–relevant items were uncorrelated with scores on novel–irrelevant items (r (465) = -.06, ns), whereas scores on familiar–relevant items were moderately correlated with scores on familiar–irrelevant items (r = .56, p = .000).

Examination of reaction time data also confirms the expectation that novel–relevant items may be more cognitively demanding than are familiar–relevant items. The mean latencies on novel–relevant items were significantly greater than the mean latencies on familiar–relevant items, as shown by a two-tailed, paired-sample test of mean differences, t (1,465) = 10.491, p = 0.000). In contrast, the mean latencies on novel–irrelevant items, as compared to familiar–irrelevant items, were not, (t (1,465) = -1.51, ns).

In sum, the CFAV is made up of two dimensions: One dimension comprises times with familiar–relevant, familiar–irrelevant, and novel–irrelevant premises, and the other comprises items with novel–relevant premises. Items in the novel–relevant dimension are more difficult to process as evidenced by mean differences in accuracy scores and greater latency scores.

External Validation Analyses. To explore construct validity, CFAV accuracy scores were correlated with scores on cognitive ability (Table 38) and personality measures (Table 39). With regard to cognitive ability, it was expected that there would be small to moderate correlations between the CFAV and BIS creative component subtests (Figural–ZF, Verbal–AM, and Numerical–ZG) and processing capacity (BG) scores. Similarly, small to moderate positive correlations were expected between the CFAV and French Kit induction factor scores (Letter and Location). With regard to personality measures, small correlations were expected between CFAV and NEO Openness and CFS scores.

As can be seen in Table 38, CFAV accuracy scores correlated with BIS numerical subtest scores (ZG) (r (151) = .25, p = .00), BIS verbal subtest scores (AM) (r (151) = .21, p = .01), and BIS processing capacity (BG) scores (r (150) = .23, p = .00), consistent with predictions. However, CFAV accuracy scores did not correlate as expected with BIS (BF) figural subtest scores. In addition, CFAV scores positively correlated with KIT induction factor scores (Letter: r (151) = .32, p = .00; Location: r (149) = .20, p = .01).

Table 38
CFAV: Correlations With Cognitive Ability Measures

	CFAV Accuracy	CFAV Latency	BIS zg	BIS am	BIS zf	BIS bg	KIT Letter	KIT Location
CFAV Accuracy	1.00							
CFAV Latency	.32**	1.00						
BIS zg (Numerical)	.25**	-.01	1.00					
BIS am (Verbal)	.21**	.09	.05	1.00				
BIS zf (Figural)	.05	-.06	.16*	.22**	1.00			
BIS bg (Process)	.23**	-.05	.15	.08	.12	1.00		
KIT Letter	.32**	-.02	.22**	.11	-.07	.24**	1.00	
KIT Location	.20*	.01	-.13	.09	.08	-.02	-.10	1.00

Note. N per cell varies between 149 and 151; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

As can be seen in Table 39, CFAV scores correlated with NEO Openness scores (accuracy: r (111) = .20p = .04; latency: r (111) = .24, p = .01), as expected. CFAV accuracy scores also correlated positively with NEO Agreeableness scores (r (120) = .18, p = .05), negatively with NEO Extraversion scores (r (110) = -.30, p = .00), and negatively with NEO Neuroticism scores (r (118) = -.26, p = .00).

Table 39
CFAV: Correlations With Personality Measures

	CFAV Accuracy	CFAV Latency	CFS	Open	Consc	Extra	Agree	Neuro
CFAV Accuracy	1.00							
CFAV Latency	.32**	1.00						
CFS	-.04	.08	1.00					
Openness	.20*	.24**	.20	1.00				
Conscientious	-.05	.08	.23	-.02	1.00			
Extraversion	-.30**	.11	.50**	-.18	.47*	1.00		
Agreeableness	.18*	.11	.07	.24	.31	-.03	1.00	
Neuroticism	-.26**	-.05	-.15	-.06	-.29	-.16	-.47**	1.00

Note. N per cell varies between 105 and 151. Participants were randomly assigned 2 of the 5 NEO personality subtests, resulting in a low *n* per cell; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

To examine the cognitive processes that contribute to performance on the CFAV, CFAV scores were regressed on BIS creativity component scores, and BIS processing scores and KIT induction factor scores entered together. The regression was significant and the predictors explained 17% of total variance in CFAV scores (*F change* (3, 144) = 9.76, p =.00). Significant predictors in the model were KIT induction factor scores (β = 1.705, t = 3.121, p = .00) and BIS creative component (β = 1.895, t = 2.727, p = .00). BIS processing was not a significant contributor to the regression.

To explore criterion-related validity, CFAV accuracy scores were correlated with college GPA, self-reported flexible performance and creative award. As can be seen in Table 40, CFAF accuracy scores correlated positively with college GPA (r (407) = .22, p = .00) and self-reported flexible performance (r (459) = .18, p = .00), but not with creative awards.

Table 40
CFAV: Correlations With Criterion Measures

	CFAV Accuracy	College GPA	SR Flex Performance	Creative Award
CFAV Accuracy	1.00			
College GPA	.22**	1.00		
SR Flexible Performance	.18**	.13*	1.00	
Creative Award	.07	.10*	.19**	1.00

Note. *N* per cell varies between 406 and 458; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

In sum, the CFAV showed small to moderate correlations with measures of divergent and convergent fluid intelligence. Correlations with criterion measures of college GPA and self-reported flexible performance were small.

Insight. This test of solving novel problems contains 9 open-response verbal, figural, and numerical insight problems dichotomously scored for accuracy as correct or incorrect. Insight problems require the capacity to restructure elements of a problem in novel ways. As such, the Insight test is expected to measure a type of mental flexibility.

Internal Test Analyses. Results of classical item analyses of the Insight test are presented in Table 41. Item difficulties and discrimination indices are computed as detailed above.

Table 41
Insight Test: Estimates of Item Difficulty and Discrimination

	Difficulty	Discrimination
insight_1	.17	.32
insight_2	*.09*	.25
insight_3	.13	.34
insight_4	.55	.76
insight_5	*.03*	*.12*
insight_6	.33	.65
insight_7	.67	.68
insight_8	.35	.59
insight_9	.38	.80

As can be seen in Table 41, difficulties range from .03 to .67 and discrimination indices range from .12 to .80. Cronbach's alpha internal consistency estimate of the Insight test is $\alpha = .64$ (N of items = 9) and the inter-item correlation is $r = .18$, which suggests adequate reliability (Streiner, 2003). A principal-components factor analysis with Varimax rotation produced three factors differentiated by item domain (verbal, figural, and numerical) that together explain 49.61% of total variance. Table 42 displays these results.

Table 42
Insight Test: Principal-Components Factor Analysis With Varimax Rotation

Item	Factor 1	Factor 2	Factor 3
insight_1	-.093	.099	**.761**
insight_2	.118	**.646**	.224
insight_3	.320	.221	.392
insight_4	**.688**	.150	.010
insight_5	-.018	**.716**	.196
insight_6	.246	.029	**.625**
insight_7	**.718**	-.002	.064
insight_8	.242	**.693**	-.155
insight_9	.571	.211	.358

Note. Factors in bold are above .30 and not double-loaded by more than one half (Stevens, 1996).

As can be seen in the table, the first factor, which explained 17.3% of shared variance, loaded on two numerical items, the second, which explained another 17.1%, loaded on three figural items, and the third factor, which explained 15.2%, loaded on two verbal items.

In sum, a number of the items on the Insight test were rather high in difficulty. The factorial structure of the test seems to be reflective of domain.

<u>External Validation Analyses</u>. To explore construct validity, Insight scores were correlated with scores on cognitive ability (Table 43) and personality measures (Table 44). As can be seen in Table 43, Insight test scores correlated with BIS numerical subtest scores (ZG) (r (153) = .23, p = .00), BIS verbal subtest scores (AM) (r (153) = .27, p = .00), BIS processing capacity (BG) scores (r (153) = .28, p = .00), and French Kit induction factors test scores (Letter: r (153) = .35, p = .00; Location: r (151) = .35, p = .00). However, contrary to expectations, Insight test scores did not correlate with BIS figural subtest scores.

Table 43
Insight Test: Correlations With Cognitive Ability Measures

	Insight	BIS zg	BIS am	BIS zf	BIS bg	KIT Letter	KIT Location
Insight	1.00						
BIS zg (Numerical)	.23**	1.00					
BIS am (Verbal)	.27**	.05	1.00				
BIS zf (Figural)	.13	.16*	.22**	1.00			
BIS bg (Processing)	.28*	.15	.08	.12	1.00		
KIT Letter	.35**	.29**	.17*	.06	.29**	1.00	
KIT Location	.35**	.29**	.07	.08	.34**	.46**	1.00

Note. N per cell ranges between 151 and 153; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

As can be seen in Table 44, the correlation between Insight test scores and NEO Openness scores approached significance (r (11) = .17, p = .07). Insight scores were negatively correlated with NEO Conscientiousness (r (123) = - .22, p = .01) and NEO Extraversion scores (r (111) = - .21, p = .03).

Table 44
Insight Test: Correlations With Personality Measures

	Insight	Open	Consc	Extra	Agree	Neuro
Insight	1.00					
Openness	.17	1.00				
Conscientiousness	-.22*	-.02	1.00			
Extraversion	-.21*	-.18	.47**	1.00		
Agreeableness	.10	.24	.31	-.03	1.00	
Neuroticism	-.14	-.06	-.29	-.16	-.47**	1.00

Note. N per cell ranges from 105 to 145. Participants were randomly assigned 2 of the 5 NEO personality subtests, resulting in a low *n* per cell; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

To examine the cognitive processes that contribute to performance on the Insight test, Insight scores were regressed on BIS creativity component scores, BIS processing scores, and KIT induction factor scores entered together. The regression was significant and the predictors explained 22% of total variance (F *change* (3, 147) = 13.81, p = .00). Significant predictors in the model were KIT induction factor scores (β = .754, t = 3.91, p = .00) and BIS creative component (β = .689, t = 2.81, p = .01). BIS processing was not a significant contributor to the regression.

To explore criterion-related validity, Insight scores were correlated with college GPA, SR flexible performance and creative award. As can be seen in Table 45, Insight scores correlated positively with college GPA (r (409) = .20, p = .00), SR Flexible Performance (r (458) = .23, p = .00), and creative awards (r (459) = .09, p = .05).

Table 45
Insight Test: Correlation With Criterion Measures

	Insight	College GPA	SR Flexible Performance	Creative Award
Insight	1.00			
College GPA	.20**	1.00		
SR Flexible Performance	.23**	.19*	1.00	
Creative Award	.09*	.10*	.19**	1.00

Note. *N* per cell is 458; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

In sum, the Insight test showed small to moderate correlations with measures of divergent and convergent fluid intelligence, and moderate correlations with tests of pattern recognition. Correlations with criterion measures of college GPA and self-report flexible performance were small.

Test Battery Analyses. The mental flexibility test battery is composed of classification and analogy (FI and FM) tests designed to assess mental flexibility according to the componential subtheory, and two tests—one analogy (CFA—Figural and Verbal) and one problem solving (Insight)—designed to assess mental flexibility according to the experiential subtheory. It is expected that one mental flexibility factor should explain variance at the test-battery level. This factor should be structurally distinct from the latent structure underlying tests of cognitive ability and pattern recognition. It also is expected that the test battery will explain variance in criterion-related measures of mental flexibility over and above variance explained by measures of divergent and convergent cognitive ability, pattern recognition, and personality. Finally, it is expected that a structural equation model that specifies two first-order latent factors that correspond to the componential and experiential subtheories and one second-order latent mental flexibility factor will fit the mental flexibility test battery data.[1]

Convergent and Discriminant Validity. Correlations among mental flexibility tests are all positive and ranged from .44 to .98, as displayed in Table 46. The table of correlations suggests reasonable convergent and discriminant validity among the Insight, CFAF, and CFAV tests. However, FI and FM tests were highly correlated, which suggests poor discriminant validity.

[1] It was not possible to test the underlying two-factor sub-theory model (Componential and Experiential) of the test battery with a Confirmatory Factor Analysis (CFA). Results were inconclusive because the high correlation between scores on the FI and FM tests resulted in empirical under-identification of the model.

Table 46
Correlations Among Mental Flexibility Tests

	CFAF	CFAV	Insight	FM	FI
CFAF	1.00				
CFAV	.49**	1.00			
Insight	.44**	.54**	1.00		
FM	.61**	.68**	.55**	1.00	
FI	.60**	.67**	.54**	.98**	1.00

Note. N ranged from 419 to 452 per cell; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

To examine the latent structure of the mental flexibility test battery, a principal components factor analysis was conducted using FI, FM, CFAF, and CFAV accuracy scores and Insight test scores. Results of the analysis were one latent factor that explained 70% of total variance. Component loadings are displayed in Table 47. A follow-up principal axis factor analysis confirmed a one-factor solution.

Table 47
Results of PCA of Mental Flexibility Test Battery

	Principal Component
Insight	.529
CFAF Accuracy	.562
CFAV Accuracy	.671
FM Accuracy	.878
FI Accuracy	.859

To assess the discriminant validity of the mental flexibility test battery from other tests of convergent and divergent cognitive ability, a second principal-components factor analysis with Varimax rotation was conducted, in which the BIS creativity subtest (ZG, AM, and ZF), BIS processing (BG), and KIT induction subtest (Letter, Location) were included. Results of the analysis were three latent factors that explained 62% of total variance. All mental flexibility tests loaded on the first component, which explained 31.45% of total variance. Primarily cognitive processing tests (BG, ZG, Letter, Location) loaded on the second component, which explained 18.63% of total variance. Two BIS creative component tests (AM, ZF) loaded on the third component, which explained 11.82% of total variance. Component loadings are displayed in Table 48. Double loadings above .30 are italicized in the table. It should be noted that the Insight test double-loaded on the second component and the KIT induction letter double-loaded on the first component, suggesting shared processing components in these tests. A follow-up principal axis factor analysis confirmed a three-factor solution. These results provide preliminary evidence of discriminant validity of the mental flexibility battery.

Table 48
Results of PCA With Varimax Rotation on Mental Flexibility and Cognitive Ability Tests

	Component 1	Component 2	Component 3
Insight	**.563**	*.312*	.264
CFAF Accuracy	**.746**	.235	.062
CFAV Accuracy	**.795**	.130	.097
FM Accuracy	**.928**	.197	.060
FI Accuracy	**.917**	.201	.041
BIS zg Numerical	.190	**.505**	.062
BIS am Verbal	.241	-.028	**.756**
BIS zf Figural	-.025	.164	**.776**
BIS bg Processing	.039	**.719**	.126
KIT I Letter	*.337*	**.625**	.110
KIT I Location	.206	**.781**	-.078

To assess the discriminant validity with respect to pattern recognition, two principal-components analyses with Varimax rotation were conducted. The first analysis included the mental flexibility test battery, GEFT, PFBT, and Minnesota Clerical tests. The second included the mental flexibility test battery, GEFT, and SI tests. The first analysis resulted in a one-component solution that explained 65% of total variance. The second revealed two largely overlapping components that explained 59% of total variance. Discriminant validity with respect to pattern recognition is not established.

Incremental Validity. To assess the incremental validity of the mental flexibility test battery, a series of regressions was conducted to examine the contribution of mental flexibility tests to criterion measure variance above and beyond variance explained by cognitive ability tests, personality measures, pattern recognition measures, and mental flexibility tests entered last.

In the first set of regressions, criterion measures were regressed on BIS creativity standardized aggregate scores, with KIT induction standardized aggregate scores entered first, and all of the mental flexibility tests entered second. In the first regression, with college GPA as the dependent variable, the first model with BIS creativity and KIT induction as predictors was not significant. The second model, with the addition of the mental flexibility test battery, was significant (*F change* (5, 116) = 3.04, $p = .01$) and explained 11.5% of variance. In the second regression, with creative award as the dependent variable, again the first model with BIS creative and KIT induction as predictors was not significant. The second model, with the addition of the mental flexibility test battery, was significant (*F change* (5, 130) = 2.58, $p = .03$) and explained 9% of variance. In the third regression, with SR flexible thinking as the dependent variable, only the first model with BIS creative and KIT induction as predictors was significant (*F change* (2, 135) = 3.14, $p = .05$) and explained 4.4% of variance. In the final regression, with SR flexible behavior, as the dependent variable, neither model was significant.

In the second set of regressions, criterion measures were regressed on GEFT, with PFBT entered first and all of the mental flexibility tests entered second. The regressions with college GPA, creative award, and SR flexible behavior as dependent variables were not significant. In the regression with SR flexible thinking as the dependent variable, only the first model with GEFT and PFBT as predictors was significant (*F change* (2, 11) = 3.38, $p = .4$); it explained 5.7% of variance.

In the third set of regressions, criterion measures were regressed on NEO Openness entered first and all of the mental flexibility tests entered second. In the regression with college GPA as the dependent variable, the first model with NEO Openness as a predictor was significant (*F change* (1, 87) = 3.89, $p = .05$) and explained 4.3% of the variance. The second model with the mental flexibility measures as predictors was also significant (*F change* (5, 82) = 2.50, $p = .04$) and explained 12.7% of the variance. In the regression with creative award as the dependent variable, only the first model with NEO Openness was significant (*F change* (1, 98) = 9.01, $p = .00$); it explained 8.4% of the variance. In the regression with self report flexible performance as the dependent variable, the first model with NEO Openness as a predictor was significant (*F change* (1, 98) = 4.43, $p = .04$) and explained 4.3% of the variance. The second model with mental flexibility measures as predictors was not significant.

Pattern Recognition. In this investigation, pattern recognition is explored as a basic process that may give rise to mental flexibility. Pattern recognition is conceptualized as a dynamic cognitive process of connecting cues to form meaningful configurations (patterns) in a given context (Margolis, 1987). To examine the role of pattern recognition as a contributor to performance on each of the newly developed mental flexibility tests, FI, FM, CFA-Verbal, CFA-Figural, and Insight tests were regressed on GEFT scores and SI sensitivity scores entered together in five separate regression analyses[2]. Table 49 displays the variance explained by the pattern recognition predictors in each of the regressions with mental flexibility test as the dependent variable.

Table 49
Mental Flexibility Test Variance Explained by Pattern Recognition Measures

	Predictors: GEFT & SI		
Dependent Variable	R^2	*F* Change	*P* value
FI	.46	71.61	.00
FM	.47	72.32	.00
CFAF	.40	59.73	.00
CFAV	.39	56.93	.00
Insight	.27	33.14	.00

As can be seen in Table 50, the results of each of the regressions were significant. In the regression with FI scores as the dependent variable, predictors explained 46% of total variance (*F change* (2, 165) = 71.61, p = 0.00). In the regression with FM scores as the dependent variable predictors explained 47% of total variance (*F change* (2, 165) = 72.32, p = .00). In the regression with CFAF scores as the dependent variable the predictors explained 40% of total variance. (*F change* (2, 178) = 59.73, $p = .00$). Similarly, in the results of the regression with the CFAV

[2]It was not possible to include PFBT scores because participants who took the SI test did not take the PFBT.

scores as the dependent variable predictors explained 39% of total variance (*F change* (2, 178) = 56.93, p = .00). Finally, in regression with Insight test scores predictors explained 27% of total variance (*F change* (2, 177) = 33.14, p = .00). GEFT and SI pattern recognition test predictors were both highly significant contributors in each of the regressions.

Table 50

Correlations Among Mental Flexibility and Pattern Recognition Measures

	FM Acc	FM Lat	FI Acc	FI Lat	CFAF Acc	CFAF Lat	CFAV Acc	CFAV Lat	Insight	GEFT	PFBT	MC Name	MC Num	SI Bias	SI Sens
FM Acc	1.00														
FM Lat	.54**	1.00													
FI Acc	–	–	1.00												
FI Lat	–	–	.53**	1.00											
CFAF Acc	–	–	–	–	1.00										
CFAF Lat	–	–	–	–	.71**	1.00									
CFAV Acc	–	–	–	–	–	–	1.00								
CFAV Lat	–	–	–	–	–	–	.32**	1.00							
Insight	–	–	–	–	–	–	–	–	1.00						
GEFT	.47**	.20**	.47**	.20**	.51**	.12*	.49**	.10	.37**	1.00					
PFBT	.37**	.28**	.33**	.28**	.25**	.19*	.34**	.23**	.26**	.51**	1.00				
MC Name	.12	.03	.12	.05	.13	-.01	.16	-.09	.13	.18**	.32**	1.00			
MC Num	.02	.01	.01	-.01	.07	.02	.03	-.18*	.09	.08	.22**	.69**	1.00		
SI Bias	.04	.07	.04	.05	.16*	.05	.08	.01	.05	.10	.26**	.13	-.02	1.00	
SI Sensitivity	.47**	.29**	.45**	.28**	.53**	.53**	.40**	.15**	.31**	.33**	.31**	.08	.04	-.12	1.00

Note. *N* per cell ranged from 127 to 466; **significant at the 0.01 level (two-tailed), *significant at the 0.05 level (two-tailed).

To further explore the role of pattern recognition, standardized aggregate scores of the BIS creativity subtests and KIT subtests were formed. Using a small subsample of participants who took both cognitive ability and pattern recognition tests ($n = 48$), a series of regressions were undertaken, in which scores for each mental flexibility test were regressed on BIS creativity component scores entered first, KIT induction factor scores entered second, and SI sensitivity scores entered last. Results of the regression with CFAF scores as the dependent variables approached significance. All of the predictors in the model (BIS creativity, KIT induction and SI sensitivity) explained 15% of total variance (F change $(4, 44) = 2.37$, $p = .08$). SI sensitivity scores were the only predictor that approached significance in the model ($\beta = 2.654$, $t = 1.712$, $p = 0.09$), explaining 6% of the variance over and above variance explained by cognitive ability scores. Results of the regressions with FI, FM, and Insight tests scores as dependent variables were not significant. However, sample size limits the power to detect small effect size in multiple regressions with three predictors.

In sum, regression analyses and correlation analyses suggest that pattern recognition test scores are significantly related to newly developed mental flexibility test scores.

Discussion. The newly developed mental flexibility tests of Flexible Inference (FI), Flexible Mapping (FM), Counterfactual Analogies–Figural (CFAF), Counterfactual Analogies–Verbal (CFAV), and Insight showed adequate reliability and preliminary evidence of construct and criterion-related validity as measures of the ability to cope with novelty. FI and FM tests, designed to measure performance components of flexible cognition, showed a consistent and expected pattern of association with measures of divergent and convergent cognitive ability and criterion measures. Correlations between FI and FM test scores and scores on divergent thinking tests (BIS creativity component subtests) and convergent thinking tests (BIS processing and KIT induction factor subtests) were small to moderate in size and explained 25% of the variance in regression analyses. Small correlations between FI and FM test scores and criterion measures (college GPA, SR flexible thinking, SR flexible behavior) were found, as expected. CFAF and CFAV tests, designed as an experiential assessment of flexible cognition, also showed a consistent and expected pattern of association with measures of divergent and convergent cognitive ability and criterion measures similar to that of FI and FM tests. Correlations between CFAF and CFAV test scores and scores on divergent thinking tests (BIS creativity component subtests) and convergent thinking tests (KIT induction factor subtests) were small to moderate in size and explained 21% of the variance in regression analyses. As expected, there were small correlations between CFAF and CFAV test scores and criterion measures (college GPA, SR flexible thinking, SR flexible behavior). The Insight test, also designed as an experiential assessment of flexible cognition, showed a similar and expected pattern of association with measures of divergent and convergent cognitive ability and criterion measures. Correlations between Insight scores and scores on divergent thinking tests (BIS creativity component subtests) and convergent thinking tests (KIT induction factor subtests) were small to moderate in size and explained 22% of the variance in regression analyses. Small correlations between Insight scores and criterion measures scores (college GPA, SR flexible thinking, SR flexible behavior) were also found.

The mental flexibility test battery showed strong evidence of convergent and discriminant validity. One factor explained 70% of the variance in factor analysis of the test battery. Moreover, the latent mental flexibility test factor was structurally differentiated from the latent cognitive ability factors. Intercorrelations among individual tests in the mental flexibility battery suggest adequate convergent and discriminant validity, with the exception of FI and FM tests, which were highly correlated. The FM test showed a slightly stronger relation to criterion measures and may be the better alternative.

The mental flexibility test battery showed evidence of incremental criterion-related validity. Taken together the mental flexibility tests explained 11.5% more variance in college GPA and 9% more variance in creative award over and above cognitive ability measures in regression equations.

Unexpected Results. There were a few unexpected results worthy of discussion. First, the BIS creativity component figural subtest (ZF) scores were not associated with mental flexibility test scores and KIT Induction factor subtest scores. It is possible that the translation of this subtest was problematic. The BIS subtests were scored by raters who reported the ZF as the most difficult to score. Second, the Soluble/Insoluble bias scores were not associated with pattern recognition test scores and mental flexibility test scores. It is possible that the test functioned more like a forced-choice task than a rating task. Forced choice tasks are only suitable for measuring sensitivity because the comparison does not involve a criterion (Stanislaw & Todorov, 1999). Test takers were presented with four answer options and an insoluble response option. The four answer options may have made use of a criterion for guessing unlikely.

Sex differences were found on the Insight test with males scoring higher than females. Sex differences have been reported on the GEFT (Witkin et al., 2002) and PFBT (Likert & Quasha, 1970). However, these tests did not show evidence of sex differences in this investigation. Moreover, sex differences were not found in tests of cognitive ability. Therefore, it seems unlikely the sex differences found on the Insight test could be attributable to selection bias. Insight test questions were spatial in nature and required both the capacity to think spatially and formulate novel solutions to spatial problems. It is possible that the increased cognitive demand for novel manipulations of spatial problems may account for sex differences on the Insight test.

In regard to personality measures, newly developed mental flexibility tests did not relate as expected with the Cognitive Flexibility Scale (CFS), which is a self-report survey designed to measure three components of cognitive flexibility including: (a) awareness of available options and alternatives; (b) willingness to be flexible and adapt to situations, and (c) self-efficacy in being flexible on a 6-point scale, ranging from 1 (strongly disagree) to 6 (strongly agree). The CFS has been shown to be related to communication competence, confidence, assertiveness, and responsiveness. In this research, it correlated with NEO Extraversion but not NEO Openness. Due to the self-report nature of the CFS, social desirability may have impacted scores. Alternatively, CFS may be more purely a measure of initiative rather than flexibility.

Pattern Recognition. Newly developed mental flexibility tests showed a consistent and strong pattern of association with measures of pattern recognition. Correlations between FI, FM, CFAF, and CFAV test scores and pattern recognition test scores (GEFT, PFBT, and SI sensitivity) were small to moderate in size, ranging from .33 to .54. In regression analyses, pattern recognition measures explained 39% to 46% of variance in the mental flexibility test scores. The correlation between Insight test scores and pattern recognition scores was also strong, with correlations ranging from .26 to .37, with pattern recognition measures together explaining 27% of variance in Insight test scores. Factor analysis did not reveal a factor that differentiated mental flexibility tests from pattern recognition tests. Research design did not permit a more in-depth analysis of the relationship, but findings suggest that pattern recognition may indeed be an important predictor of mental flexibility.

GENERAL DISCUSSION

The mental flexibility test battery can be distinguished from traditional tests of fluid ability in that it is theory–based and designed to measure flexibility at multiple levels of analysis. Flexible Inference (FI), Flexible Mapping (FM), and Counterfactual Analogies Tests–Figural and Verbal (CFAF and CFAV) build on traditional fluid ability tasks and incorporate an additional aspect of flexible thinking in each assessment. With Flexible Inference and Flexible Mapping tests, based on traditional classification and analogy tasks, a mental shift in the class and type of stimuli is required to correctly solve problems. With CFAF and CFAV tests, also based on traditional analogy tasks, counterintuitive assumptions must be effectively applied to correctly solve problems. The Insight test is a unique application of mind puzzles that require mentally restructuring information to solve them correctly.

Another way that the mental flexibility test battery is distinguished from traditional measures of fluid intelligence is the dynamic testing method of assessment. Dynamic assessment can be a more sensitive measure of individual differences in mental processing. With FI and FM, test items are made up of three item–parts that require a shift in mental sets. Performance is measured three times on both accuracy and latency. In addition, test-takers are given the opportunity to reflect on their reasoning in FI and FM tests, which can further distinguish individual differences associated with the theorized capacity to develop mental flexibility. Thus, accuracy and latency scores may reflect more specific components of mental processing involved in flexible thinking.

The mental flexibility test battery seems to measure something more than traditional tests of fluid ability, as suggested by findings suggestive of incremental criterion-related validity (11.5% more variance explained in college GPA and 9% more variance explained in creative award). The strong association found between newly developed mental flexibility tests and measures of pattern recognition, stronger than with traditional measures of fluid ability in some tests, raises the question as to whether there may be a shared flexible ability factor that has not yet been identified.

The preliminary findings are promising and further refinement and testing of the mental flexibility test battery seems warranted. First the tests could be further revised to reduce test length and remove redundancies. For example, because FI and FM tests are so highly correlated and reflect an analogous pattern of results in regard to construct-validity, one of these tests could be omitted from the test battery. In addition, with the CFAV test, results suggest that items with irrelevant premises fall within the same latent dimension as items with familiar premises and, therefore, could be considered for removal from the test.

Field testing a revised mental test battery with a broader sample from the adult population would provide the opportunity to further explore construct and criterion-related validity. In particular, results tempt further examination of the contribution of pattern recognition relative to traditional measures of fluid ability. In addition, continued field testing would permit more in-depth analysis of the specific contribution of individual tests in prediction criterion measures. Future research should examine questions about group differences to ensure that measurement is culturally fair.

CONCLUSION

Sternberg's (1985) theory of successful intelligence articulates a rich theoretical framework from which to develop instruments that predict real-world performance beyond traditional tests of intelligence. This project represents an initial attempt to create a theory-based mental flexibility test battery that measures the ability to cope with novelty more broadly than traditional measures of fluid intelligence. This mental flexibility test battery expands our understanding of the underlying mental processes that link flexible thinking to behavior. Further development and testing of this instrument promises to someday make it possible to select military leaders who are highly capable of coping with novelty. Next steps could include assessing the predictive validity of the test battery to performance measures of adaptive leader behaviors in the field. Potential applications to U. S Army Leadership could include testing for job classification and adaptation of testing methods to leadership training and development protocols.

REFERENCES

Andrew, D. M. & Paterson, D. G. (1959). *Minnesota Clerical test manual* (1959 revision). New York: The Psychological Corporation.

Bal, S (1988). Creativity, cognitive style, and academic achievement among university students. *Psychological Studies*, 33, 1, 10-13.

Barron, F. & Harrington, D. M (1981). Creativity, intelligence, and personality. *Annual Review of Psychology,* 32, 439-476.

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). *European Journal of Psychological Assessment*, 18, 2, 97-112.

Brand, C. (1996). *The g factor: General intelligence and its implications.* Chichester, England: Wiley.

Brody, N. (2000). History of theories and measurements of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp.16-33). New York: Cambridge University Press.

Bucik, V., & Neubauer, A. C. (1996). Bimodality in the Berlin Model of Intelligence Structure (BIS): A replication study.

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

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.

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

Cattell, R. B., & Cattell, H. E. P. (1963). *Measuring intelligence with the Culture Fair Tests.* Champaign, IL: Institute for Personality and Ability Testing.

Costa, P.T, Jr., & McCrae, R.R. (1992*). The NEO Personality Inventory-R: Professional manual.* Lutz, FL: Psychological Assessment Resources, Inc.

Crocker, L. & Algina, J. (1986). *Introduction to classical & modern test theory*. New York: Harcourt Brace Jovanovich College Publishers.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-references cognitive tests*. Princeton, NJ: Educational Testing Service.

Embretson, S. E., & McCollam, K. (2000). Psychometric approaches to the understanding and measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 423-444). New York: Cambridge University Press.

Embretson, S. E., & Prenovost, L. K. (2000). Dynamic cognitive testing: What kind of information is gained by measuring response time and modifiability? *Educational and Psychological Measurement, 60*(6), 837-863.

Fixx, J. F. (1972). *More games for the super-intelligent.* New York: Warner Books.

Grigorenko, E. L., & Sternberg, R. J. (1995). Thinking styles. In D. H. Saklorske and M. Zeidner (Eds.), International Handbook of Personality and Intelligence (pp. 205-229). New York: Plenum Press

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267-293.

Guilford, J. P. (1967). *The nature of human intelligence.* New York: McGraw-Hill.

Guilford, J. P. (1982). Cognitive psychology's ambiguities: Some suggested remedies. *Psychological Review, 89,* 48-59.

Guilford, J. P. (1985). A sixty-year perspective on psychological measurement. *Applied Psychological Measurement*, 37, 11-19.

Guilford, J. P., Berger, R. M., & Christensen, P. R. (1955a). A factor-analytic study of planning, I, Administration of tests and analysis of results. Report No. 12 from the Psychological Laboratory. University of Southern California, Los Angeles.

Guilford, J. P., Frick, J. W., Christensen, P. R., & Merrifield, P. R. (1957). A factor-analytic study of flexibility in thinking. Report No. 18 from the Psychological Laboratory. University of Southern California, Los Angeles.

Guilford, J. P., Kettner, N. W., & Christensen, P. R. (1955). The relation of certain thinking factors to training criteria in the U. S. Coast Guard Academy. Report No. 13 from the Psychological Laboratory, University of Southern California, Los Angeles.

Guilford, J. P., Wilson, R. C., & Christensen, P. R. (1952). A factor-analytic study of creative thinking, II. Administration of tests and analysis of results. Report No. 8 from the Psychological Laboratory, University of Southern California, Los Angeles.

Gustafsson, J.-E., Demetriou, A., & Efklides, A. (1994). Hierarchical models of intelligence and educational achievement. In *Intelligence, mind, and reasoning: Structure and development.* (pp. 45-73). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Guthke, J. R., Beckmann, J. F., Lidz, C., & Elliott, J. G. (2000). The learning test concept and its application in practice. In *Dynamic assessment: Prevailing models and applications*. (pp. 17-69). New York, NY, US: Elsevier Science.

Guthke, J. R., Beckmann, J. F., Sternberg, R. J., Lautrey, J., & Lubart, T. I. (2003). Dynamic assessment with diagnostic problems. In *Models of intelligence: International perspectives*. (pp. 227-242). Washington, DC, US: American Psychological Association.

Herrnstein, R. J, & Murray, C. (1994). *The bell curve*. New York: Free Press.
Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *The encyclopedia of human intelligence (Vol. 1)* (pp. 443-451). New York: Macmillan.

Horn, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.) Intelligence: Measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G. Humphreys, (pp. 29-72), VIII, Champaign, IL: University of Illinois Press.

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107-129.

Intelligence and its measurement: A symposium. (1921). *Journal of Educational Psychology, 12*, 123–147, 195–216, 271–275.

Jäger, A. O. (1982). Mehrmodal Klassifikation von Intelligenzleistungen: Experimentall kontrollierte Weiterentwichklung eines deskriptiven Intelligenzstrukturmodells [Multimodal classification of intelligence tests: Experimentally controlled development of a descriptive model of intelligence structure]. *Diagnostica*, *28*, 195-225.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven [Research on intelligence structure: Competing models, new developments, perspectives]. *Psychologische Rundschau*, *35*, 21-35.

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

Kleine, D., & Jäger, A. O. (1987). Replikation des Berliner intelligenzstrukturmodells (BIS) bei brasilianischen schülern and studentem. *Diagnostica, 33*, 14-29.

Kleine, D., & Jäger, A. O. (1989). Kriteriumsvalidität eines neuartigen tests zum Berliner intelligenzstrukturmodell: Eine untersuchung an brasilianischen schülern and studentem. *Diagnostica, 35*, 17-37.

Likert, R., & Quasha, W. H. (1970). *Revised Minnesota paper form board test manual, 1970 revision*. New York: The Psychological Corporation.

Lord, F. M. (1952). The relationship of the reliability of multiple choice items to the distribution of item difficulties. *Psychometrika*, 18, 181-194.

Lubart, T. I., & Sternberg, R. J. (1995). An investment approach to creativity: Theory and data. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 269-302). Cambridge, MA: MIT Press.

Margolis, H. (1987). Patterns, thinking, and cognition: A theory of judgment. Chicago: The University of Chicago Press.

Marr, D. B., & Sternberg, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and nongifted children to relevant and irrelevant novel stimuli. *Cognitive Development, 1*(1), 53-72.

Martin, M.M., & Anderson, C.M. (1998). The cognitive flexibility scale: Three validity studies. *Communication Reports*, *11*, 1.

Martin, M.M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports*, *76*, 623-626.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 288-294.

Paterson, D. G., Elliott, R. M., Anderson, L. D., Toops, H. A., & Heidbreder, E. (1930). *Minnesota mechanical ability tests*. Minneapolis: University of Minnesota Press.

Rosas, R. (1991). *Replikation des Berliner intelligenz-struktur-modells (BIS) und vorhersagbarkeit des stdienerfolgs bei chienischen studentem.* Unpublished doctoral dissertation, Freien Universität Berlin, Fachbaereich Erziehungs und Unterrichtwissenschaften, Berlin, Germany.

Seifert, C. M., & Patalano, A. L. (1991). Memory for incomplete tasks: A re-examination of the Zeigarnik effect. In Proceedings of the Thirteen Annual Conference of the Cognitive Science Society (pp. 114-119). Hillsdale, NJ: Erlbaum.

Snodgrass, J. G., & Corwin, J. (1988). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture test. *Perceptual and Motor Skills*, *67*, 1, 3-36.

Solso, L. R. (2003). *The Psychology of Art and the Evolution of the Conscious Brain*. MIT Press: Bradford Books.

Spearman, C. (1923). *The nature of intelligence and the principles of cognition.* London: Macmillan.

Spearman, C. (1927). *The abilities of man.* New York: Macmillan.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137-149.

Sternberg, R. J. (1981). Intelligence and nonentrenchment. *Journal of Educational Psychology, 73*, 1-16.

Sternberg, R. J. (1982). Nonentrenchment in the assessment of intellectual giftedness. *Gifted Child Quarterly, 26*, 63-67.

Sternberg, R. J. (1983). Componential theory and componential analysis: Is there a "Neisser" alternative? *Cognition, 15*, 199-206.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1987). Teaching intelligence: The application of cognitive psychology to the improvement of intellectual skills. In J. B Baron, and R. J. Sternberg, (Eds). Teaching thinking skills: Theory and practice. (pp. 182-218). Xi.. NY: W H Freeman/Times Books/ Henry Holt & Co.

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

Sternberg, R. J. (1997). *Successful intelligence.* New York: Plume.

Sternberg, R. J. (1999) The theory of successful intelligence. *Review of General Psychology, 3,* 292-316.

Sternberg, R. J., & Davidson, J. E. (1982). The mind of the puzzler. *Psychology Today, 16*, 37-44.

Sternberg, R. J., & Detterman, D. K. (1986). *What is intelligence?* Norwood, N.J.: Ablex Publishing Corporation.

Sternberg, R. J., & Gardner, M. K. (1982). A componential interpretation of the general factor in human intelligence. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 231-254). Berlin: Springer-Verlag.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. *Journal of Experimental Psychology: General, 112*, 80-116.

Sternberg, R. J., & Gastel, J. (1989a). Coping with novelty in human intelligence: An empirical investigation. *Intelligence, 13,* 187-197.

Sternberg, R. J., & Gastel, J. (1989b). If dancers ate their shoes: Inductive reasoning with factual and counterfactual premises. *Memory and Cognition, 17*, 1-10.

Sternberg, R. J., & Gastel, J. (1989c). If dancers ate their shoes: Inductive reasoning with factual and counterfactual premises. *Memory & Cognition, 17*(1), 1-10.

Sternberg, R. J., & Gastel, J. (1989d). Coping with novelty in human intelligence: An empirical investigation. *Intelligence, 13,* 187-197.

Sternberg, R. J., & Grigorenko, E. L. (Eds.) (1997). *Intelligence, heredity, and environment.* New York: Cambridge University Press.

Sternberg, R. J., & Lubart, T. I. (1991). An investment theory of creativity and its development. *Human Development, 34*(1), 1-31.

Sternberg, R. J., & Lubart, T. I. (1995*). Defying the crowd: Cultivating creativity in a culture of conformity*. New York: Free Press.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist, 51*(7), 677-688.

Sternberg, R. J., & Williams, W. M. (1997). Does the Graduate Record Examination predict meaningful success in the graduate training of psychologists? A case study. *American Psychologist, 52*, 630-641.

Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. L. (2001). Confirmatory Factor Analysis of the Sternberg Triarchic Abilities Test (Multiple-Choice Items) in Three International Samples: An Empirical Test of the Triarchic Theory of Intelligence. *European Journal of Psychological Assessment, 17,* 1-16.

Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude-treatment interaction. *European Journal of Psychological Assessment,15(1)*, 1-11.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Suen, H. (1990). *Principles of test theories*. NJ, England: Lawrence Erlbaum Associates, Inc.

Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Tetewsky, S. J., & Sternberg, R. J. (1986). Conceptual and lexical determinants of nonentrenched thinking. *Journal of Memory and Language, 25*, 202-225.

Thurstone, L. L. & Thurstone, T. G. (1939) *American council on education psychological examination for college freshman*. Series V, Volume IV, No. 2. Princeton, NJ: Educational Testing Service.

Thurstone, L. L. (1925). The psychological test program. *Educational Record*, *6*, 69-

Thurstone, L.L. (1926). The psychological test program. *Educational Record*, *7*, 114-

Thurstone, L. L. (1944). *A factorial study of perception. Chicago: University of Chicago Press.*

Vernon, P. E. (1971). *The structure of human abilities*. London: Methuen.

Witkin, H. A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, *19*, 1-15.

Witkin, H. A. (1973). The role of cognitive style in academic performance and in teacher-student relations. Unpublished report, Educational Testing Service, Princeton, NJ.

Witkin, H. A. (1975). Some implications of research on cognitive style for problems of education. In J. M. Whitehead (Eds.), *Personality and learning*. London: Hodder and Stoughton.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. (1962). *Psychological differentiation*. New York: Wiley.

Witkin, H, Oltman, P, Raskin, E, & Karp, S, (1971). A manual for the Embedded Figures Test. Consulting Psychologists Press, Inc.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. (2002). *Group embedded figures test manual*. Menlo Park, CA: Mindgarden, Inc.

APPENDIX A

Description of Reference Tests

Test Name	Abbreviation	Validity
divergent calculus	(BIS-DR)	divergent thinking – numerical
number riddle	(BIS-ZR)	divergent thinking – numerical
drawing completions	(BIS-ZF)	divergent thinking – figural
object designs	(BIS-OJ)	divergent thinking – figural
multiple uses	(BIS-AM)	divergent thinking – verbal
Masselon	(BIS-MA)	divergent thinking – verbal
Bongard	(BIS-BG)	classification – figural
letter set	(FKit-LS)	classification – verbal; induction
making groups	(FKit-MG)	classification - verbal
toothpick test	(FKit-TP)	adaptive flexibility - figural
location test	(FKit-LC)	rule inference

APPENDIX B

Scoring Soluble/Insoluble Sensitivity and Bias Indices

Sensitivity index (Pr)

Pr = Hit rate – False alarm rate

Hit rate = correct identification of soluble items indicated by providing an answer whether correct or incorrect since accuracy is a matter of reasoning ability.

$$\text{Hit rate} = \frac{(\#\text{correct soluble} + 0.5)}{(\#\text{soluble items} + 1)}$$

False alarm rate = response tendency to "see" an item as soluble when uncertain.

$$\text{False alarm rate} = \frac{(\#\text{incorrect insoluble} = 0.5)}{(\#\text{insoluble items} + 1)}$$

Pr ranges between -1.0 and 1.0

Pr = 0 represents "zero knowledge" (not recognizing any pattern)

Pr = -1.0 represents maximal erroneous "knowledge"

Pr = 1.0 maximum in recognizing "patterns"

Bias index (Br)

$$B_r = \frac{\text{false alarm rate}}{(1-P_r)}$$

Bias index = response tendency if one must guess whether an item is soluble or insoluble.

Bias ranges between 0 and 1, neutral Bias at Br = 0.5

If Br < 0.5 then conservative Bias, that is: if not sure then rather saying: not soluble

If Br > 0.5 then liberal Bias, that is: if not sure, then tendency to guess and to provide an answer where there is none.

Underlying two high threshold model (Snodgrass & Corwin, 1988):