



Michigan Technological University
Create the Future Digital Commons @ Michigan Tech

Dissertations, Master's Theses and Master's
Reports - Open

Dissertations, Master's Theses and Master's
Reports

2014

COMPONENT NUMERACY SKILLS AND DECISION MAKING

Saima Ghazal
Michigan Technological University

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Cognitive Psychology Commons](#), and the [Quantitative Psychology Commons](#)

Copyright 2014 Saima Ghazal

Recommended Citation

Ghazal, Saima, "COMPONENT NUMERACY SKILLS AND DECISION MAKING", Dissertation, Michigan Technological University, 2014.

<https://doi.org/10.37099/mtu.dc.etds/964>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Cognitive Psychology Commons](#), and the [Quantitative Psychology Commons](#)

COMPONENT NUMERACY SKILLS AND DECISION MAKING

By

Saima Ghazal

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2014

© 2014 Saima Ghazal

This dissertation has been approved in partial fulfillment of the requirements for the Degree of DOCTOR OF PHILOSOPHY in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Dissertation Co-Advisor: *Edward T. Cokely*

Dissertation Co-Advisor: *Rocio Garcia-Retamero*

Committee Member: *Adam Feltz*

Committee Member: *Shari Stockero*

Department Chair: *Susan L. Amato-Henderson*

To
Adil
Ashar
&
Azeem

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	11
ABSTRACT	13
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: NUMERACY.....	5
2.1 MEASURING NUMERACY	9
2.1.1 <i>Advances in Numeracy Assessment</i>	12
2.1.2 <i>The Berlin Numeracy Test</i>	14
2.1.3 <i>Other Numeracy Assessments</i>	15
CHAPTER 3: NUMERACY AND DECISION MAKING	17
3.1 CONFIDENCE & DELIBERATION	18
3.2 TRANSFER AND IDENTICAL ELEMENTS.....	24
CHAPTER 4: STUDY 1.....	27
4.1 INTRODUCTION.....	27
4.2 TEST CONSTRUCTION	28
4.2.1 <i>Item Selection</i>	28
4.2.2 <i>Phase I: Preliminary Studies</i>	29
4.2.3 <i>Phase II: The 40 Item Test</i>	30
4.2.3.1 Participants	30
4.2.3.2 Materials and Procedure.....	30
4.2.4 <i>Data Analysis</i>	33
4.2.4.1 Model Selection	33
4.2.4.2 Item Calibration.....	35

4.2.4.3 Item analysis of Individual subscales (10 items each)	36
<i>4.2.5 Phase III: A 20 Item Test</i>	38
4.2.5.1 Subscales Analysis (5 items per scale)	43
4.2.5.2 Demographic analysis	46
<i>4.2.6 Summary of the Study 1</i>	46
CHAPTER 5: STUDY 2	51
5.1 INTRODUCTION	51
5.2 GENERAL HYPOTHESES.....	54
5.3 METHOD	55
5.3.1 <i>Participants</i>	55
5.3.2 <i>Material</i>	56
5.3.3 <i>Procedure</i>	60
5.4 DATA ANALYSIS	61
5.4.1 <i>Psychometric Analysis</i>	61
5.4.1.1 Reliability Analysis.....	67
5.4.1.2 Factor Analysis:	68
5.4.2 <i>Component Analysis</i>	71
5.4.2.1 Data and Descriptive Analysis	71
5.4.2.2 Theoretical Models	73
5.4.2.3 Individual Models (Components as Single Predictors)	77
5.4.3 <i>Composite and Interaction Analysis</i>	78
5.4.3.1 Interaction Analysis.....	81
5.4.3.2 Composite Analysis	82
5.4.4 <i>Competitive Analysis</i>	84
5.4.4.1 Overall Superior Decision Making Skill (SDMS)	85
5.4.4.2 Risk Literacy	87

5.4.4.3 Standard Risky Decisions	88
5.4.4.4 Ecological Decision Making	89
5.4.4.5 ADMC	90
<i>5.4.5 Model Fitting.....</i>	<i>95</i>
5.4.5.1 Generating profiles.....	99
5.4.5.2 Model Recovery	100
<i>5.4.6 Summary of the Study 2.....</i>	<i>103</i>
5.5 DISCUSSION	103
<i>5.5.1 Theoretical Implications.....</i>	<i>105</i>
<i>5.5.2 Emerging and Future Applications.....</i>	<i>108</i>
<i>5.5.3 Concluding Remarks.....</i>	<i>111</i>
REFERENCES	113
APPENDIX A	133
APPENDIX B	137
APPENDIX C	139
APPENDIX D	141
APPENDIX E.....	145
APPENDIX F.....	151
APPENDIX G	221
APPENDIX I.....	235
APPENDIX J	241
APPENDIX K.....	251
APPENDIX L.....	257
APPENDIX M	269

Acknowledgements

I thank my Advisor Edward T. Cokely for his patience, guidance, support, and encouragement throughout my studies. I have been extremely lucky to have an advisor like him who invests such a huge amount of quality time into his students, who cares for his students and considers their success as his own success. This dissertation would not have been possible without him. I thank my co-advisor Rocio Garcia-Retamero whose guidance and assistance I greatly appreciate. I also thank the members of my dissertation committee; Adam Feltz and Shari Stockero for their feedback.

I am thankful to Ahmad Baniabedalruhman for his math expertise and guidance in developing the BNT-C test items. Also, I would like to thank all the participants who completed 3 hours of testing, and made this happen.

I must express my deepest thanks to Azeem, my loving husband, for his dream to see me as a “doctor”, and for his belief in me. He set me on this road. I am very grateful for his support and encouragement throughout this journey.

Finally, I am very thankful to my two little boys Adil, and Ashar, who patiently watched their mom working. I am thankful to my family, my brothers; Tariq, Amjad, Safdar, and my special sister Rani, for their endless prayers for me.

I am also grateful to the Higher education commission of Pakistan and University of the Punjab Pakistan, for providing me with funding for the PhD studies.

“The great body of physical science, a great deal of the essential fact of financial science, and endless social and political problems are only accessible and only thinkable to those who have had a sound training in mathematical analysis, and the time may not be very remote when it will be understood that for complete initiation as an efficient citizen of one of the new great complex worldwide States that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and write”

H. G. Wells (1904) *Mankind in the Making*

Abstract

Numeracy—i.e., one’s practical understanding of mathematics in context—is one of the strongest predictors of people’s general decision making skill, independent of other cognitive abilities (e.g., intelligence, working memory, attentional control). Despite notable scientific progress on the nature of numeracy and decision making, the cognitive and decision sciences have yet to investigate individual differences in numeracy components (e.g., algebra versus probability). In this dissertation, I report on my efforts to develop new measurement technology and quantitative models of cognitive and decision skills. Analyses include the first known investigations of the relations between the major adult component numeracy skills and general decision making skills. Specifically, Study 1 built on an established theoretical framework from adult education and used a two-parameter logistic IRT model to create the Berlin Numeracy Components Test (BNT-C) for college educated samples. Behavioral and analytic results indicated that the test efficiently measured full-scale adult numeracy and component numeracy skills (i.e. operations, probability, geometry, and algebra), with superior psychometric performance (e.g., difficulty, discriminability, and sensitivity). In Study 2, I investigated the links between the BNT-C, other numeracy tests, general cognitive abilities (e.g., intelligence, cognitive impulsivity), and general decision making skills (e.g., Adult Decision Making Competencies, Risk Literacy). Predictive modeling of behavioral data revealed that with few exceptions the BNT-C explained all types of decision skill better than any other individual ability assessment (e.g., intelligence v. impulsivity v. other numeracy tests). The BNT-C additionally outperformed the optimal linear combination of all combined ability tests when predicting overall general decision making skill, a finding that is consistent with a causal cognitive account of the relations between numeracy and general decision making skill. In accord with leading theory, component analyses indicated that operations and probability skills were robustly and uniquely tied to risk literacy and to nearly all general decision making sub-skill competencies independent of the influence of other cognitive abilities (e.g., intelligence). Analyses also provided the first evidence that algebra and geometry sub-skills are uniquely linked to some essential

general decision skills in educated adults (e.g., confidence calibration, sunk costs, ratio bias). Discussion focuses on theoretical implications and factor analytic modeling of the relations between numeracy, its component skills, and superior decision making. Adaptive test construction and potential applications in training and personalized decision support are also briefly discussed.

Chapter 1: Introduction

It is well established that mathematical skills are among the most influential educational factors contributing to economic prosperity in industrial countries (Hunt & Wittmann, 2008). In part, the link between mathematical skills and economic prosperity reflects the influence of specialized human capital (i.e., expertise) in advanced STEM-related fields (i.e., science, mathematics, engineering and technology). However, beyond the influence of these domain-specific factors, research indicates that mathematical skills may give rise to economic prosperity via their influence on domain-general superior decision making. Individual differences in numeracy—i.e., one’s practical understanding of mathematics in context—have been found to be one of the strongest domain-general predictors of superior judgment and decision making across common and high-stakes financial, health, and consumer domains (Banks, O’Dea, & Oldfield, 2010; Cokely & Kelley, 2009; Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Lipkus & Peters, 2009; Peters & Levin, 2008; Peters, Vastfjall, Slovic, Mertz, Mazzocco, & Dickert, 2006; Reyna, Nelson, Han, & Dieckmann, 2009). Moreover, although mathematical skills are robustly linked to full scale IQ and general fluid intelligence, research indicates that numeracy tends to be a unique predictor of superior decision making independent of the influence of fluid intelligence, working memory and attentional control capacities. Research further suggests that even small differences in numeracy can confer considerable benefits across a wide range of socioeconomic conditions. In short, numeracy matters: From influencing the likelihood of contracting HIV among villagers in rural Ghana to predicting surgeons’ ability to understand and

communicate surgical risks (Garcia-Retamero, Wicki, Cokely, & Hanson, 2014; Ghazal, Cokely, & Garcia-Retamero, 2014; Peters, 2012; Reyna et al., 2009).

For more than 50 years, researchers have studied the causes and consequences of *numeracy* (Huff & Geis, 1954; Paulos, 1988), including extensive longitudinal studies conducted in large diverse samples such as the National Assessment of Adult Literacy (NAAL; Kutner et al., 2006) and the Program for International Student Assessment (PISA; OECD, 2012). There is wide agreement that numeracy is somewhat different from pure mathematical skill, referring instead to mathematical or quantitative literacy with emphasis on “mathematics in context” akin to that described in the *US Common Core State Standards Initiatives* (Steen, 1990; see also Nelson, Reyna, Fagerlin, Lipkus, & Peters, 2008, and Reyna et al., 2009). Numeracy is also a broad construct that includes various components and sub-skills. For example, Ginsberg et al. (2006) describe four major components of adult numeracy, including number sense, probability, geometry, and algebra. Unfortunately, modern research instruments used in the cognitive and decision sciences tend to focus on assessing differences in full scale numeracy or otherwise assess one single subcomponent (i.e. statistical numeracy). Although there is a growing body of research linking numeracy and superior decision making, there is virtually no research available investigating the relative importance of component numeracy sub-skills with respect to superior decision making. To begin to address this limitation, in what follows I present a psychometric optimization study where I designed and created a brief test of numeracy and its major sub-skills for use with educated samples from industrialized countries i.e. Berlin Numeracy Component Test (BNT-C). I

then using this tool investigated the links between BNT-C, standard numeracy instruments, component numeracy sub skills, general cognitive abilities, and decision making skill.

Chapter 2: Numeracy

Researchers and philosophers have found it difficult to agree on an exact and uncontroversial theoretical definition of mathematics – which is ironic for a field otherwise characterized by its emphasis on elegance and precision. Thankfully, quantitative skills are easier to operationalize and measure, and there is a long history of psychological and educational investigation of mathematical cognition. Exact theoretical definitions of numeracy vary in their phrasing and emphasis (see for example: Coben, 2000; Cockcroft, 1982; Crowther, 1959; Gal, van Groenestijn, Manly, Schmitt, & Tout, 2003; Johnston, 1994; Lindenskov & Wedege, 2001; and Steen, 2001). For example, on one view the construct of numeracy is thought to reflect the “array of mathematically related proficiencies that are evident in adults’ lives... including a connection to context, purpose, or use... for active participation in the democratic process and... in the global economy” (Ginsburg, Manly, Schmitt, 2006). Others state that “Numeracy describes the aggregate of skills, knowledge, beliefs, dispositions, and habits of mind — as well as general communicative and problem solving skills, that people need in-order to effectively handle real world situations or interpretive tasks with embedded mathematical or quantitative elements” (Gal, 1995, p. 9). Whereas pure mathematics is abstract and context free, “Unlike mathematics, numeracy does not so much lead upward in an ascending pursuits of abstraction as it moves toward an ever richer engagement with life’s diverse context and situations” (Orrill, 2001, p. xv iii).

A careful review of various accounts of the construct of numeracy indicates that there are important nuanced differences (e.g., emphasis on the relative sophistication of

mathematics versus numeracy). There is evidence that essential conceptual distinctions have been subject to change over time with advances in knowledge and with shifts in lifestyles, technology, and educational practices (Steen, 1990). Specific differences also emerge as a function of the domain in which numeracy is situated and studied (e.g., health literacy v. risk comprehension; see also Reyna, Nelson, Han, & Dieckman, 2009). Nevertheless, overall some general consensus emerges. Numeracy tends to be defined as a practical or functional application and use of mathematical knowledge and skills for effective functioning in one's group or community (Beazley, 1984). In short, numeracy helps people make good decisions (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012; Peters, 2012). Despite the presence of many carefully articulated models and frameworks, there is no single unifying theory systematically linking numeracy, mathematics, and cognition. Unfortunately, there is also relatively little data documenting the relations between component mathematical abilities and the essential functions of numeracy (e.g., effective decision making).

Theoretically, at the more basic levels numeracy is said to involve an understanding of the "real number line, time, measurement, and estimation" whereas higher levels often emphasize an "understanding of ratio concepts, notably fractions, proportions, percentages, and probabilities" (Reyna et al., 2009). More generally, mathematical knowledge has been classified in number of ways that vary depending on one's reference class. Steen (1990) defined numeracy in terms of six broad categories: quantity, dimension, shape, pattern, uncertainty and change. Gal et al., (2005) who conducted the Adult Literacy and Life Skills (ALL) survey, defined and assessed

numeracy as a set of five fundamental ideas that characterize mathematical knowledge including: quantity and number, dimension and shape, pattern, function and relationship, data and chance, and change. Another well-established framework presented by Golbeck (2005) defined numeracy in the context of health. Golbeck's elements of numeracy included four overlapping categories of numerical information, namely: basic (e.g., ability to identify and read numbers), computational (e.g., counting and arithmetic), analytical (e.g., inference, estimation, proportion, percentage, frequencies, basic graphs), and statistical (e.g., basic probability, statistics, and risk assessment).

With regard to adult literacy and numeracy, perhaps the most comprehensive framework of numeracy and its components is provided by Ginsburg and colleagues (2006). Ginsburg et al. used the existing models, frameworks, and governmental standards—including US and International standards—to conceptualize and describe a “core” collection of numeracy components among adults. In total, based on systematic conceptual analysis of 29 adult mathematical and numeracy frameworks, they identified four major mathematical competencies thought to be essential components of numeracy. The four numeracy components are as follows:

- i) **Number and Operation Sense:** includes mainly the mathematical operations such as, place value, computation, estimation, rates, ratios, proportions and percentages.
- ii) **Statistics and Probability:** collection, organization, and display of data, analysis and interpretation of data, chance and probability, and inferential reasoning.

- iii) **Measurement and Shape:** includes measurement units, Pythagorean Theorem or trigonometric ratios, angles and lines, shapes, perimeter, area, and volume, length, width, height, and radius.
- iv) **Pattern Functions and Algebra:** includes tables, diagrams, and algebraic expressions, algebraic symbols, equations, mathematical modeling, and functions.

Although there are likely many reasonable, useful conceptualizations of adult numeracy and its components, the components identified by Ginsburg and colleagues (2006) are noteworthy for several reasons. First, they represent major classes of skills identified in nearly all other taxonomies. Second, they largely accord with the components identified in standards set out by the Core Common State Standards Initiative (CCSSI). The CCSSI is a major recent effort to create modern US educational standards, including standards for mathematics training and content, sponsored by the National Governors Association and the Council of Chief State School Officers. Finally, the components accord with general theoretical frameworks developed by the Educational Testing Service and the College Board following years of very large scale, sophisticated psychometric analyses of the mathematical abilities of people who were preparing to enter college in the United States. Surprisingly, however, despite the widespread use, evaluation, and consistent interpretations of component numeracy skills, modern research instruments designed to assess numeracy in the cognitive and decision sciences tend to focus almost exclusively on assessing differences in full scale numeracy or on assessing only one specific subcomponent.

2.1 Measuring Numeracy

Within the cognitive and decisions sciences, efforts to understand and measure numeracy tend to focus primarily on one's understanding of probability or statistical numeracy. The reasons for emphasis on statistical numeracy are both theoretical and practical. For example, patients often struggle to understand numerical, probabilistic information (e.g., risks of treatments or side effects) and thus it can be efficient to directly assess these specific skills when designing interventions. Generally, efforts to measure numeracy include both subjective and performance assessments. For example, one validated subjective assessment of numeracy often used in health and medical domains asks participants eight questions in which they judge their personal levels of numeracy (e.g., "How good are you at working with fractions;" Fagerlin, Zikmund-Fisher, Ubel, Jankovic, Derry, & Smith, 2007; Zikmund-Fisher, Smith, Ubel, & Fagerlin, 2007; and for subjective graph literacy see Garcia-Retamero, Cokely, & Ghazal, In preparation). Several studies indicate moderate-to-high correlations between objective and subjective measures (Fagerlin et al., 2007; Liberali, Reyna, Furlan, Stein, & Pardo, 2012; Weller, Dieckmann, Tusler, Mertz, Burns, & Peters, 2012; Zikmund-Fisher et al., 2007). Studies further show the subjective test can provide unique predictive power beyond intelligence test scores (Låg et al., 2013). Nevertheless, other research indicates that people can be highly overconfident in reporting their subjective numerical ability. For example, Sheridan, Pignone, and Lewis (2003) showed that 70% of subjects reported that they consider themselves to be "good with numbers", while only 2% of those respondents

correctly answered three objective numeracy questions (see also Dunning, Heath, & Suls, 2004).¹

Performance based numeracy assessments are the most commonly used methods in the allied decision sciences. The longest-standing and most widely used assessments of numeracy are based on classical testing theory, which estimates theoretical differences in abilities based on one's relative test score (Novick, 1966; see also Cokely et al., 2013, in press; Lipkus, Samsa, & Rimer, 2001; Peters et al., 2006; Schapira, Walker, & Sedivy, 2009; Schwartz et al., 1997). To illustrate, in 1997, Schwartz et al. conducted a seminal randomized cross-sectional numeracy study investigating the relations between numeracy and relative risk perceptions. Five hundred women were initially mailed the study stimuli and asked to participate. Respondents included 287 mostly older adult women (mean age 68 years) who were veterans with modest incomes (e.g., less than \$25,000 per year). The majority of participants had also completed high school (96%) and about a third had completed at least some college. Numeracy was assessed with three items that were similar to and based on items used in the NAAL survey (see previous section). Once scored, these items were used to predict the women's understanding of data presented in one of four formats (e.g., relative risk reduction versus absolute risk reduction with baseline). The women were asked to interpret the material provided and to report on the risks/benefits of mammography screening (e.g., "Imagine 1000 women exactly like you.

¹ The three items were from the test by Schwartz et al. (1997). These results suggest that subjective instruments are likely best suited for specific purposes, including rapid, rough numeracy assessment among people who have some math anxiety.

Of these women what is your best guess about how many will die from breast cancer during the next 10 years if they are not screened every year for breast cancer?”). Results indicated that about half of the women (i.e., 54%) accurately answered two questions, while only 20% accurately answered all three (i.e., most could not convert 1 in 1000 to 0.001). As expected, results also revealed a moderate positive correlation between participants’ final score and their relative risk reduction interpretations, providing evidence of decision-related criterion validity for the brief assessment.

The results of Schwartz et al. (1997) and the subsequent results provided by Lipkus et al. (2001) were timely for a number of reasons (for reviews see Cokely et al., 2012, in press).² First, the results provided additional evidence that among community samples in the United States some sizable proportion of individuals were likely to be statistically innumerate (e.g., 20% failed questions dealing with risk magnitude), a result that accords with findings from the NALS and NAALS national longitudinal surveys. Such findings are important as many efforts designed to support informed and shared decision making rest on an assumption that decision-makers are numerate (or at least sufficiently statistically numerate; see also Edwards & Elwyn, 2009, and Guadagnoli & Ward, 1998). Second, results indicated that domain framing (e.g., medical, financial, or abstract gambles) did not tend to affect test performance or comprehension. This finding indicates that various domain-specific items (e.g., items framed in terms of financial or

² There are also a number of performance measures of numeracy that assess one’s approximate number system—a related but independent theoretical construct. For a recent example of these tests see Lindskog, Winman, and Juslin (2013).

medical or gambling risks) can provide a reasonable basis for the assessment of domain-general statistical numeracy skills, although it is theoretically possible that domain familiarity will confer some additional decision performance advantages (Levy, Ubel, Dillard, Weir, & Fagerlin, 2014).

2.1.1 Advances in Numeracy Assessment

After more than a decade of research using classical tests of numeracy, research in the decision sciences has turned to modern psychometric testing paradigms—i.e., Item Response Theory (IRT) and its variants. In contrast to classical testing theory, item response theory requires modeling of probabilistic distributions over test taker's responses to specific items. The focus of test development is on the *item* rather than on the pooled responses to items as in classical testing theory. A full description of the theory is beyond the scope of this paper (see Lord, 1980; Van der Linden & Hambleton, 1997); however, it is useful to note that IRT tests improve predictive performance by eliminating item redundancy with estimated parameters including item difficulty (e.g., how hard is any particular item for a given trait level), discrimination (e.g., how sharply and consistently does an item distinguish individuals at higher versus lower trait levels), and guessing (e.g., true/false items will be guessed correctly 50% of the time). To illustrate, Schapira et al. (2012) developed the Numeracy Understanding in Medicine Instrument (NUMi) to provide a higher-fidelity assessment of basic health numeracy among less educated patient samples. The 20 item test was developed using a two-parameter IRT approach integrating four numeracy sub-skills (e.g., graph literacy, statistical numeracy). Results revealed that the NUMi test is robust and provides good

psychometric sensitivity that is suitable for use with less numerate individuals (e.g., older adult patient samples). Results also provided evidence of construct validity and unique predictive power (e.g., independent of the predictions of general intelligence tests).

Using a Rasch analysis, which is akin to a one parameter IRT-type approach, Weller et al. (2012) developed an eight item numeracy measure optimized for use with the general population of the United States. Test development involved comparison of 18 items taken from existing measures of numeracy and a cognitive reflection test. Specifically, items were drawn from tests developed by Lipkus et al. (2001) (which includes the items of Schwartz et al., 1997), and tests developed by Peters et al. (2007), and Frederick (2005). The resulting scale provides greatly improved psychometric discriminability when used with the general population of the United States. Evidence also indicates that the test provides stronger predictive validity for risk judgments (i.e., Lag, Bauger, Linburg, Friborg, 2013; Lipkus et al., 2001). Despite these notable improvements, one limitation of the Weller et al. (2012) abbreviated numeracy scale, as well as the test items analyzed by Lag and colleagues (2013), is that they combine two distinct types of test items with differential ranges of sensitivity to improve psychometric sensitivity the numeracy assessments. In particular, they include: (1) some relatively difficult items designed to measure cognitive impulsivity/reflection (i.e., the CRT by

Frederick, 2005) and (2) some relatively easy items designed to measure statistical numeracy.³

2.1.2 The Berlin Numeracy Test

Building on the work of Lipkus et al. (2001) and Schwartz et al. (1997), Cokely and colleagues (2012) developed a fast psychometric test of differences in *statistical numeracy* among educated samples of adults living in diverse industrialized countries (e.g., college students, working professionals, and computer literate adults). The test was created using new statistical numeracy items selected from a large pool of candidate items. All items were subjected to think aloud protocol analysis to control for potential confounds from factors such as linguistic confusion. The test was then developed using a decision tree application from the predictive modeling software DTREG (Sherrod, 2003). The analysis yielded several versions of the test (see www.RiskLiteracy.org for links and test format recommendation tools), including (i) the adaptive test that adjusts item difficulty based on a test-takers previous responses (2-3 items; about 2.5 minutes duration) and (ii) a traditional 4 item paper-and-pencil test (4 items; < 5 minutes duration). Psychometrically the decision tree's assessment approximates an item response

³ Although confirmatory factor analysis has indicated that the constructs can be considered one factor, there is reason to be cautious with this interpretation. The two types of items have been found to dissociate in theoretically notable ways, differentially predicting financial judgments, reasoning, and risk comprehension (Cokely et al., 2012; Cokely, Parpart, & Schooler, 2009; Di-Girolamo, Harrison, Lau, & Swarthout, 2014; Låg et al., 2013; Liberali et al., 2012). Recent results also indicate the two types of items can load on different factors (Liberali et al., 2012) and that statistical numeracy alone can capture all reliable variance associated with the CRT in some tasks involving highly educated individuals (Låg et al., 2013). Differences in item type are also responsible for differences in psychometric discrimination at different ranges (e.g., CRT items are harder and numeracy items are easier; Låg et al., 2013; Weller et al., 2013).

theory analysis identifying items with high levels of discriminability across a specified range of item difficulty, with a guessing parameter of zero.

The construct validity, reliability, and psychometric sensitivity of the Berlin Numeracy Test (BNT) was initially established in 21 studies ($n=5336$) of participants from 15 countries including assessments of diverse groups (e.g., US medical professionals, community samples, Mechanical Turk web-panels). Validation studies have since been extended to participants from 60 countries and include several patient and physician samples from all over the world (Garcia-Retamero, Cokely, & Ghazal, in preparation; Garcia-Retamero, Wicki, Cokely, & Hanson, 2014). Initial and subsequent analyses indicate that the test offers robust sensitivity, with optimal performance among those who have some college education.⁴ The test was also found to be the strongest predictor of understanding everyday risks (e.g., evaluating claims about products and treatments; interpreting forecasts), doubling the predictive power of other numeracy instruments and accounting for unique variance beyond other cognitive tests (e.g., cognitive reflection, working memory, intelligence).

2.1.3 Other Numeracy Assessments

Beyond the recent advances in the assessment of numeracy in the cognitive and decision sciences, there are also many other notable numeracy tests available for measurement of individual differences in numeracy. For example, as discussed in the

⁴ Sensitivity was poorest among students at an elite university in China. About 75% of those participants answered all questions correctly.

introduction, influential national and international surveys have been designed to assess and classify overall numeracy, as well as monitor changes in numeracy over time (e.g., NALS, NAALS, and PISA). Standardized intelligence assessments (e.g., Woodcock Johnson Test of Cognitive Abilities, WAIS) also commonly include tests of general quantitative abilities along with extensive behavioral and performance norms (e.g., standards for younger and older adults, etc.). Unfortunately, many of the numeracy tests developed for large scale assessment are not appropriate for educated samples. Other tests take a long time to administer and many others are subject to proprietary law and use restrictions. Perhaps most importantly for the current thesis, to my knowledge there is currently no modern IRT based brief numeracy test validated for assessment of numeracy sub-skills in educated samples. Accordingly, although there is a growing body of research linking numeracy and superior decision making, there is virtually no research available investigating the relative importance of numeracy components with respect to the underlying cognitive mechanisms that give rise to superior decision making.

Chapter 3: Numeracy and Decision Making

The theoretical construct of numeracy is multifactorial including (i) a practical understanding of numbers and mathematical procedures, and (ii) the skills necessary for effective problem solving and self-regulated learning (e.g., metacognition and thinking about thinking; Flavell, 1979; Garofalo & Lester, 1985; Cornoldi, 1997; see also Dunlosky & Metcalfe, 2009). To some extent, numeracy tests are thought to predict a wide range of behavior because they simultaneously assess both mathematical knowledge and the metacognitive processes involved in effective thinking (Halpern, 1998; Schoenfeld, 1992; Schraw, 1998; but for related theory in decision making see Baron, 1985, 2008; Baron, Badgio, & Gaskins, 1986; Stanovich, 2012; Stanovich, West, & Toplak, 2011; Toplak, West, & Stanovich, in press a, in press b). For example, in a series of previous studies examining two large samples of highly numerate, highly educated (e.g., 50% held advanced graduate degrees) professionals living in Europe, I with my colleagues (Ghazal et al., 2014) observed links between numeracy, confidence, deliberation, and superior judgment and decision making performance in key paradigmatic tasks (i.e., risky lotteries, intertemporal choices, medical risk interpretation, and judgment self-assessment). Because most participants were highly numerate the differences in decision performance were not likely attributable to differences in the availability of requisite mathematical skills. Nearly all participants were numerate enough to accurately calculate all expected values, discount rates, and relative proportions. Differences were also unlikely to reflect variation in levels of short-term motivation or task goals, as all participants volunteered and logged-on so they could test

their numeracy and decision making. Rather than differences in goals, motivation, or minimum mathematical understanding, the observed performance differences appeared to be more metacognitive in nature.⁵ Those participants who had a more accurate subjective sense of their judgment performance (i.e., estimated confidence) and those who tended to spend more time deliberating during decision making tended to perform better. While there are likely many other important metacognitive and numeracy-related skills at play (Peters, 2012; Peters, Meilleur, & Tompkins, in press; Reyna & Farley, 2006; Reyna et al., 2009), these data accord with previous research suggesting that deliberation and accurate self-monitoring often play central roles in domain-general superior judgment and decision making.⁶

3.1 Confidence & Deliberation

The relationship between confidence and superior judgment and decision making is well-established (Bruine de Bruin et al., 2007), as are the relations between confidence, numeracy, and intelligence (Stankov, 2000). Research indicates that one's subjective estimate of confidence tends to derive from two factors—i.e., self-consistency (e.g., how reliably and quickly a judgment comes to mind) and the breadth of information that comes to mind (Koriat, 2012; see also Pleskac & Busemeyer, 2010). Together with my

⁵ Appropriate cognitive representations, rather than explicit math skills, can play a crucial role in superior performance, as can be seen with the influence of simple visual aids that eliminate large performance differences between more and less numerate participants (Garcia-Retamero & Cokely, 2011, 2013, in press; see also Gigerenzer, Gaissmaier, Kruz-Mickle, Schwartz, Woloshin, 2007, and Peters et al., in press).

⁶ There are many theories about the causal mechanisms that give rise to the link between domain-general abilities and superior performance, as well as many compelling critiques of those theories (Kahneman, 2003, 2011; Reyna et al., 2009; Stanovich & West, 2000, 2008).

mentors (Ghazal et al., 2014), I was the first to demonstrate that confidence can partially mediate the relationship between numeracy and superior judgment. Given that the sample was highly numerate and the measurement scale was simplified (i.e., Likert rather than probability based), individual differences in observed calibration could not reasonably be attributed to differences in scale-comprehension (e.g., different people did not likely use the scale in systematically different non-interval ways). Of note, the relationship between confidence and performance was found to be curvilinear reflecting a robust unskilled yet unaware effect, such that those participants who had the lowest levels of numeracy also showed the poorest ability to assess their own skills (Ehrlinger & Dunning, 2003; Ehrlinger et al., 2008).

The confidence results reported in Ghazal, Cokely, & Garcia-Retamero (2014) accord with a variety of factor-analytic studies indicating that one's self-assessment via confidence judgments can operate as a domain-general skill that will be correlated with but also an independent predictor of general abilities and personality traits (Baker, 2010; Schraw, 2010; Stankov, 2000; Stankov & Lee, 2008). Previous findings also accord with metacognitive theory suggesting confidence tends to be useful specifically because it is instrumental in self-regulation—i.e., the monitoring and control of cognition (Nelson & Narens, 1990; see also Metcalfe & Finn, 2008). For example, Koriati and Goldsmith (1996) describe how confidence accumulates and then is checked against a criterion in order to decide what type of information will be output in from memory. Related studies of factors such as “feeling of correctness” show that confidence-type judgments predict differences in information search and elaboration. In addition to predicting performance

judgments about the correctness of one's answer, one's feeling of correctness tends to be related to "rethinking" times and the likelihood of changing one's initial answer during reasoning (Thompson, Prowse Turner, & Pennycook, 2011). These studies and others suggest that factors related to how one uses and assesses one subjective confidence may often be an essential mediating factor determining the extent to which one deliberates during judgment and decision making (e.g., how much evidence does one require in order to feel confident in one's decision?).

The links between deliberation and various types of superior cognitive performance are also well-established. Deliberation is related to and can even cause differences in domain-general cognitive abilities, such as intelligence and attentional control (Baron, 1978, 1985; Cokely, Kelley, & Gilchrist, 2006; Hertzog & Robinson, 2005; Stanovich, 2012). Deliberation is thought to be an essential component of rational thinking (e.g., the search-inference framework, reflectiveness and active open-minded thinking; Baron, 1985, 2008; Baron et al., 1986) and cognitive process tracing studies suggest that differences in deliberation are not likely to result primarily from differences in normative decision strategies.⁷ Consider the protocol analysis conducted by Cokely and Kelley (2009) examining deliberative processes in simple risky lotteries. Although a pilot study indicated that most college students could perform the required math (e.g., "what is 3% of 7000"), less than 5% of their sample calculated expected value during

⁷ For related experimental evidence see the study by Peters et al. (2006) showing that while numeracy is related to superior performance it is also predictably related to biases, reflecting the influence of heuristic processes (e.g., influenced by affective precision).

decision making. Analyses of formal decision models, reaction times, and retrospective memory reports indicated that the ability-to-performance relationship was fully mediated by large differences in heuristic-based deliberation and elaborative processing (Cokely & Kelley, 2009; see also Pachur & Galesic, 2013). Better risky decision making followed from differences in *how participants thought* about the decision (e.g., meaning-oriented elaborative processes such as imagining how the changes in wealth could affect one's life and how that might feel in contrast to others who treated the task as if it was just a game of chance; see also Reyna et al., 2009). Better risky decision making also followed from differences in *how much participants thought* about the decision (e.g., elaborating multiple reasons for each decision, imagining how various outcomes would feel, transforming probabilities, and reframing outcomes). Similar results have been found in other protocol analyses, eye-tracking studies, and memory analyses used to examine some medical and economic judgments and decisions (Barton et al., 2009; Woller-Carter et al., 2012). Protocol analysis also suggests that during move selection in chess, the systematic use of more deliberation tends to be associated with large performance advantages for both novices and experts (Moxley, Ericsson, Charness, & Krampe, 2012; see also Ericsson, Prietula, & Cokely, 2007).

It is important to note that theoretically deliberation, confidence, and performance likely reflect a host of early selection metacognitive processes (Cokely & Kelley, 2009) that may function somewhat independent of basic cognitive capacities (e.g., attentional control, intelligence, and executive functions). Research shows that individuals who score higher on domain-general cognitive ability measures often spend more time

preparing for tasks and also more elaborately process information, deliberately building richer cognitive representations in long-term memory in order to provide better monitoring and control during subsequent task performance (Baron, 1978, 1985; Cokely & Kelley, 2009; Cokely et al., 2006; Ericsson & Kintsch, 1995; Hertzog & Robinson, 2005; Sternberg, 1977; Vigneau, Caissie, & Bors, 2005). To further illustrate, consider an analogy. In manufacturing one can improve the quality of goods sent to market by (a) improving inputs (e.g., higher quality materials and plans), (b) improving outputs (e.g., careful inspection and repair), or (c) doing both. In the metacognition literature these quality control efforts are referred to in terms of (a) early selection versus (b) late correction processes (Jacoby, Kelley, & McElree, 1999; Jacoby, Shimizu, Daniels, & Rhodes, 2005). Early selection uses controlled processing (e.g., System 2) to generate goals, strategies, and mental contexts that qualitatively alter the output of automatic processes (e.g., System 1) before biased intuitions are generated (e.g., approaching the task more carefully). In contrast, late correction processes attempt to detect and repair (e.g., System 2) the output of faulty automatic processes (e.g., System 1), such as cognitive and decision biases, after one forms a biased impression.

Research suggests that when early selection metacognitive processes are recruited they tend to involve a number of cognitive strategies involving deliberation, heuristic search, and elaborative encoding in long-term memory. For example, participants who tend to make better decisions will tend to contextualize a problem by deeply thinking about the various aspects of the problem and its potential implications (e.g., What will happen to me? How would I feel about that?). They may reframe the problem by

considering different perspectives or alternative approaches (e.g., a 5% chance of dying is also like a 95% chance of survival). They will also search for disconfirming evidence and may be more likely to double-check their work. The cognitive processes required to use these strategies essentially involve elaborative encoding, a process known to be one of most powerful means of moving information from working memory into more enduring representations in long-term memory. As a result, the decision maker builds a more complex intuitive understanding of the problem, freeing-up limited attentional resources that can be used for cognitive monitoring and self-assessment (Cokely et al., 2006; Craik & Lockhart, 1972). In previous research my colleagues and I have speculated that such representations may be similar in important respects to those described in Fuzzy-Trace Theory as gist-based representations (e.g., a rich highly-contextualized but not necessarily precise representation of “meaning”; Reyna & Brainerd, 1991,2005; Reyna et al., 2009).

Those participants who rely on early selection metacognitive processes and elaborative encoding processes create a feedback loop such that confidence calibration can be improved because more detailed representations provide more diagnostic cues for accurate cognitive monitoring (Mitchum & Kelley, 2010). Note, again, however, that mere deliberation does not guarantee improved performance. Performance incentives that increase deliberation often fails to improve calibration or decision making performance because participants tend to search for evidence that confirms their current beliefs

(Koriat, Lichtenstein, & Fischhoff, 1980; see also Nickerson, 1998).⁸ Improving calibration typically requires either changing task structures or training with individualized feedback. This type of training can lead to nearly perfect calibration. However, confidence calibration will tend to be highly domain-specific unless training also focuses on transferable skills (Arkes, 1991).

3.2 Transfer and Identical Elements

Cognitive skills and abilities only generalize to the extent that similar elements of the skills are present on *training* and *transfer* tasks. Transfer requires shared elements (Woodworth & Thorndike, 1901; see also Blume, Ford, Baldwin, Huang, 2010). Many skills are highly domain-specific and so they are unrelated to performance outside a narrow band of expertise (e.g., surgical skill is not related to managerial decision making; Ericsson, Charness, Feltovich, & Hoffman, 2006; Ericsson et al., 2007). Numeracy is different.

In the modern world, mathematical concepts are ubiquitous: Numeracy is an essential component of risk literacy and scientific thinking (Bruine de Bruin & Bostrom, in press; Cokely et al., 2012; Gigerenzer 2002; 2012). Beyond the contributions of one's mathematical competence, numeracy tests have been found to predict superior judgment and decision making because they assess (i) heuristic-based deliberation and metacognition (e.g., Cokely & Kelley, 2009; Cokely et al., 2012; see also Stanovich,

⁸ See Cokely and Kelley (2009) for a more detailed discussion of deliberative early selection versus late correction cognitive control processes.

2012; reflectiveness, Baron, 1985), (ii) affective numerical intuition (Peters, 2012; Peters et al., 2006; Slovic, Finucane, Peters, MacGregor, 2002), and (iii) meaningful intuitive understanding (e.g., Reyna & Brainerd, 1991, 2005a; Reyna et al., 2009) among other factors (e.g., test anxiety; for a review see Cokely et al., 2012). More research is needed to investigate and model training and transfer across numeracy, metacognition, and decision tasks. What aspects of numeracy improve metacognition and how does training metacognition improve numeracy? More central to the current thesis: How are different aspects of numeracy (i.e., numeracy sub-skills) related to differences in decision making competency and risk literacy? In order to investigate these and many other theoretical questions, there is a need for simple, efficient testing technologies that quickly and accurately assess individual differences in numeracy and its sub-skills. We also need a better understanding of the how the various sub-skills relate to essential decision making biases and competencies. Investigations of the links between numeracy sub-skills and superior decision making may allow for a higher-fidelity theoretical understanding of underlying shared elements including key strategies, capacities, and skill-sets that give rise to transfer and domain-general differences in superior decision making.

Chapter 4: STUDY 1

Measuring Numeracy and Component Numeracy Skills

4.1 Introduction

Based on the reviewed numeracy frameworks (Gal et al., 2003; Ginsberg et al., 2006; National Assessment of Adult Literacy (NAAL), 2003) the goal of Study 1 was to develop a psychometrically sound, brief, and comprehensive test of adult numeracy and its four major components: operations, probability, geometry, and algebra. The test was developed using modern psychometric modeling and evaluation methods—i.e. Item Response Theory (IRT) (see Table 4-1 for sample items in each component).

Specifically, I conducted a multiple-phase iterative test development process involving (1) numeracy test item selection followed by (2) numeracy test item analysis and modeling. My goal was to create a test that is as brief as possible yet still provided simultaneous robust estimates of an individual's (i) overall numeracy and (ii) comparable specific numeracy sub-skills. To preview, I conducted a series of studies and analyses beginning with a pool of 100 candidate numeracy test items that were iteratively reduced from 100 items to 80, 40, and ultimately 20 items.

Table 4-1: *Sample items for each component; correct answer is in bold*

Components	Sample Items
Operations	The odometer of a new automobile functions improperly and registers only 2 miles for every 3 miles driven. If the odometer indicates 48 miles, how many miles has the automobile actually been driven? A) 144 B) 72 C) 64 D) 32 E) 24
Probability	Imagine that you are throwing 6 dice up in the air. What is the probability that all of them would land on even numbers? A) 1/432 B) 3/216 C) 1/64 D) 3/6 E) 1/216
Algebra	If $(3x + 2)(2x - 5) = ax^2 + kx + n$. What is the value of $a - n + k$? A) 5 B) 8 C) 9 D) 10 E) 11
Geometry	A, B, C, and D are points on a line, with D the midpoint of BC. The lengths of AB, AC and BC are 10, 2, and 12, respectively. What is the length of AD? A) 2 B) 4 C) 6 D) 10 E) 12

4.2 Test Construction

4.2.1 Item Selection

Test development began with an extensive review of existing measures of numeracy. The review included emphasis on the key numeracy components taught in mathematics education from grade 9th to 12th. Four numeracy sub-skills were selected for assessment based on the well-established existing frameworks for the construct of adult numeracy (Ginsberg et al., 2006). Math course books and SAT study guide books

were consulted for item generation and selection in the four content areas of numeracy, namely (1) Number Sense and Operations, (2) Statistics and Probability, (3) Geometry and Measurement, (4) Algebra and Functions. Initially a pool of 100 items was generated, with 25 items per numeracy component. Following careful review 80 items were selected to be included in Phase I of the test development.

4.2.2 Phase I: Preliminary Studies

In Phase I, data were collected from 30 undergraduate students from Michigan Technological University (for course credits) who completed an extensive numeracy test including all 80 items. Following standards for SAT exams, during the test session participants were allowed to use calculators. Students signed up for the study using the subject pool (SONNA) system, and were tested in the laboratory. Items were administered in a paper pencil format. The majority of the items were multiple choice items. A small proportion of the items required answering open ended questions.

The initial test performance and item analyses were conducted via SPSS because the sample size was too small for IRT ($n=30$). Difficulty, discrimination, and consistency parameters were calculated using descriptive analysis, item-total correlations, and reliability analysis, respectively. Based on this item analysis, 40 items were identified (10 items in each component) with the broadest range of difficulty (varying from .96 to .15 where higher value indicates easier item) and with desirable discrimination levels (item-total correlation $> .3$). The total scale (all 80 items) had a high reliability coefficient (cronbach alpha = .91) as did individual sub-scales for all four components (Operations =

.84, Probability= .79, Geometry= .84, and Algebra= .87). See Appendix A for the complete 80 item analysis.

4.2.3 Phase II: The 40 Item Test

4.2.3.1 Participants

In Phase II, data were collected from Michigan Technological University undergrad students ($n=135$) who were recruited through psychology subject pool (SONNA) system. The students signed up for the study and completed the test administered in a paper pencil format. Following standards for SAT exams, during the test session participants were allowed to use calculators. Seventy five percent of the total sample was male. The mean sample age was 20 years ($SD=1.4$).

Standards for sample sizes and power estimates in IRT depend on a number of factors including model complexity, item types, sample types and others. Although “There is no gold standard or magic number that can be proposed” for sample size (Morizot et al., p. 411), a general heuristics is that more power tends to provide more accurate estimation of the parameters, with diminishing marginal returns in very large samples (e.g., 500+). A general heuristic that is common practice recommends a sample size of about $n=100$ for a one parameter model and about $n=200$ for a two-parameter model, given dichotomous response formats (Moritoz, et al., 2007). A final sample size of $n=165$ including the pilot data ($n=30$) and new responses were used in Phase II.

4.2.3.2 Materials and Procedure

An optimal 40 item test (10 items per sub-skill) was identified via item analysis. Multiple choice items were scored as correct or incorrect (0, 1 coding, where 1 represent the correct answer). Following standard conventions, multiple choice items were scored dichotomously by treating one option as correct and treating all the distractors (other options) as equally wrong, which allows model development via IRT one, two, or three parameter logistic models (Kang & Cohen, 2007; NCES; Allen et al., 1997). Demographic information including participant age, sex, SAT/ACT scores, and SAT/ACT math scores and academic major was also collected (See Appendix E for all study material).

Data were analyzed using the statistical programming environment R. Specifically, the R package “ltm” (latent trait models) was used for all IRT analyses. Two and three-parameter models were analyzed for total score and subscale estimation. In IRT, a two-parameter logistic (2-PL) model measures the probability of answering an item correctly, given one’s theoretical ability level (i.e., a non-linear function of theoretical trait differences), determined via estimates of item difficulty and discrimination—i.e., how well it can discriminate between various levels of the underlying trait. The model can be expressed as Equation (4-1)

$$P(X=1|\theta) = 1 / 1 + e^{-1.7a(\theta-b)} \quad (4-1)$$

Where θ = the theoretical ability of a particular examinee

a = the discrimination parameter

b = the difficulty parameter

Difficulty and discrimination parameter values can then be used to identify items with optimal performance across a range of ability levels. Theoretically, item difficulty ranges from negative infinity to positive infinity, although practically the range is between -4 and $+4$. A difficulty parameter value of -4 indicates that the item is extremely easy whereas a difficulty level of 4 indicates that the item tends to be extremely hard. Item discrimination theoretically ranges from 0 to infinity, where lower values indicate a less discriminating item. Items with very low discrimination parameters cannot differentiate between examinees with low or high underlying trait levels, even when items are difficult (e.g., the probability that a person with higher ability levels will correctly answer the question does not differ much from the probability that someone with low ability levels will correctly answer the problem).

A three parameter logistic model (Birnbau, 1968) may also confer a number of benefits for test development, especially given dichotomous or multiple choice items. The third parameter that is modeled is called 'Pseudo guessing'. This parameter estimates and eliminates variance owing to examinees who answer moderate to hard items correctly purely as a function of guessing. Theoretically guessing parameters range from 0.0 to 1.0 , with typical ranges below 0.3 (Harris, 1989).

Data from model comparisons suggested that a two-parameter model provided a significantly improved fit over the more complex three-parameter model (See Table 4-2). Thus IRT two-parameter model was used for all analysis.

4.2.4 Data Analysis

4.2.4.1 Model Selection

The R package “ltm” (latent trait models) module was used for IRT analysis. One parameter (Rasch equivalent model), two-parameter, and three parameter logistic models were fitted to the data (all four subscales with 10 items each and full scale 40 items). As noted, analyses indicated that the two-parameter model was superior to the three parameter model based on their information statistics (see Table 4-2) including Akaike’s Information Criteria (AIC, Akaike, 1974). Smaller AIC values indicate better fit in terms of principle of parsimony (e.g., increases in model complexity can only be justified via substantive improvements in model fit; Lin & Dayton, 1997; Kang & Cohen, 2007).

Table 4-2: *Likelihood Ratio Table showing information each model is providing (for total numeracy score on 40 items test)*

	AIC	BIC	Log.lik	LR	df	P value
One-parameter	7340.3	7467.6	-3629.1			
Two-parameter	7254.3	7502.8	-3547.2	164.0	39	.001
Three- parameter	7265.9	7638.7	-3513.0	68.3	40	.03

Thus the two-parameter model was used for all item analyses across each of the four subscales. (See Appendix C for model comparison/selection for each subscale). Of note, the ‘ltm’ package uses Marginal Maximum Likelihood Estimation (MMLE) for

fitting the models and calculating information, consistent with general requirements for accurate and robust interpretation of results from the AIC criterion in model selection.

Table 4-3: *Item analysis, using IRT two-parameter model; showing full scale (40 items) and individual subscales (10 items each) for four components*

	Difficulty	Discrimination	P(x=1 z=0)
N1	-3.30	0.85	0.94
N2	0.48	1.08	0.37
N3	-0.64	1.70	0.75
N4	-1.02	1.29	0.79
N5	-0.59	1.97	0.76
N6	-0.42	1.42	0.65
N7	0.01	1.57	0.50
N8	-0.69	0.63	0.61
N9	1.08	0.81	0.29
N10	-0.75	1.17	0.71
P1	-0.46	1.48	0.66
P2	-0.67	0.81	0.63
P3	-2.49	1.19	0.95
P4	-1.16	1.11	0.78
P5	-0.54	1.48	0.69
P6	3.99	0.24	0.28
P7	-0.59	1.05	0.65
P8	0.52	0.73	0.41
P9	1.73	0.60	0.26
P10	-0.11	1.19	0.53
G1	-1.00	1.51	0.82
G2	0.01	1.48	0.50
G3	-0.95	1.58	0.82
G4	-0.19	0.57	0.53
G5	-2.58	0.59	0.82
G6	1.59	0.15	0.44
G7	0.39	1.35	0.37
G8	-0.61	1.24	0.68
G9	0.03	0.78	0.49
G10	-0.13	1.56	0.55
A1	-0.87	0.63	0.63
A2	-0.46	2.20	0.74
A3	-0.58	1.49	0.70
A4	0.36	1.53	0.37
A5	-1.02	1.53	0.83

A6	0.34	1.31	0.39
A7	0.80	1.72	0.20
A8	1.10	1.34	0.19
A9	1.33	1.08	0.19
A10	0.44	1.77	0.31

4.2.4.2 Item Calibration

All 40 items were analyzed using the 'ltm' two-parameter model. Analysis revealed that most of the items provided good discrimination among individuals at high and low ability levels. Overall, items reflected a desirable and broad range of difficulty (-3.3 to 3.9), they provided good discrimination (mostly above .65). However, there were few problematic items as well (discrimination index $<.65$; see Table 4-3 for all item analyses showing beta (difficulty) and alpha (discrimination) coefficients along with theoretical theta probability estimates for each item)

The long 40 item test showed a good distribution, providing 56% of the total information about low latent trait levels (-4 to 0), and 40% of the total information about high trait levels (0 to +4) (see Figure 4.1), with exceptional reliability (Cronbach alpha = .97).

Figure 4-1 shows item and test analysis; the first curve is Item Characteristic Curve (ICC) that gives you information about item parameters. What you hope to see here is a sigmoid curve with a range of difficulty and a lower asymptote (guessing close to 0 or .1). Second curve is Item Information Curve (ICC), it gives you information about the person parameter. What we want to see here are peaked distributions across a range of ability levels. More peaked distributions indicate more precise ability estimates.

Third curve is Test Information Function (TIF) curve; it tells you that the test is most useful, informative, and precise at its peaked point. Here, TIF curve is indicating that my test is very well-suited for a wide range of college educated individuals.

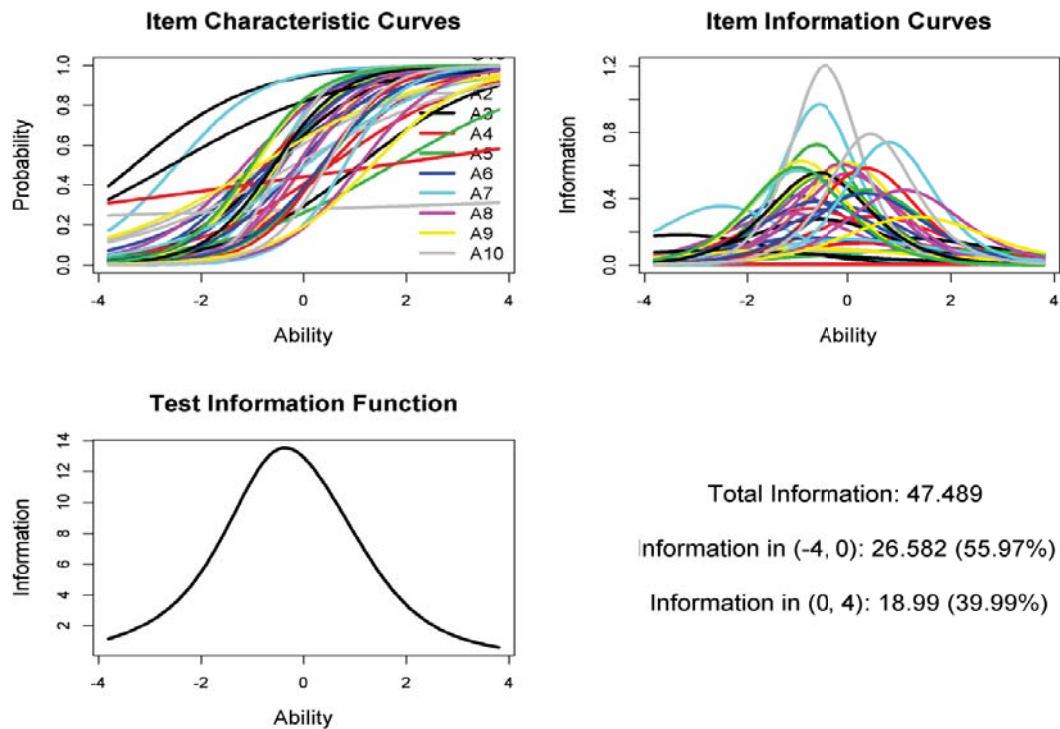


Figure 4.1. Item Information and test information curves for 40 item numeracy test

4.2.4.3 Item analysis of Individual subscales (10 items each)

I further conducted item analyses, using IRT two-parameter logistic model, on separate sub scales (10 items each). The goal was to estimate the parameter values for each component as an independent sub-scale (See Table 4-4, Table 4-5, Table 4-6, and

Table 4-7 for subscale item analysis, also see Figure 4.3 for information functions of the four subscales).

Table 4-4: *Item analysis for sub-scale “Operations” (10 items)*

Number	Difficulty	Discrimination	p/z_N
1	-2.4	1.3	0.96
2	0.43	1.2	0.37
3	-0.59	2.0	0.76
4	-1.00	1.4	0.80
5	-0.55	2.2	0.77
6	-0.41	1.5	0.65
7	0.01	2.2	0.5
8	-0.67	0.7	0.61
9	1.2	0.7	0.30
10	-0.82	1.0	0.70

Table 4-5: *Item analysis for sub-scale “Probability” (10 items)*

Prob.	Difficulty	Discrimination	p/z_P
1	-0.38	2.2	0.70
2	-0.69	0.8	0.63
3	-2.0	1.8	0.97
4	-1.2	1.1	0.78
5	-0.46	2.0	0.72
6	3.7	0.39	0.26
7	-0.45	1.6	0.68
8	0.34	1.3	0.39
9	1.2	0.9	0.25
10	-0.1	1.2	0.54

Table 4-6: *Item analysis for sub-scale “Geometry” (10 items)*

Geom.	Difficulty	Discrimination	p/z_G
1	-1.00	1.6	0.83
2	0.01	1.7	0.49
3	-0.85	2.0	0.84
4	-0.23	0.5	0.53
5	-2.00	0.8	0.83
6	1.34	0.2	0.44
7	0.38	1.4	0.37
8	-0.57	1.4	0.69
9	-0.02	0.9	0.50
10	-0.12	1.7	0.55

Table 4-7: *Item analysis for sub-scale “Algebra” (10 items)*

Algebra	Difficulty	Discrimination	p/z_A
1	-0.61	1.0	0.65
2	-0.43	2.5	0.75
3	-0.46	2.6	0.77
4	0.37	1.4	0.37
5	-0.82	2.6	0.89
6	0.35	1.2	0.40
7	0.82	1.5	0.22
8	0.90	1.9	0.16
9	1.20	1.3	0.19
10	0.42	1.9	0.31

4.2.5 Phase III: A 20 Item Test

Based on the item analysis of all 40 items, ideal items were identified (with range of difficulty, good discrimination and probability estimates for each item) providing a shorter test of overall numeracy including its four major components—controlling for

potential test bias as a result of differential test sensitivity. Identification of the best-suited items primarily emphasized theoretical theta ability estimates (i.e., probability estimates $P(x=1|z=0)$ or the probability of a correct response for the i th item for an average ability individual). This provided a principled way to select the final items because the measurement scale for the ability/theta estimates is theoretically independent of the items, which makes theta more useful when comparing different tests of same ability (Lord, 1980). This approach is known to result in more efficient test designs. Final analysis used the 'ltm' package to transform the parameter estimates to probability estimates with a function called 'coef()' that uses MMLE.

The final cumulated numeracy test included 20 items with 5 items in each subscale. Again, items were selected such that the psychometric sensitivity across all four components remained consistent and comparable in terms of probability and parameter estimates (see Table 4-8 for all 20 items analysis).

Table 4-8: *Item analysis for final 20 items with four subscales*

Items	Difficulty	Discrimination	P/Z ($x=1 z=0$)
Operations			
N4	-1.1	1.2	.78
N6	-.4	1.6	.65
N7	.00	2.7	.50
N2	.4	1.2	.37
N9	1.3	.65	.31
Probability			
P4	-1.3	1.1	.78
P5	-.43	2.5	.74
P1	-.3	2.1	.69
P10	.1	1.1	.53
P8	.34	1.3	.37
Geometry			
G1	-1.0	1.8	.83
G8	-.62	1.2	.68
G10	-.1	1.7	.55
G2	.00	1.7	.49
G7	.38	1.3	.37
Algebra			
A3	-.46	2.8	.73
A1	-.61	1.0	.65
A6	.36	1.1	.40
A4	.40	1.2	.37
A10	.45	1.7	.32

Theoretically, each sub-skill estimates one's sub-skill level across an equivalent range of psychometric sensitivity (i.e., relative to the same average numerical ability and ability range). To the extent that an individual shows high or low test scores on any of the final 5 item tests, differences should reflect differences in theoretical sub-skill levels not differences in sub-test difficulty or discriminability.

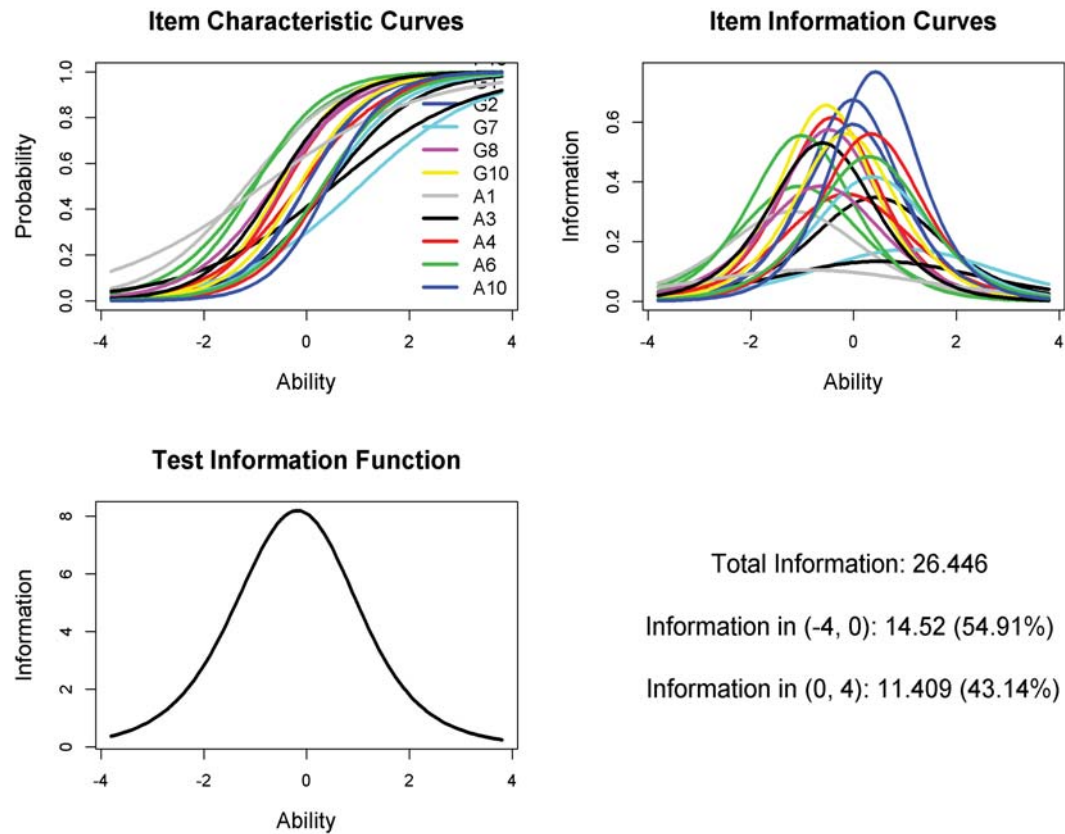


Figure 4.2. Item and Test Information Function for brief 20 items numeracy test

The new short test of numeracy (20 items) showed a good distribution which provided 55% of the total information for low latent trait levels (-4 to 0), and 43% of the total information for high trait levels (0 to +4) (see Figure 4.2), with good reliability (Cronbach alpha = .86). The short comprehensive test was also a remarkably strong predictor of numeracy test scores on the longer 40 items test ($r = .97$). Sub-skill inter-correlations showed robust positive associations between the four subscales (ranging between .47 to .58), without indicating perfect co-linearity (i.e., sub-scale performance tended to be correlated but also showed considerable dissociation wherein a large

proportion of the reliable differences in performance variance should be expected to be independent across sub-skills).

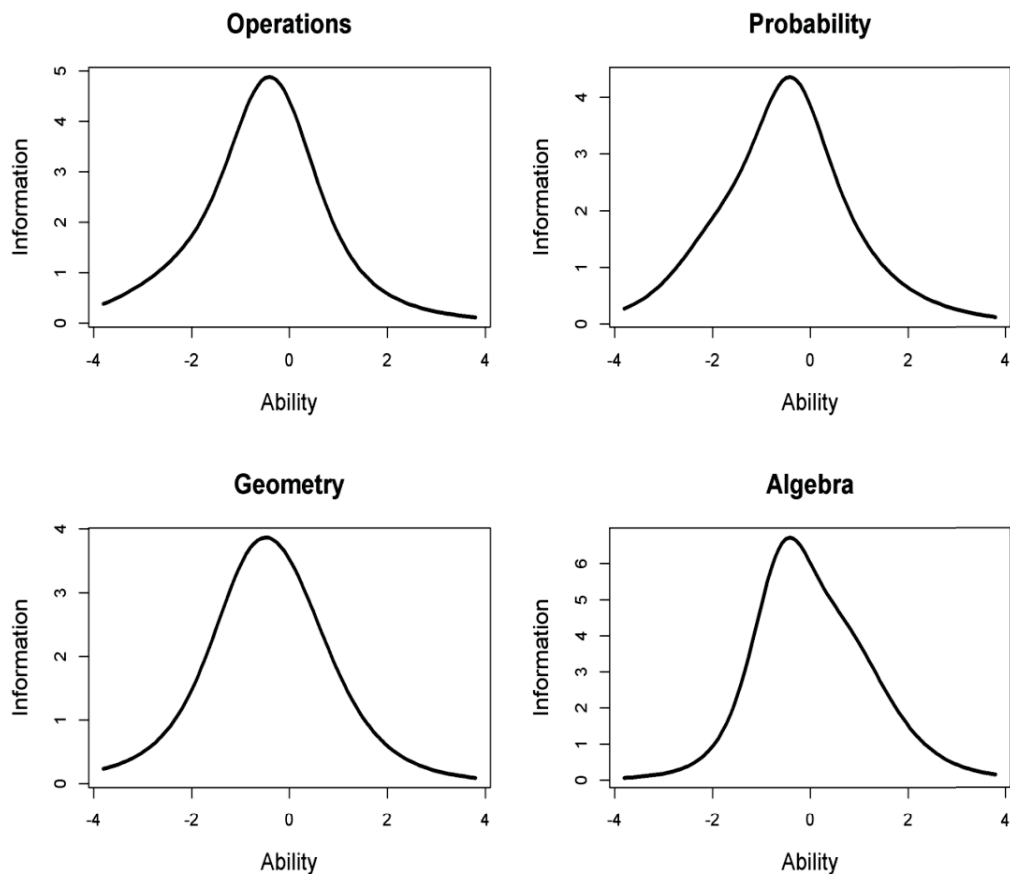


Figure 4.3. Test Information Functions for four subscales (10 items each)

The test information function (TIF) curve for new brief numeracy test shows that it is providing a good estimate of a wide range of ability levels. The peak of the curve is close to average ($\theta = 0$) ability level (see Figure 4.2) indicating that the new test is providing most information and is most discriminating at an average ability level, a very desirable property for a psychometric test. The specific item information curves also indicate that most items provide maximal information between -2 to +2 standard

deviations (i.e., high sensitivity for about 95% of all test takers), another desirable finding.

4.2.5.1 Subscales Analysis (5 items per scale)

I next analyzed the data evaluating each numeracy subscale (i.e., numeracy components) separately following the same procedure I did in phase II for 40 items test. First, IRT item analysis was performed on the 5 *operation* items. Difficulty, discrimination, and probability estimates were analyzed. Difficulty parameters for all five items ranged from -1.1 to 1.3, and discrimination parameter ranged from .65 to 2.7. This set of five items provided 53% of the total information for low (0 to -4) latent trait levels, and 45% of the total information for high latent trait levels, a relatively good balance. The reliability coefficient for the five operations items was fair-to-good at .60.⁹

⁹ Shorter tests will always show attenuation of reliability because they attempt to assess a wider range of skill with a few items.

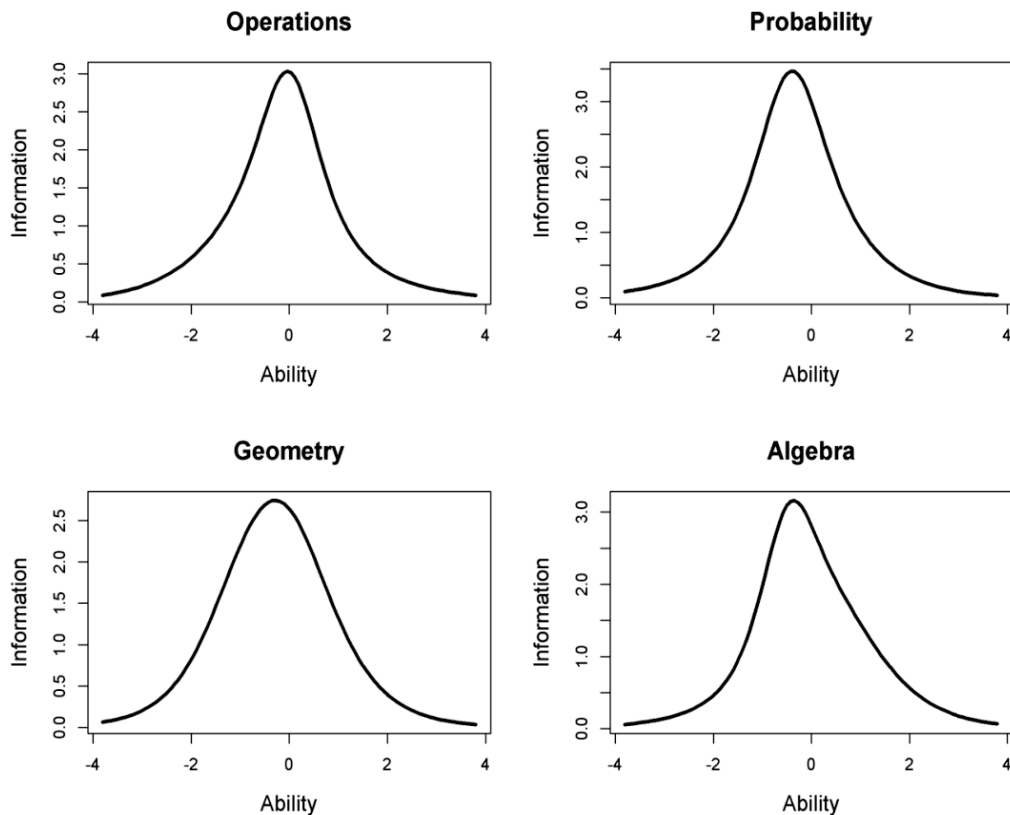


Figure 4.4. Test Information Functions for four subscales (5 items each)

I next applied the same process used to analyze the operations subscale to each of the other three subscales (5 items each), including *Probability*, *Geometry* and *Algebra*. (See Appendix B for comprehensive IRT Item analyses of all four subscales). The test information functions for all four subscales showed desirable, nearly identical distributions peaked near the average ability level ($\theta=0$). These analyses indicate that each subscale should provide the maximum information and discrimination among individuals with near-average ability levels (i.e., for most people) (see Figure 4.4). Across the four subscales other indices of psychometric parameters were highly-

comparable in terms of their difficulty, discrimination, and particularly in their probability estimates (theta). Subscale performance ranged at the low-end from probability of .31-.37 and at the high-end from probabilities of .73-.83 (see Table 4-9).

Table 4-9: *Item statistics for all four subscales*

Subscales	Difficulty	Discrim	Probab.	Info_ low	Info_ High	Cron bach
Operations	-1.1– 1.3	.65 – 2.7	.31 – .78	53%	45%	.60
Probability	-1.3 – .34	1.1 – 2.5	.37 – .78	63%	35%	.65
Geometry	-1.0 – .38	1.3 – 1.8	.37 – .83	59%	40%	.65
Algebra	-.46 – .45	1.0 – 2.8	.32 – .73	54%	45%	.64

In my sample ($n=165$), the mean score on new 20 items numeracy test was found to be 10.8 ($sd= 5.0$), i.e. close to median. Also, mean and standard deviation for the four subscales was comparable (see Table 4-10), indicating that four subscales are not very different in terms of difficulty.

Table 4-10: *Raw score, means, and standard deviations on four subscales*

Components	Mean	<i>Sd.</i>
Operations	2.6	1.5
Probability	2.9	1.6
Geometry	2.8	1.6
Algebra	2.5	1.5

4.2.5.2 Demographic analysis

No significant difference was found between performance of males and females on total numeracy score ($t(132) = 1.0, p = .33$). College major was a significant predictor of performance on total numeracy test scores. Participants from engineering majors did significantly better on total numeracy and on all four sub scales as compared to participants with Biological/Life-Science majors and others ($F(2, 135) = 11.24, p < .01$). The brief comprehensive numeracy test total scores were also found to be strongly related to SAT reported scores ($r = .60, p < .001$) and SAT math scores ($r = .64, p < .001$), providing additional evidence of convergent test validity. Time spent working on the test ($r = .42, p < .001$) was a significant predictor of performance (i.e., those who spend more time did much better). Age was unrelated to performance, reflecting our restricted range of participant ages ($r = .09, p = .90$).

4.2.6 Summary of the Study 1

Using a multiphase iterative test development procedure, I developed an optimized brief comprehensive Berlin Numeracy Component Test (BNT-C) for use with college educated samples. The test is based on a two-parameter logistic IRT model that identified the 20 test items that allow simultaneous assessment of one's overall numeracy level as well as differences in one's numeracy sub-skills. The brief test was found to offer desirable characteristics across the range of essential psychometric properties (e.g., similar levels of difficulty, discriminability, and sensitivity).

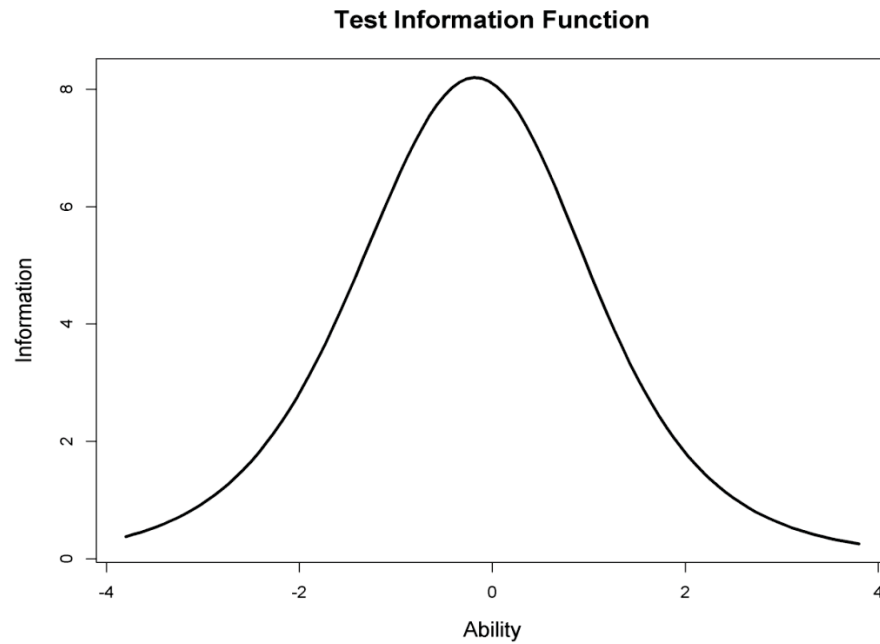


Figure 4.5. TIF for brief comprehensive test of numeracy (20 items)

The Test information function for the new scale shows that the test provides considerable information about an individual's overall skill level, with maximum information and discrimination close to the theoretically ideal of theta level = 0.

The brief comprehensive test (20 items) provided more discriminating items, and more comparable subscales in terms of psychometric properties as compared to long 40 item numeracy test (see Figure 4.6 and Figure 4.7 for comparison of the two long and short tests).

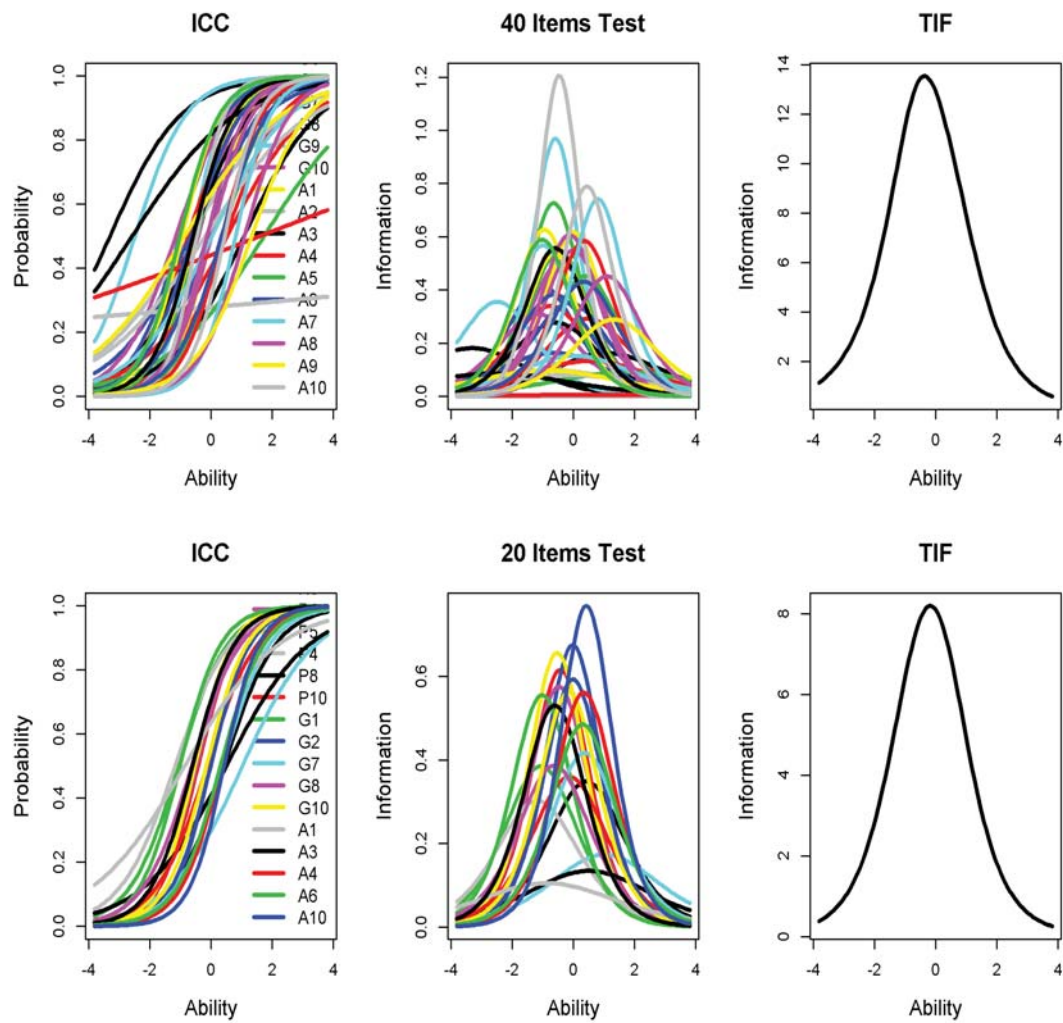


Figure 4.6. Item and Test Information Functions for long 40 items and brief 20 items numeracy tests

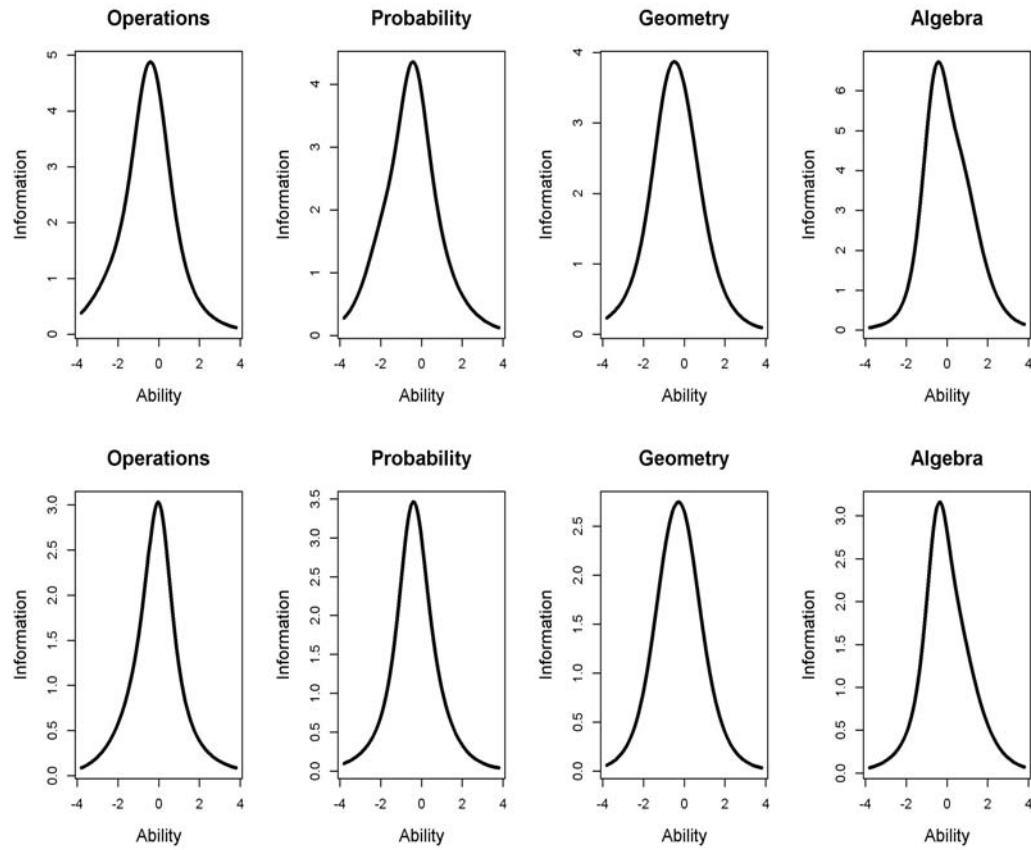


Figure 4.7. Test information Functions for four subscales both for long 40 items and brief 20 items numeracy tests

Chapter 5: STUDY 2

Numeracy Components and Decision Making Skill

5.1 Introduction

For more than a century, decision making under risk has been an area of considerable interest in the decision sciences, resulting in important empirical findings and the refinement of integrative normative theories (for review see Edwards, 1961; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Human decision making is complex and does not always follow normative prescriptions of rational choice. Nevertheless, individual differences in cognitive abilities and skills predict normatively superior decision making (Cokely & Kelly, 2009; Fredrick, 2005; Ghazal et al., 2014; Peters & Levin, 2008; Stanovich & West, 2000). Accurate assessment of these individual differences is crucial for better understanding of the relationship between these skills and superior decision making, and for modeling and mapping the nature of underlying cognitive abilities.

There are now a number of psychological tests and tasks available that are relatively well-understood and known to be important predictors of general decision making behavior, risk literacy, decision competency, health outcomes, financial outcomes, relationship outcomes, etc. (Baker, Williams, Parker, Gazmararian, & Nurss, 1999; Bruine de Bruin et al., 2007; Cokely et al., 2012; Kahneman & Tversky, 1979). A growing body of evidence suggests that one's numeracy tends to be among the strongest

single domain-general predictors of superior decision making. Research also indicates that the predictive power of numeracy tends to be independent and often is much stronger than that of general fluid intelligence, attentional control capacities, personality variables, and cognitive styles (Cokely et al., 2012; Låg et al., 2012; Lindberg, & Friborg, 2013; Schapira et al., 2012; Weller, et al., 2012).

To my knowledge, there is no study that has systematically assessed the co-variation and unique exploratory power of overall numeracy compared with the unique exploratory power of specific numeracy sub-skills in decision making. In part, this may reflect the fact that efficient tools like my newly developed brief numeracy test are not available or else are not sufficiently validated. To the extent that we can quickly, accurately, and robustly assess individual differences in one's general numeracy and numeracy component sub-skills, a host of important and interesting theoretical questions can be addressed. For example, is risk literacy more strongly related to differences in one's understanding of key concepts in statistics? What types of mathematical skills are associated with the ability to avoid overconfidence and ratio bias effects? Do subcomponent skills interact, in the statistical sense, to inoculate decision makers from costly common biases?

The goal of the study was to quantitatively model and systematically estimate the extent to which individual differences in numeracy and its sub-skills are related to decision making competency, risk literacy, decisions outcomes, and risky choice. Specifically, my goal was to assess predictive or concurrent validity of a variety of

numeracy assessments across a wide and representative range of theoretically crucial decision making tasks, including:

- a) The Adult Decision Making Competence Test (ADMC)
- b) The Decision Outcome Inventory (DOI)
- c) Paradigmatic Risky Choices (lotteries, intertemporal choice)
- d) Ecological Risk Literacy (e.g., direct to consumer advertising)
- e) Denominator neglect and ratio bias

In addition to the newly developed Berlin Numeracy Components Test (BNT-C), other numeracy, intelligence, and cognitive style tests were also investigated. Additional tests included the Berlin Numeracy Test (Cokely et al., 2012), the Numeracy Test by Weller et al. (2013), the NUMi test (Shapira et al., 2012), the short form of the Raven's advanced progressive matrices for assessment of fluid intelligence, and an assessment of cognitive impulsivity known as the Cognitive Reflection Test (Frederick, 2005). The battery of numeracy tests represents the currently best available numeracy instruments used in the decision sciences. One specific goal was to estimate the relative overall and unique predictive power of each numeracy component and numeracy assessment, with respect to essential decision making skills and tasks (e.g., interpreting information about loans, avoiding overconfidence, etc.). The design also allowed for a component analysis of the nature of the relations between numeracy and other general cognitive ability and cognitive style tests (e.g., is the Berlin Numeracy Test primarily a test of probabilistic numeracy skill or is it a mixture? Which types of math skills are more strongly related to

fluid intelligence and which types are more likely to predict differences in reflective thinking styles?).

5.2 General hypotheses

I hypothesize that basic attributes, including reliability, factor structure, information function and performance on four subscales would be consistent with the findings of Study 1 (college educated samples).

- 1) I hypothesize a strong positive relationship between the new brief comprehensive Berlin Numeracy Component Test (BNT-C), the BNT and other numeracy tests available (Shwartz, Lipkus, Weller, NUMi).
- 2) I hypothesize that BNT will be more strongly related to the probability component of numeracy as compared to other three components, although higher-order interactions are also likely (e.g., one needs a sufficient level of skill with probability theory before the benefits of skill in algebra may interact with test performance).
- 3) I hypothesize that the probability component of numeracy will provide the strongest predictive power for risk literacy tasks (e.g., interpreting loans, selecting lotteries, making intertemporal choices, avoiding denominator neglect) compared to other three components because risk is fundamentally a ratio concept that requires consideration of reference classes and ratio relations.
- 4) I predict that one's overall numeracy level will be robustly related to differences in all types of decision making competency, although numeracy subscales will differentially predict various decision competency factors (e.g., framing effects may

- be more strongly associated with algebra because abstract thinking and symbol manipulations may be cognitively similar to the process of converting from one frame to other).
- 5) In contrast to dominant theory, I hypothesize that overall numeracy will be the stronger predictor for superior decision skills as compared to sub-skills, because multi-colinearity of the components will increase the error term.
 - 6) I hypothesize that overall numeracy will show unique predictive power for all four criteria measures over and above the other predictors such as intelligence, CRT, and personality.
 - 7) I predict that overall performance on the numeracy tests will be essentially uncorrelated with personality (discriminant validity) although math anxiety may attenuate predictive relations (e.g., anxiety may reduce motivation to engage in elaborative encoding of numerical information, and those with no anxiety may instead experience higher levels of affective meaning from numbers which could inspire further encoding).

5.3 Method

5.3.1 Participants

Data were collected from Michigan Tech undergraduate students in two phases. In phase I, 216 students signed up for the study through the psychology subject pool system, and completed the experiment in lab. In phase II, they all were sent a follow-up link for study part II, however only 160 students responded to that follow up link. After deleting

the incomplete data points, along with those who didn't do the study seriously (10 participants wrote in the comment box that they didn't do it seriously), and those who completed the study in less than 5 minutes (26 participants), the final data set used for analysis consisted of 124 participants. Total average time for both Part I and Part II was about 2.5 hours.

5.3.2 Material

- a) **Adult Decision Making Competence (ADMC):** The “Adult Decision Making Competence” test (Bruine de Bruin et al., 2007) includes a battery of seven categories of decision tasks, namely; Resistance to framing, Recognized social norms, Under/Over confidence, Applying decision rules, Consistency in risk perception, Resistance to sunk cost, and Path independence. Below is the brief description of each sub-scale
- i) *Resistance to Framing:* measures whether value assessment is effected by irrelevant variations in how the problem is presented. There are seven risky choice framing tasks with formally equivalent gain and loss versions of decision problems. Additionally there are seven *attribute framing* tasks that ask the participants to rate positively or negatively described versions of seven normatively equivalent events. Performance is judged by the absolute difference between the ratings of gain and loss versions of each item.
- ii) *Recognized Social Norms:* measures how well participant assess peer social norms e.g. “Is it sometimes OK to engage in each of 16 undesirable behaviors such as stealing under certain circumstances.” A second set of questions measures the

- participants perceived social norms for these behaviors (e.g. Out of 100 people of your age, how many of you would say it is sometimes OK to steal under certain circumstances)
- iii) *Under/Over Confidence*: This measure presents 34 true false statements, and after each statement participants are asked how sure they are of their answer on a scale ranging from 50% to 100% confident. Overall score reflects the reversed absolute discrepancy from perfect calibration.
 - iv) *Applying decision Rules*: task specifically evaluates the ability to apply decision rules of varying complexity. Participants are asked to choose between DVD players with different numeric ratings on multiple features and to select one or more options following a specific decision rule.
 - v) *Consistency in Risk Perception*: task ask participants to judge the probability of various events (e.g. getting into a car accident, driving accident-free) happening in different time frames (e.g. in the next year, in the next 5 years). Performance is then assessed by evaluating the congruency of the participants' judgments with basic probability principles, which is deemed an important prerequisite for making decisions involving risk.
 - vi) *Resistance to Sunk Cost*: task measures the capacity to discontinue financial and time investments that are no longer paying off. Performance is measured by average ratings across 10 items.
 - vii) *Path Independence*: Presents pairs posing normatively equivalent choices between gambles. Choice should not be affected by normatively irrelevant changes in how

they are played. Performance is measured by the percentage of consistent responses regardless of whether options are presented as a single stage or two stage gambles.

b) Ecological Risk Literacy Tasks

Via a systematic and extensive review, I identified a number of tasks taken directly from the natural ecology that theoretically should require various levels of risk literacy to interpret (e.g., direct-to-consumer advertising, information about loan terms). For a more detailed account of the process of sampling and item identification, see Appendix D.

c) Decision making Outcome Inventory (DOI)

The DOI, developed and introduced by Bruine de Bruin et al. (2007), is a self-report measure of decision making that assesses successful decision making as evidenced by avoiding negative decision outcomes. The inventory includes 41 negative decision outcomes, with 35 of the outcomes preceded by a question asking if participants have made decisions that would make the outcome possible (e.g., Invested in stocks?... Lost more than \$1000 on stock investment). The DOI provides a weighted number of negative outcomes that respondents have avoided out of those they had opportunity to experience it.

d) Paradigmatic Risky Decision Tasks

I included risky choice tasks (hypothetical gambles and lotteries) and time preference tasks (intertemporal choice) that are known to be systematically related to a wide range of “real world” risky decision making behaviors, including those in

finance, law, and health. After reviewing the literature, I carefully selected items that vary in range of expected value magnitude and complexity, so the test provides better sensitivity across different levels of ability. For risky lotteries, I selected the items that vary in terms of complexity (3 vs 2 option lotteries), affect (rich vs poor affect) and magnitude (varying the amount of gain or loss). I selected items from the standard studies by Cokely et al., 2009; Fredrick, 2005; Kahneman & Tversky, 1979; Pachur & Galesic, 2012; Peters & Levin, 2008 (see Appendix F).

For inter-temporal choice items, I included items with varying time intervals and reward amounts. I included some complex items to see discounting by delay and discounting by intervals (subadditivity). Items were taken mainly from Fredrick, 2005 and Read, 2006. (See Appendix F).

For denominator neglect and ratio bias, I used hypothetical risky decision scenarios taken from Alonso & Fernandez-Berrocal, 2003; Denes-Raj & Epstein, 1994; Okan, Garcia-Retamero, Cokely, & Maldando, 2012; Yamgishi, 1997. The scenarios were carefully selected with varying difficulty levels.

Overall, I selected the items in the above said three paradigmatic risky tasks (i.e. to provide more psychometric sensitivity in a comparable way (See appendix F for all measures).

e) Other measures:

I also included the best available measures (potential predictors) to see how BNT-C and its four components are predicting superior decision making differently as compared to these potential predictors, as well as how BNT-C and its four

components are related to these predictors/ability measures (convergent validity). I have included the following measures:

- i) *The Cognitive Reflection Test* (Fredrick, 2005)
- ii) *Global Personality Traits*: A ten item assessment of a Big Five personality traits (Gosling, Rentfrow, and Swann, 2003)
- iii) *Fluid Intelligence*: The short form of the Raven's Advance Progressive Matrices__ a 12 items test of Fluid Intelligence (Bors and Stokes, 1998)
- iv) *Math Anxiety*: Mathematics Anxiety Rating Scale-Revised (MARS-R) (Brewer, 2011).
- v) *The Abbreviated Numeracy Scale*: A Rasch analysis optimized for the general population (Weller et al., 2012)
- vi) *Numeracy Understanding in Medicine Instrument* (NUMI; Schapira et al., 2012), a test designed to assess numeracy skills that are essential parts of health decision making.
- vii) *The Schwartz et al. 3 items test of statistical numeracy* (Schwartz et al., 1997)
- viii) *The Lipkus et al. 11 items test of statistical numeracy* (Lipkus et al., 2001)
- ix) *The Berlin Numeracy Test, a brief test of statistical numeracy and risk literacy* (Cokely et al., 2012)
- x) Demographic information on age, sex, study major, SAT/ACT scores, and SAT/ACT scores on math section was also collected.

5.3.3 Procedure

Experiments were designed using online survey software (unipark.de). Students signed up for the study 2 part I through MTU Psychology subject Pool (SONA) system. They came to lab to take the experiment, where they were provided an online link to complete the experiment which included all numeracy measures (except BNT-C), ADMC and DOI. Within 15 days¹⁰ of completion of part I, they were sent a follow up link (part II) that included new test BNT-C, along with paradigmatic and ecological risky decision tasks. Total average testing time for both the parts was about 2.5 hours.

5.4 Data Analysis

The mean age of the participants ($n=124$) was 19 years ($SD= 1.7$) and 66% of the sample consisted of male students. Majority of the students (49%) were engineering students. No significant difference of gender ($t(113) = .14, P = .89$) or major ($t(121) = .10, P = .94$) was found on performance of BNT- C, as well as overall Superior Decision Making Skill (SDMS) ($t(113) = .13, P = .90$), ($t(121) = 1.6, P = .12$) respectively. Reported SAT scores were highly correlated with BNT ($r(106) = .49, p > .001$), BNT-C ($r(106) = .51, p < .001$), and intelligence ($r(106) = .35, p > .001$).

5.4.1 Psychometric Analysis

First, to test the basic attributes of the new BNT-C test, item analysis was conducted following the same procedure done in study 1. (i.e. Item Response Theory 2PL

¹⁰ The split time design has proven useful in the past, as it limits the amount of time participants are required to concentrate at any given time and also provides the opportunity to assess both concurrent and predictive validity of various measures

model). It was predicted that basic attributes of the new BNT-C, including reliability, factor structure, information function and performance distributions of four subscales would be consistent with the previous study.

Descriptive analyses were conducted for the 20 items test. Results indicated that performance was consistent across the four subscales (see Table 5-1), and was comparable to what we found in study 1, however, this time performance on operations was slightly lower as compared to study 1 (mean 2.3 as compared to 2.6).

Table 5-1: Descriptive Statistics for the four subscales and full BNT-C

	N	Minimum	Maximum	Mean	SD
Operations	124	.00	5.00	2.3	1.3
Probability	124	.00	5.00	2.7	1.7
Geometry	124	.00	5.00	2.4	1.4
Algebra	124	.00	5.00	2.4	1.5
BNT-C	124	2.00	19.00	9.8	4.6

Further Item Analysis was conducted following the Item Response Theory two parameter logistic model, using ltm R package. Overall, all four components subscales provided almost equal amount of information, with maximum information around the mean, as observed in study 1. However, psychometric sensitivity of the four components

was not similar and deviated slightly from that observed in study 1¹¹; operations component was found to have more sensitivity as compared to the algebra component (see Table 5-2 for Item parameters and probability estimates).

Table 5-2: *Item analysis for the Berlin Numeracy Component Test (BNT-C)*

Operations	Difficulty	Discrimination	P (x=1/z=0)
BNT-C-1	-2.42	0.67	0.83
BNT-C-2	-0.02	4.9	0.53
BNT-C-3	0.62	0.38	0.44
BNT-C-4	0.87	1.7	0.18
BNT-C-5	1.7	0.74	0.22
Probability			
BNT-C-6	-0.85	2.05	0.85
BNT-C-7	-0.23	1.7	0.60
BNT-C-8	-0.12	1.5	0.55
BNT-C-9	0.13	4.7	0.34
BNT-C-10	0.23	1.9	0.39
Geometry			
BNT-C-11	-1.15	0.63	0.67
BNT-C-12	-0.47	1.15	0.63
BNT-C-13	-0.45	0.62	0.57
BNT-C-14	0.7	1.7	0.23
BNT-C-15	0.93	1.5	0.19
Algebra			
BNT-C-16	-0.5	0.92	0.61
BNT-C-17	-0.07	3.3	0.56

¹¹ I developed and used parallel items of what I use in study 1 (because of copyright issues), the differences observed in psychometric properties might be contributed to that factor.

BNT-C-18	-0.03	1.2	0.51
BNT-C-19	0.44	0.75	0.42
BNT-C-20	0.63	1.35	0.29

BNT-C test (20 items) showed similar Test Information Functions (TIF) in the two studies; the amount of information provided was almost same (26% vs 25%) with maximum information at average ability level (both curves peaked at the center, close to average ($\theta = 0$) ability level). However, the distribution in study 2 was found to be more leptokurtic, as compared to the normal distribution in study 1. Nevertheless, both distributions have their peaked point in the center (near $\theta = 0$), indicating that test is most suitable, informative, and discriminating for average ability college educated samples (see Figure 5.1 and Figure 5.2 for TIF of the two studies).

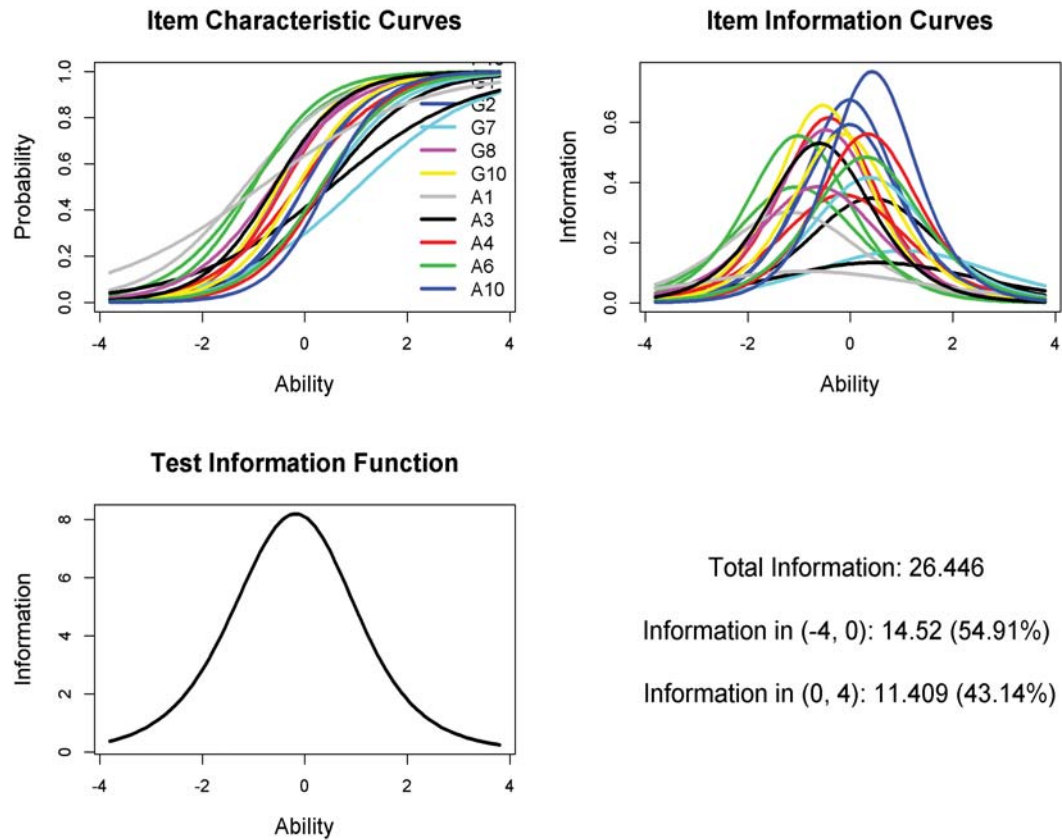


Figure 5.1. Item and Test Information Function for brief 20 items numeracy test (study 1)

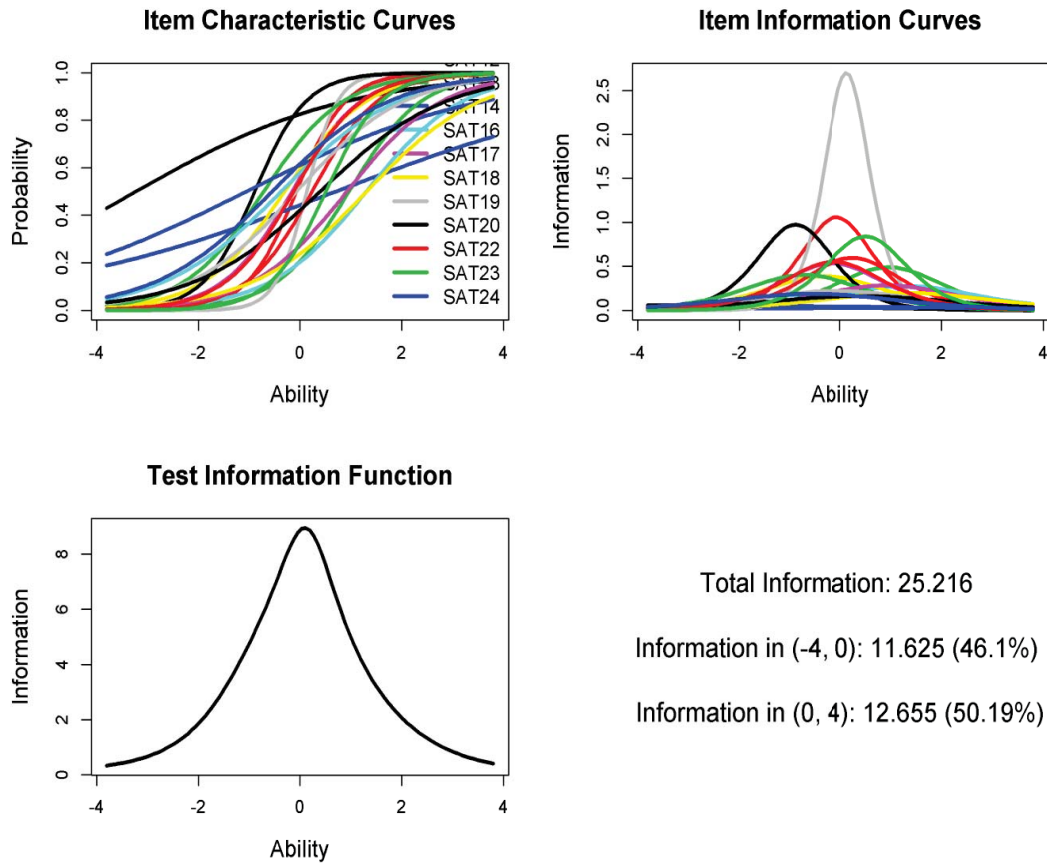


Figure 5.2. Item and Test Information Function for brief 20 items numeracy Test (BNT-C, Study 2)

Subscales analysis revealed that the four components provided a similar amount of total information, however, one notable difference found was in psychometric sensitivity: operations and probability components provided more sensitivity as compared to geometry and algebra. This difference may be attributed to use of parallel items or small sample size (see Figure 5-3 for comparison of subscale distributions from the two studies).

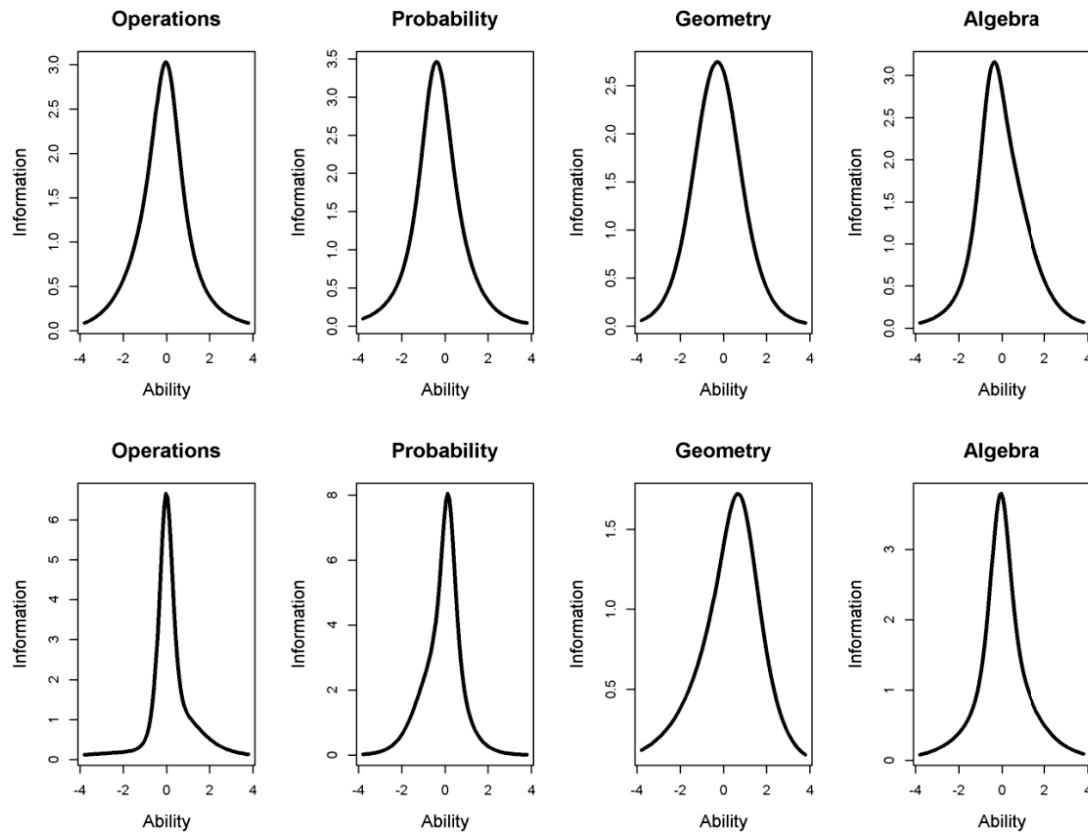


Figure 5.3. Test information functions of four numeracy components from study 1 (first row) and study 2 (second row)

5.4.1.1 Reliability Analysis

Cronbach alpha for the new BNT-C (20 items) test was = .83 (very comparable to .86 from study 1). Individual component subscales showed anticipated moderate alphas (operation, .48; probability .75; geometry .50; and algebra; .58). Cronbach alpha is mean of split half correlations (Cronbach. 1951), so for short length tests (5 items in each subscale) alpha will not necessarily be a good index of internal reliability. This is specifically because shorter tests show attenuation of reliability because they attempt to assess wider range of skills with fewer items.

5.4.1.2 Factor Analysis:

Exploratory Factor Analysis (EFA) for BNT-C 20 items test, using maximum likelihood and direct oblimin rotation, yielded a six factors solution, with the majority of items loading on first factor. The first four factors explained 34% of the variance; 22%, 5%, 4%, and 2%, respectively. In Confirmatory Factor Analysis (CFA), the four factors explained 33% of the variance: with 22%, 4%, 4%, and 3%, respectively. I next conducted separate EFAs for the four components using maximum likelihood, and direct oblimin rotation. For probability and geometry components, all 5 items, in each scale, loaded on one factor, however for operations and Algebra the EFA gave a two factor solution with majority of the items loading on first factor. Results suggest that overall, the test performed as intended with some variability in factor structure reflecting the constraints of modeling a wide-range (e.g., sensitivity and discriminability) of highly related skills (e.g., specificity).

BNT-C and its four components were found to be significantly and positively related to existing numeracy and ability measures and uncorrelated to personality measures, providing an evidence of convergent and discriminant validity (see Table 5-3 for all correlations). Overall, the BNT-C provided desirable psychometric estimates of (1) overall numeracy and (2) four essential numeracy sub-skills (i.e., operations, probability, geometry, and algebra). This out of sample cross validation study indicates that we have a technology that is psychometrically sound, robust, and reliable. The new measurement tool appears to offer good estimates of overall as well as component numeracy skills in educated samples from industrialized countries (see Cokely et al, 2012 for additional

details on generalizability to 15 other industrialized countries). Accordingly, I next addressed the second question as stated in research hypothesis: How do these numeracy component skills relate to superior decision making skill?

	BNTA	BNT	BNT-S	BNT-C	Oper	Probab	Geom	Algeb	CRT	Lipkus	Schw	NUMi	Weller
BNTA													
BNT	.94**												
BNT_S	.83**	.90**											
BNT-C	.47**	.50**	.51**										
Operation	.47**	.50**	.52**	.76**									
Probability	.37**	.38**	.43**	.81**	.47**								
Geometry	.33**	.37**	.36**	.75**	.44**	.45**							
Algebra	.30**	.32**	.31**	.81**	.53**	.52**	.49**						
CRT	.49**	.53**	.55**	.53**	.49**	.44**	.31**	.40**					
Lipkus	.31**	.36**	.63**	.34**	.39**	.26**	.23**	.19**	.45**				
Schwartz	.28**	.33**	.71**	.30**	.31**	.32**	.18**	.14	.33**	.78**			
NUMi	.31**	.37**	.45**	.54**	.40**	.48**	.41**	.38**	.51**	.43**	.37**		
Weller	.46**	.51**	.72**	.50**	.49**	.43**	.33**	.33**	.79**	.77**	.73**	.52**	
Intelligence	.37**	.38**	.44**	.51**	.51**	.44**	.37**	.29**	.44**	.34**	.33**	.36**	.49**
Extraversion	-.02	-.09	-.11	-.18	-.16	-.16	-.02	-.21	-.07	-.05	-.1	-.24*	-.12
Agreeable	-.003	-.02	.04	.02	.08	-.08	.06	.03	.03	.03	.05	.05	.04
Conscien.	-.005	-.04	-.04	-.07	-.06	-.11	.05	-.08	-.13	-.01	-.02	-.01	-.07
Openness	.04	-.03	-.07	-.15	-.06	-.17	-.02	-.19*	-.05	-.05	-.09	-.06	-.08
Emo.													
Stability	.20*	.19*	.18*	.16	.14	.09	.21*	-.08	.22*	0.15	.07	.13	.21*

** Correlation is significant at 0.01 level; * Correlation is significant at 0.05 level

5.4.2 Component Analysis

5.4.2.1 Data and Descriptive Analysis

First data were inspected for any outliers; to detect univariate outliers, numeric variables were converted to their standard score equivalents. Data was then inspected for any value above +3 or below -3, assuming normal z distribution. For multivariate outliers, Mahalanobis distances were calculated for all predictor and criteria variables. I converted the Mahalanobis distance values to probability values to identify the cases that were most distant, or different, from the other cases in the sample ($p > .05$).

The assumptions of a linear relationship and homoscedasticity of errors were tested through examination of residual vs. predicted plots. Tolerance values for all main analysis were found to be $> .1$. Also Variance Inflation Factors (VIF) were assessed for all models to determine whether VIF was > 10 for any model in order to assess multicollinearity. The assumption of no autocorrelation among residuals was examined with the Durbin-Waston test ($\alpha = 0.05$).

Descriptive statistics were calculated for all predictors, i.e., numeracy and ability measures, and all criterion variables including ADMC, standard risky, and ecological risky decisions (See appendix G for descriptive statistics for all variables). High correlations between all predictors and criteria measures indicated consistency—the standard positive performance manifold—in performance across various measures (see Table 5-4).

Table 5-2: Pearson product moment correlation coefficients showing relationship of all numeracy measures with new test BNT-C and its four components

	BNTC	BNT	BNTS	Sch	Lipkus	CRT	Welr	NUMI	Opr.	Pro.	Geo.	Alge	Intel.
Framing	.25**	.17	.21*	.19*	.15	.16	.19*	.25**	.13	.22*	.24**	.17	.16
Social Norms	.31**	.21*	.26**	.24*	.18*	.22*	.27**	.38**	.20*	.30**	.23*	.24**	.29**
Confidence	.27**	.18*	.13	-.01	-.11	-.18*	-.10	-.23*	-.17	-.19*	-.25**	-.24*	-.15
Sunk Cost	.25**	.33**	.30**	.12	.06	.15	.15	.05	.17	.26**	.10	.23*	.29**
Decision Rules	.24**	.14	.19*	.18*	.13	.22*	.28**	.32**	.28**	.20*	.08	.18*	.34**
Path Independ.	.12	.09	.10	.06	.05	.12	.10	.19*	.11	.05	.12	.10	.14
Risk Perception	.20*	.15	.17	.12	.12	.10	.11	.05	.14	.23*	.08	.16	.16
DOI	.12	.07	.01	-.10	-.10	.18*	-.001	.15	.08	.08	.12	.09	.19*
Eco Finance	.39**	.25**	.31**	.27**	.32**	.35**	.42**	.35**	.39**	.32**	.28**	.25**	.30**
Eco Medical	.41**	.21*	.22*	.14	.18	.34**	.27*	.31**	.36**	.32**	.30**	.31**	.23*
Ratio Bias	.32**	-.004	.09	.21*	.13	.13	.19*	.19*	.24*	.23**	.25**	.28**	.15
Lotteries	.13	.25**	.22*	.06	.15	.18*	.18*	.08	.08	.12	.13	.07	.10
Inter-Temp	.30**	.22**	.31**	.32**	.17	.20*	.26**	.19*	.28**	.33**	.17	.14	.17
Intelligence	.51**	.38**	.44**	.33**	.34**	.44**	.46**	.36**	.51**	.44**	.37**	.29**	1
ADMC_T	.50**	.43**	.46**	.31**	.26**	.40**	.42**	.49**	.41**	.44**	.33	.40**	.49**
Stand. Risky	.44**	.21*	.32**	.36**	.24**	.27**	.35**	.27**	.36**	.40**	.32**	.31**	.23**
Risk Literacy	.54**	.30*	.37**	.32**	.33**	.44**	.47**	.43**	.50**	.44**	.39*	.38**	.35**
SDMS_T	.58**	.35**	.42**	.34**	.35**	.48**	.50**	.47**	.51**	.48**	.41**	.43**	.41**
Predicting/18	15	12	13	11	7	11	13	14	11	15	10	12	10
Strongest pred.	6	2	0	0	0	0	1	2	0	3	1	1	2

5.4.2.2 Theoretical Models

To investigate how different numeracy component skills are involved and are related to different numeracy, ability, and decision making skills, I built theoretically informed quantitative models, following my hypothesis, both with ability and decision making tasks. Components were added to the hierarchical model following a priori assumptions; 1) for risky decisions (e.g. intertemporal choice, lotteries, path independence, risk perception, etc.) probability was entered first, then operations, and then geometry and algebra were entered; 2) For ecological decision making; operations was entered first, then probability, then geometry and then algebra; and 3) for CRT and ratio bias, algebra was entered first, then geometry, and then operations and probability; 4) and for framing and confidence bias, geometry was entered first, then algebra, operations, and probability. Separate hierarchical regression analyses were conducted for each of the ability and decision making tasks, as well as overall major decision making categories—i.e. ADMC, risk literacy, and overall Superior Decision Making Skill (SDMS) (see Appendix H for the results from hierarchical regression for each of the decision and ability task). I have summarized all the results from hierarchical models in Figures 5-4, 5-5, and 5-6.

To see how the four components accounted for overall major decision categories and Superior Decision Making Skill (SDMS), hierarchical regression models were constructed following a priori assumptions for each type of decision category. As predicted, for ADMC; probability was the strongest predictor, and for risk literacy and overall SDMS; operations provided consistent and robust explanatory power. Algebra

and geometry did not predict overall superior decision making skill (SDMS) as well as risk literacy and standard risky decisions (see Figure 5-4).

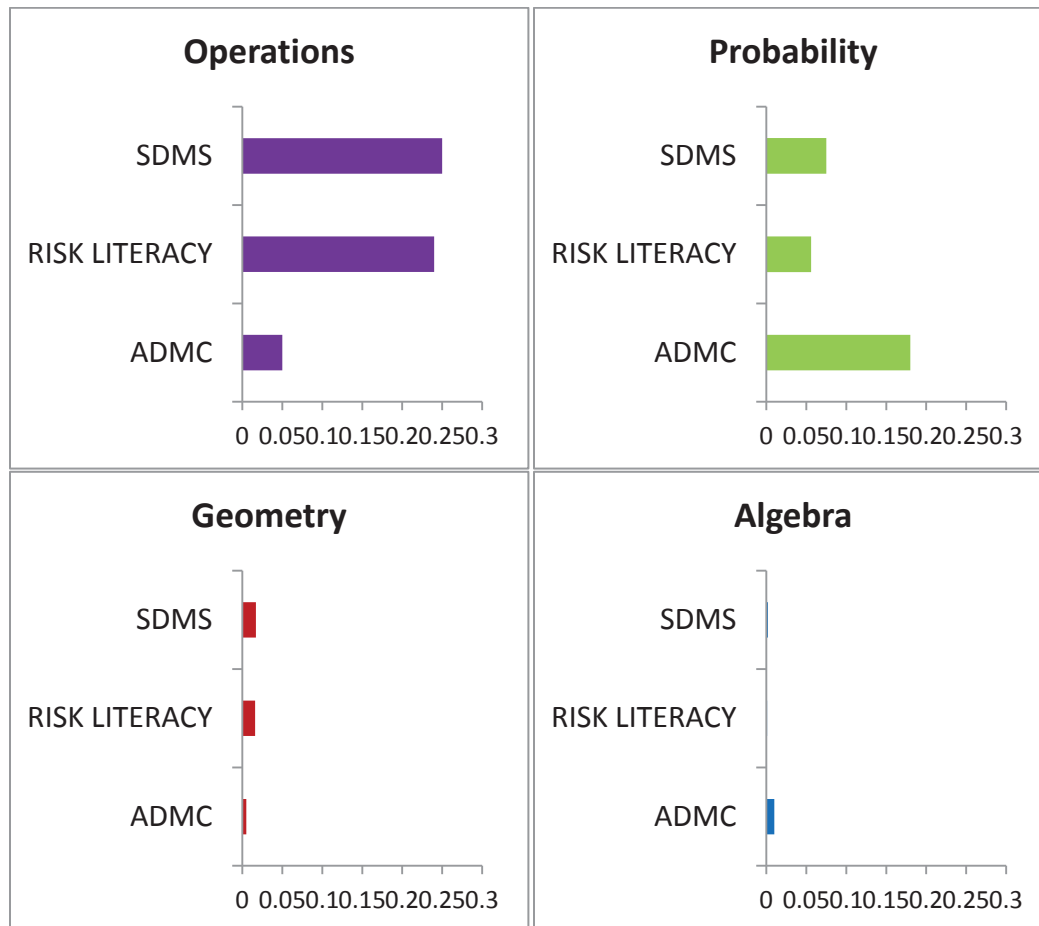


Figure 5.4. Hierarchical regression models (length of bars showing adjusted R² values) for overall superior decision making skill

Next, hierarchical regressions with four components predicting each of the individual decision tasks were conducted. Results are shown in Figure 5-5. Data supported the hypotheses; 1) probability significantly predicted intertemporal decision tasks, risk perception, better than other three components; however it also predicted sunk cost and social norms. 2) Operations, as hypothesized, best predicted ecological

decisions; 3) Geometry predicted confidence bias and framing; and 4) Algebra significantly predicted ratio bias and confidence bias (see Figure 5-5.). Overall for all decision types, probability component predicted more decision tasks as compared to other three components (see Figure 5-5.).

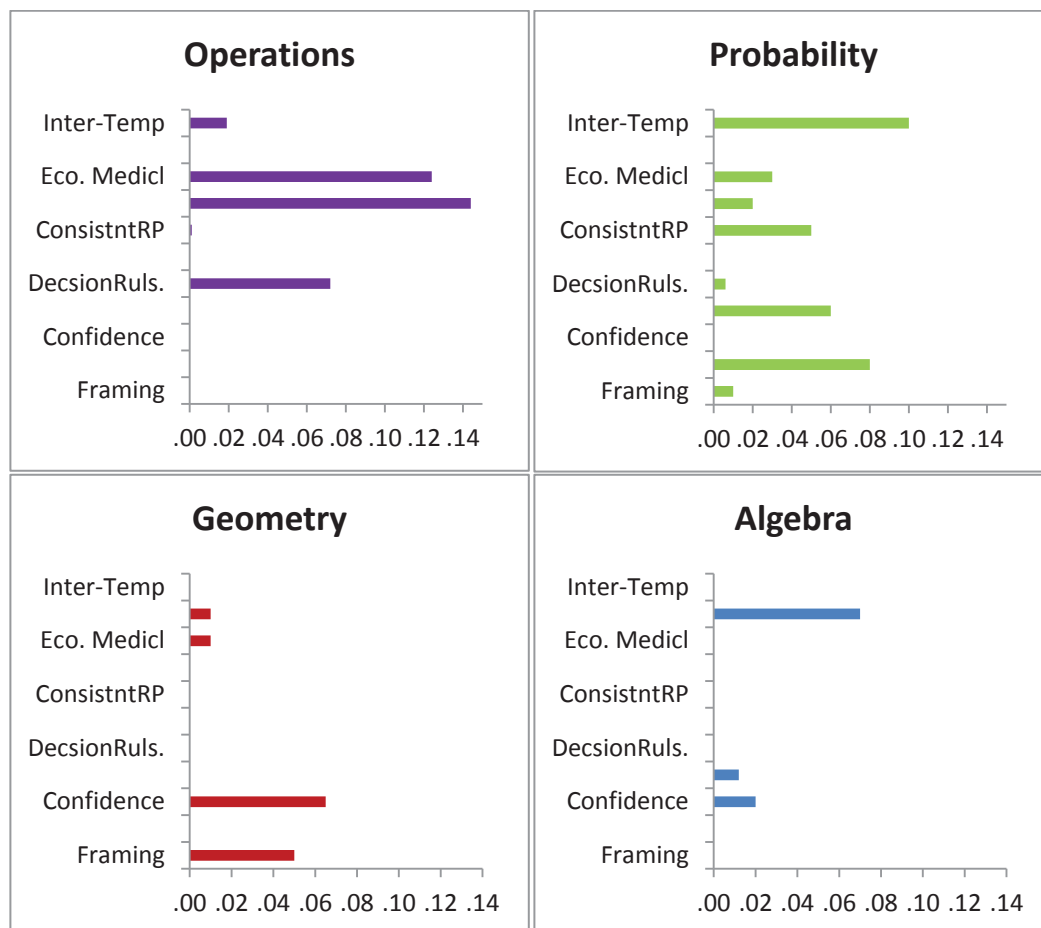


Figure 5.5. Hierarchical regression following theoretical models (length of bars showing adjusted R2) for all decision tasks

Similarly, hierarchical regression analyses were conducted for all ability measures. Because there were no previously stated assumptions about the ability measures and how they relate to different numeracy component skills, guidance was taken from step wise

regression followed by model estimation via hierarchical regression modeling (see Appendix I for results from step wise regression). Results indicated that the operations component explained most of the variance in all ability measures including intelligence and CRT. NUMi was by far most strongly, significantly, and uniquely related to probability, and geometry (see Figure 5-6).

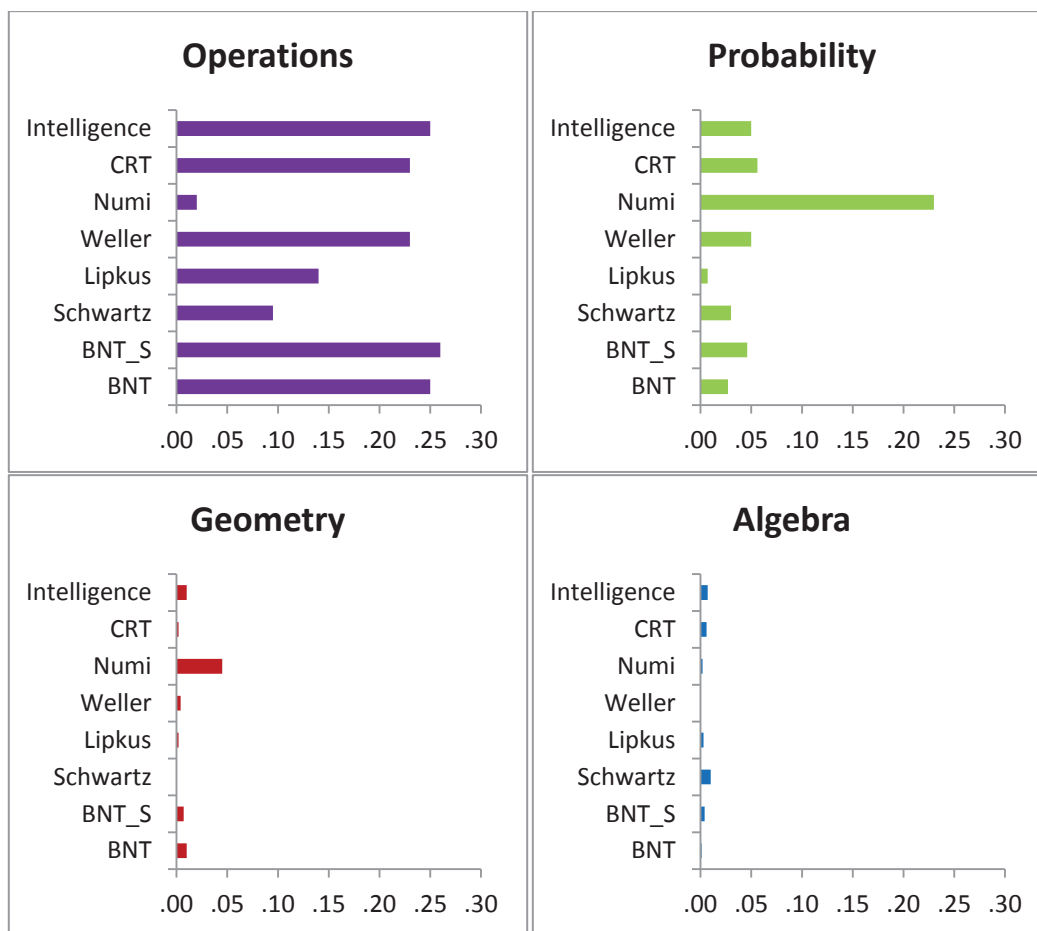


Figure 5.6. Hierarchical regression models (length of bars showing adjusted R²) for all ability measures

Overall, hierarchical regression analysis indicated that operations was the most important component, explaining majority of the predictable variance in overall Superior Decision Making Skill ($R^2 = .26$, $F(3, 116) = 40.5$, $\beta = .51$, $p < .001$). All four components in total explained 36% of the variance and operations component alone explained 72% of that explained variance (see Figure 5-7).

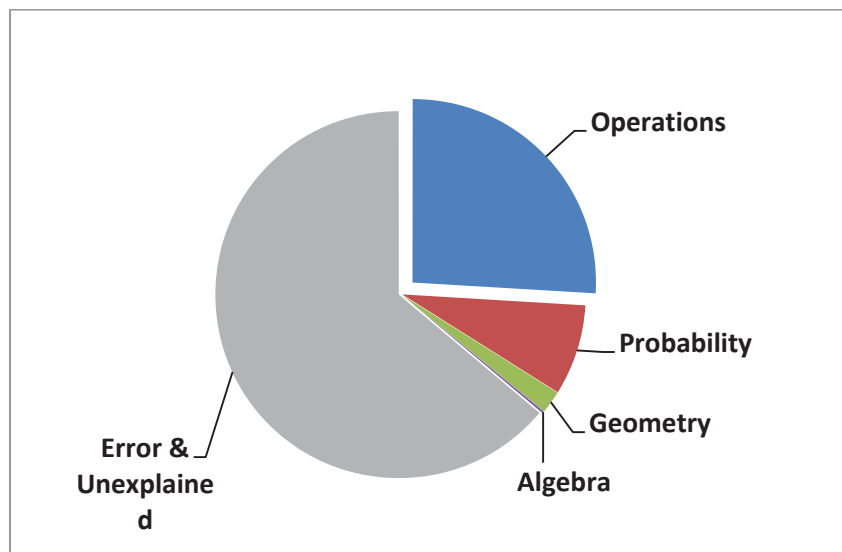


Figure 5.7. Percentage of variance explained by each numeracy component in overall Superior Decision Making Skill (Hierarchical regression models)

5.4.2.3 Individual Models (Components as Single Predictors)

Because of the high positive correlations between the four components, I further assessed the relationship of each component with each decision task and ability measure, using each component as a single predictor. Results indicated that four components were significantly and positively related (Pearson r ranging from .30 to .54) to each of the decision task, ability measure, and overall decision making skill (see Appendix M for all Figures from independent models). However, operations and probability components

explained majority of the decision and ability tasks more robustly and precisely than geometry and algebra components. Also, 1) Algebra was the strongest single predictor for ratio bias tasks; 2) geometry was the strongest single predictor for confidence bias; 3) probability was the strongest single predictor for social norms, sunk cost, risk perception, inter-temporal choice tasks, ADMC, and standard risky decision tasks; and 4) operations was the strongest single predictor for applying decision rules, ecological financial and medical decisions, and overall superior decision making skill (see Appendix M).

Similar analyses were conducted with each of the ability measure predicted by each component as a single predictor. Results indicated that BNT-S and intelligence were most strongly predicted by operations, NUMi was most strongly predicted by probabilities, and CRT was most strongly predicted by algebra (see Appendix N). However, the difference between the correlations of ability measures with four components was not statistically different (see Appendix M).

A similar analysis was conducted with overall superior decision making and major decision types, using each component as a single predictor. Results indicated that each component significantly contributed to superior decision making, however, operations and probabilities predicted more strongly and robustly as compared to geometry and algebra components (see Appendix N for individual analysis of the components).

5.4.3 Composite and Interaction Analysis

To estimate the model of best fit, a multiple simultaneous regression was conducted for each of the decision and ability task, where all four components were

entered into the model simultaneously. Results indicated, as observed before, that operations and probability were the two most explanatory and significant contributors when predicting performance on most of the decision and ability tasks. After controlling for operations and probability, algebra and geometry did not appear to add any unique predictive power in predicting performance on any of the ability measure including intelligence (except NUMi). For ability tasks, operations was the strongest predictor (with strongest regression weights) for BNT, BNT-A, BNT-S, CRT, Schwartz, Lipkus, Weller and intelligence (8 out of 9 ability tasks) and for applying decision rules, ecological financial decisions, ecological medical decisions, risk literacy, overall ecological decisions, and overall SDMS (6 out of 17 decision tasks). The probability component was the strongest predictor of performance on NUMi (1 out of 9 ability tasks) and intertemporal, sunk cost, risk perception, social norms, ADMC, and standard risky decisions (6 out of 17 decision tasks; see Appendix J for simultaneous regression models; both for decision and ability tasks, also see Figure 5-8 for quick summary).

Previous research has indicated that algebra and geometry are jointly related to students' latter achievements in STEM (Rose and Bets, 2001), therefore I combined algebra and geometry (named it as BNT-Engineering) in order to investigate the potential changes in the predictive power for each decision and ability task. Results from simultaneous regression models indicated an improvement; algebra and geometry together (BNT-engineering) predicted framing effects ($R^2 = .04$, $F(1, 3986) = 180.86$, $\beta = .21$, $p < .001$) and overconfidence bias ($R^2 = .09$, $F(3, 117) = 3.62$, $\beta = -.27$, $p < .001$). People high on BNT-engineering showed less susceptibility to framing effects with lower

levels of confidence (instead of over confidence), indicating that those having high algebra and geometry component skill levels tend to have lower levels of confidence but high performance. I hypothesize that this may reflect experience showing the value of a double checking heuristic, which could manifest as lower levels of confidence and better attention to detail and consideration of different missing variables (e.g., “well, 90% of risk is the same as 10% gain”) (see Appendix K for all simultaneous regression models, also see Figure 5-8).

As expected, results from hierarchical theoretically based (causally informed) regression models and simultaneous or stepwise models (best fit) largely converge, suggesting that the relationships among components accords with theoretical models (e.g., the central influence of statistical numeracy on decision making).

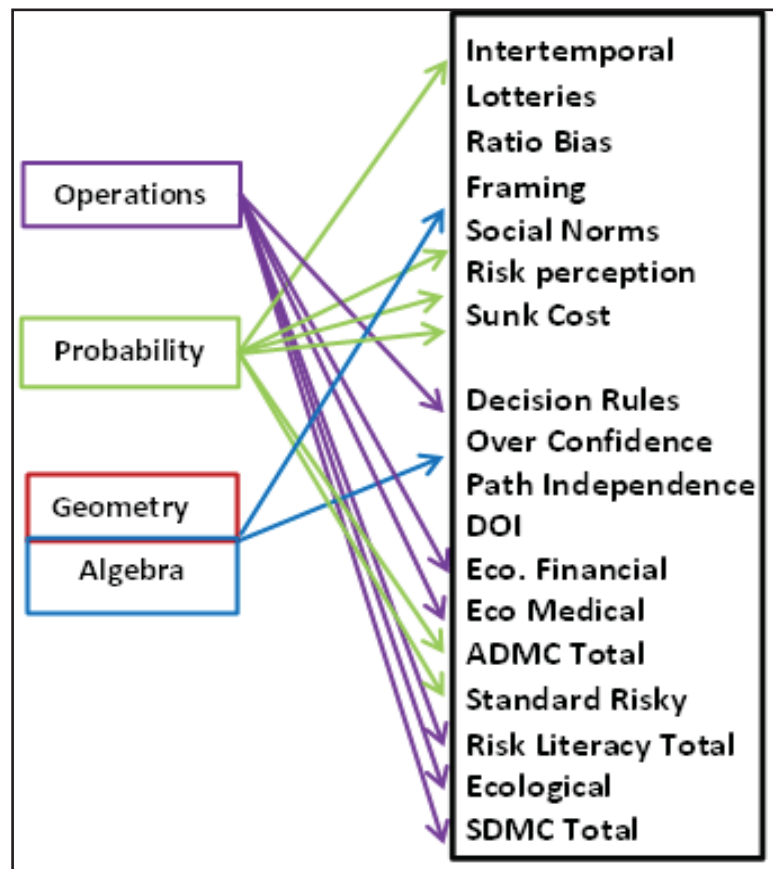


Figure 5.8. Components predicting different types of decisions (multiple simultaneous regression)

5.4.3.1 Interaction Analysis

Because adding geometry and algebra together improved the predictive power as compared to their individual predictive power, I further explored the possibility of interaction effects. I computed the 2way, 3way and 4way interaction terms. Then I conducted step wise regression, where I entered 4 components along with all 2way, 3way and 4way interaction terms together. This optimal fit model indicated that the strongest predictor, for ADMC, standard risky, ecological, risk literacy, and overall superior

decision making competence, was the interaction of operation and probability. I next compared this interaction term (operation X probability) to the combined component (BNT-stats). I conducted 2 separate hierarchical regressions; one where I entered interaction of operation*probability, and then interaction of geometry*algebra, and in second hierarchical regression I entered BNT-stats and then BNT-engineering. Results indicated that interaction terms explained numerically more variance as compared to combined term; however, this difference was not statistically significant (see Table 5-5), suggesting that main effects of each skill tends to be direct and linear, regardless of other skill levels (among educated participants).

Table 5-5: *Model comparison for interaction vs combined effects Note: Table showing R2 values from hierarchical regression models*

Decision Tasks	M1: Interaction	M2: Combined	t-value
ADMC	0.24	0.24	0
Standard Risky	0.16	0.20	1.03 n.s
Ecological decisions	0.26	0.22	2.03 n.s
Risk literacy	0.33	0.29	1.70 n.s
SDMS-Total	0.36	0.33	1.30 n.s

5.4.3.2 Composite Analysis

I further analyzed the predictive power of BNT-C as a one comprehensive numeracy test and compared it with other ability measures. I entered all existing numeracy measures, along with BNT-C, CRT, and intelligence into a stepwise regression, to predict each type of individual decision task as well as overall superior decision making competence (see Appendix K for step wise regression analysis). BNT-C outperformed all other measures, including intelligence; it predicted and explained better

and more often than other ability measures (strongest predictor for 10 out of 19 decision tasks, including overall risk literacy and superior decision making skill; see Figure 5-9).

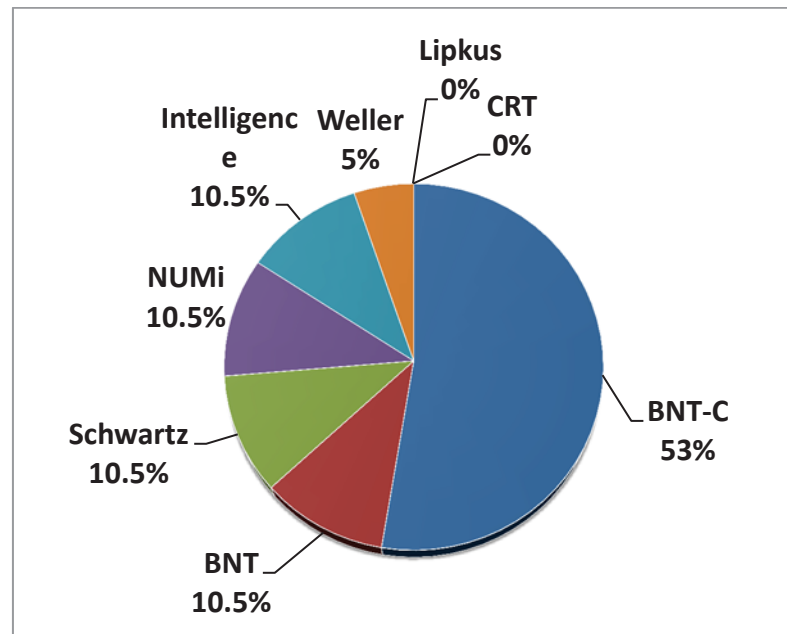


Figure 5.9. Graph showing percentage of number of times each of the ability measure showed itself as strongest predictor (stepwise regression) out of 19 decision tasks

It was hypothesized that numeracy would show unique predictive power for all criteria measures over and above the other ability measures/predictors such as intelligence, and CRT. The results supported the hypothesis; hierarchical regression results indicated that BNT-C explained 34% of the total variance, or 89% of the explained variance (see Figure 5-10) in overall superior decision making skill.

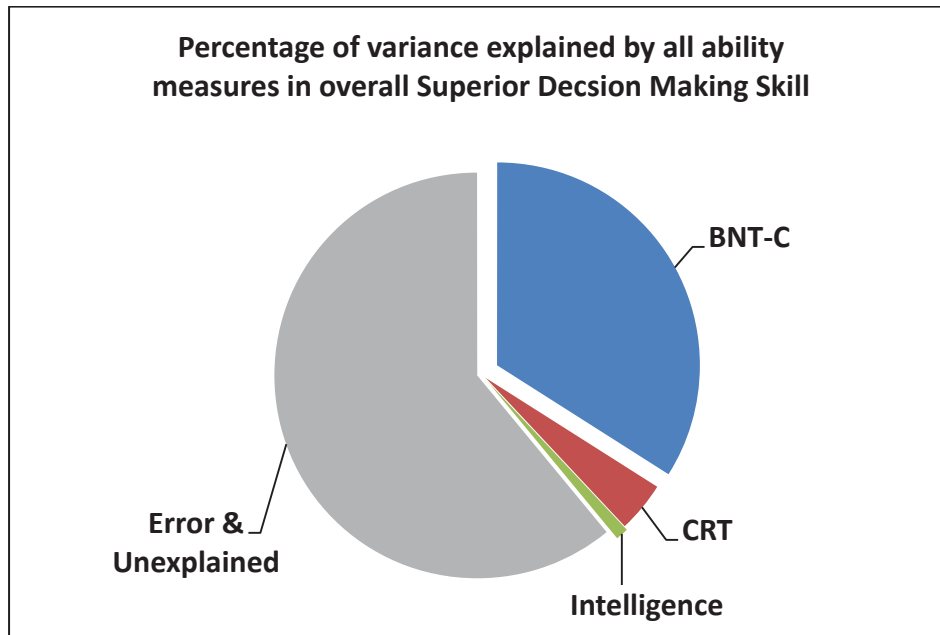


Figure 5.10. Percentage of total variance (including unexplained and error) in overall decision making competence by abilities

5.4.4 Competitive Analysis

It was hypothesized that there would be differences in math sub-skills in predicting different types of decision making tasks; e.g. probability component of numeracy will provide the strongest predictive power for risky decisions (e.g., selecting lotteries, making inter-temporal choices, avoiding denominator neglect) compared to other three components because risk is fundamentally a ratio concept that requires consideration of reference classes and ratio relations. Competitive analysis will allow to compare the four components (following theoretical models), four components as single predictors, with total BNT-C, BNT-stats, BNT-engineering, and maximum variance model (variance explained by everything together including all ability and decision tasks), to see which model explain the maximum variance in criterion measures.

Competitive analyses were conducted to generate regression equations, based on best fit models, for predicting decision and ability skills.

First simultaneous multiple regression was conducted, where all ability and all decision measures were entered into the model as predictors (excluding BNT-C and the one that is being predicted) across each of the decision task and ability measures. This provided the maximum variance that could be explained by all of the abilities and decision measures combined together. Next, simple linear regression was conducted for BNT-C total as a single predictor. I already have calculated component analysis following theoretical (hierarchical regression) and independent predictor models. For competitive analysis of these models, I will start with overall superior decision making skill and major decision making categories (e.g. risk literacy, ADMC, etc.).

5.4.4.1 Overall Superior Decision Making Skill (SDMS)

For overall superior decision making, I calculated the average weighted proportion for each of the decision category, i.e. ADMC, standard risky and ecological decisions. I have plotted the results from 5 different regression analyses (adjusted R²s) on the same bar graph (for ease of comparisons).

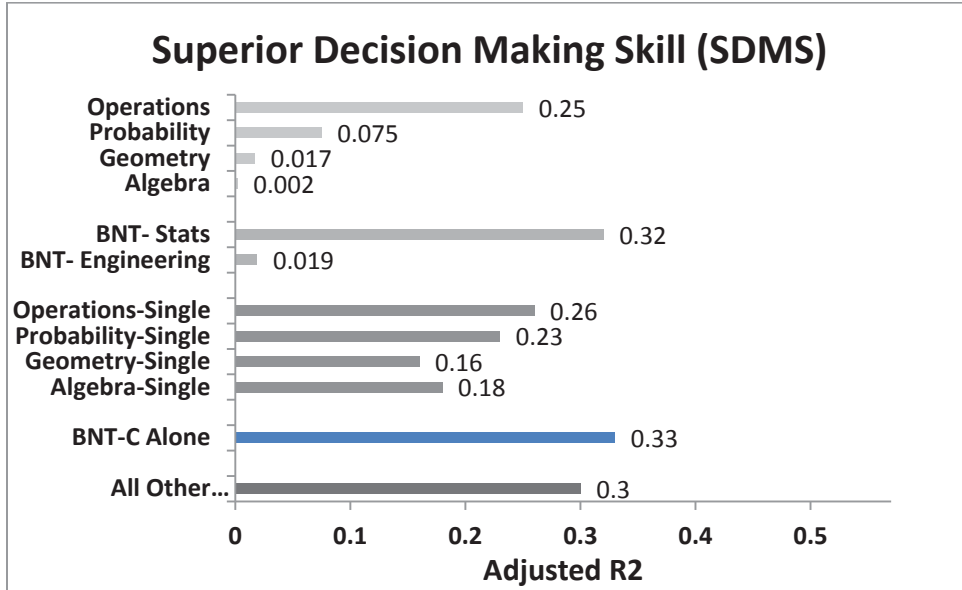


Figure 5.11. Numeracy component skills predicting overall Superior Decision Making

Note: Adjusted R² from 5 different regression analysis: (1) maximum variance explained by all ability measures combined, using simultaneous regression (2) Linear regression with BNT-C as a single predictor (3) linear regressions with four components as single predictors (4) BNT-stats and BNT engineering following theoretical hierarchical regression (5) four components following theoretical hierarchal regression models

Results from hierarchal regression analysis indicated ‘operations’ as a strongest component ($R^2 = .25, p < .001$) in predicting overall SDMS. Prediction got even better, when operations and probability were combined. BNT-stats explained 32% of the variance. Results from simple linear regressions revealed that all 4 components, as single predictors, significantly predicted SDMS (adjusted $R^2 = .16$ to $.26, p < .001$), however, operations and probability components predicted more strongly as compared to geometry and algebra components. The best model in predicting SDMS was BNT-C total score ($R^2 = .33, p < .001$). Overall, BNT-C alone outperformed all other predictors combined in

predicting superior decision making (see Figure 5-11). BNT-C alone explained 10% more variance as compared to variance explained by all other ability measures, including CRT and intelligence, combined (adjusted $R^2 = .33$ by BNT-C as compared to $R^2 = .30$ by all ability measures combined). BNT-Stats alone explained 6% more variance than all other measures combined, and Operations alone (5 questions only) explained close to what is explained by all other ability measures combined, in predicting overall superior decision making skill (see Figure 5-11).

5.4.4.2 Risk Literacy

For calculating performance on risk literacy, I computed and combined the weighted proportion of risky decisions and ecological decisions. Results from hierarchical regression indicated, that operations was the strongest predictor for risk literacy. Probability component also significantly improved the model (R^2 change = .06, $p < .001$). Algebra and geometry did not add anything significant into the model. BNT-stats and BNT-C explained equal amount of variance ($R^2 = .29$ $p < .001$) in predicting risk literacy. Overall, BNT-C (as well as BNT-stats) explained 7% more variance than the variance explained by all other ability measures combined ($R^2 = .29$ vs .27; see Figure 5-12).

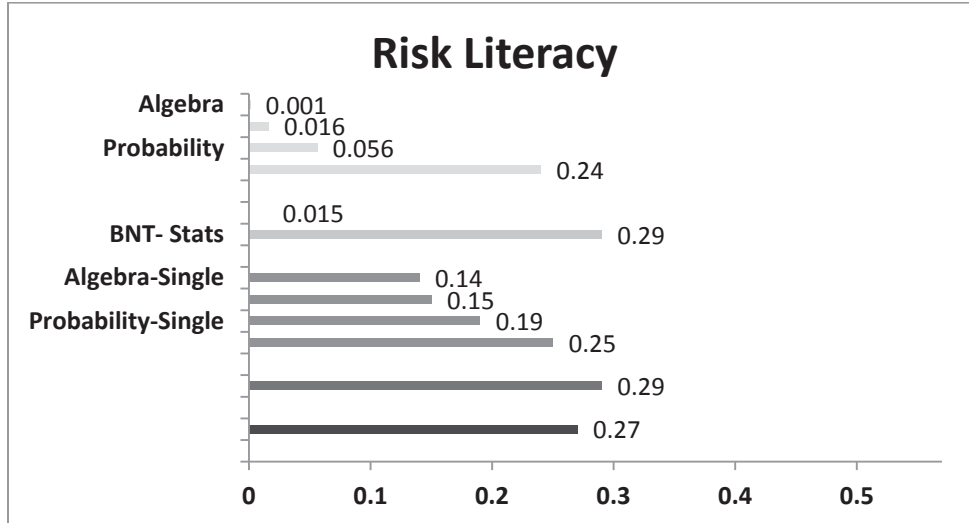


Figure 5.12. Numeracy and component skills predicting Risk Literacy.

Note: Adjusted R² from 5 different regression analysis: (1) maximum variance explained by all ability measures combined, using simultaneous regression (2) Linear regression with BNT-C as a single predictor (3) linear regressions with four components as single predictors (4) BNT-stats and BNT engineering following theoretical hierarchical regression (5) four components following theoretical hierarchal regression models

5.4.4.3 Standard Risky Decisions

I computed the weighted proportions for all paradigmatic risky decision tasks (intertemporal choice, lotteries, and ratio bias) and combined them together into standard risky decisions. Results indicated that probability was the strongest predictor for standard risky decision tasks, and explained 16% of the variance. Also BNT-stats was stronger than probability component alone (R^2 .19 vs .15). BNT-C, again, outperformed all the models, and alone explained 58% more variance than what is explained by all other ability measures combined (R^2 = .19 vs .12 max). Probability component alone (5 items) explained 25% more variance than variance explained by all ability measures combined (see Figure 5-13).

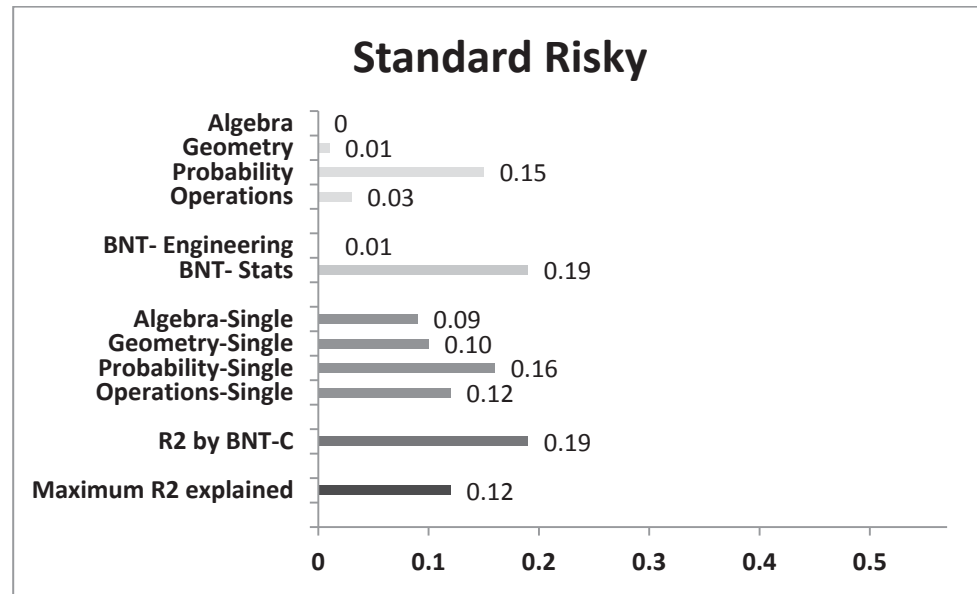


Figure 5.13. Numeracy and component skills predicting Standard Risky Decisions.

Note: Adjusted R² from 5 different regression analysis: (1) maximum variance explained by all ability measures combined, using simultaneous regression (2) Linear regression with BNT-C as a single predictor (3) linear regressions with four components as single predictors (4) BNT-stats and BNT engineering following theoretical hierarchical regression (5) four components following theoretical hierarchal regression models

5.4.4.4 Ecological Decision Making

I computed the weighted proportions for ecological medical and ecological financial decisions, and added them together. Results from hierarchal regression indicated that operations was the strongest predictor ($R^2 = .19$ $p < .001$). In addition, probability component significantly improved the model ($R^2_{\text{change}} = .034$ $p < .01$). BNT-C was the best model explaining 6% more variance than the maximum variance explained by all other ability measures combined (see Figure 5-1414).

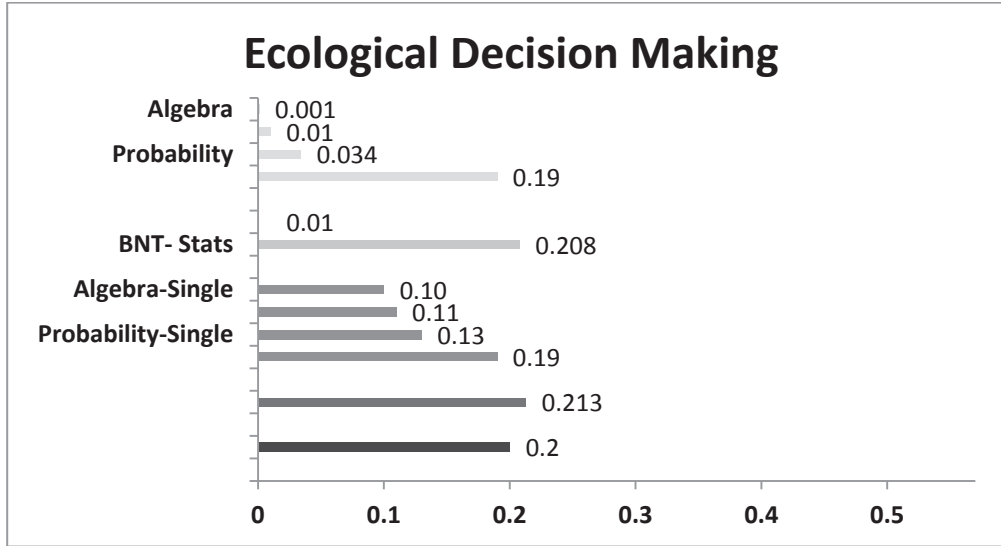


Figure 5.14. Numeracy and component skills predicting Ecological Decisions.

Note: Adjusted R² from 5 different regression analysis: (1) maximum variance explained by all ability measures combined, using simultaneous regression (2) Linear regression with BNT-C as a single predictor (3) linear regressions with four components as single predictors (4) BNT-stats and BNT engineering following theoretical hierarchical regression (5) four components following theoretical hierarchal regression models

5.4.4.5 ADMC

Weighted proportions for each ADMC subscales were calculated, except confidence¹². Results from hierarchical regression indicated that probability was the strongest predictor among all four components ($R^2 = .18$ $p < .001$). BNT-Stats (operations and probability combined) predicted even better ($R^2 = .23$ $p < .001$), once again BNT-C

¹² In ADMC original scoring, confidence was coded as absolute values, thus not providing any information about over or under confidence. I coded them manually, so that low score would reflect under confidence. Thus I analyzed it separately.

total shown to be best model ($R^2 = .25$ $p < .001$). However, maximum variance explained by other ability tasks together was higher than explained by BNT-C alone ($R^2 = .39$ $p < .001$ vs $R^2 = .25$ $p < .001$). This time BNT-C explained 65% of the maximum variance explained by all other ability measures together (see Figure 5-15).

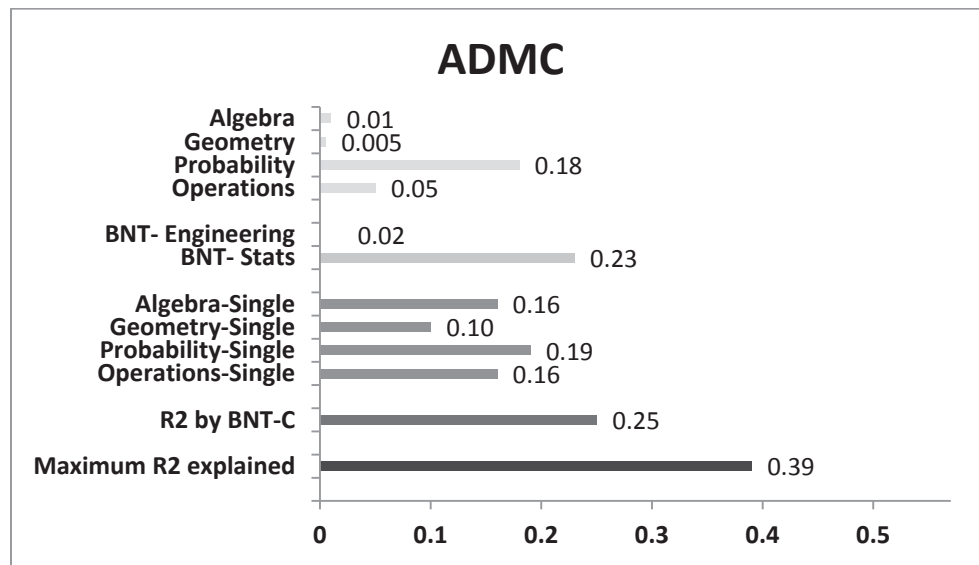


Figure 5.15. Numeracy and component skills predicting Adult Decision Making Competence (ADMC).

Note: Adjusted R² from 5 different regression analysis: (1) maximum variance explained by all ability measures combined, using simultaneous regression (2) Linear regression with BNT-C as a single predictor (3) linear regressions with four components as single predictors (4) BNT-stats and BNT engineering following theoretical hierarchical regression (5) four components following theoretical hierarchical regression models

In total, BNT-C outperformed and captured more than 100% of the variance explained by all ability measures combined, including intelligence, in predicting major decision making categories (see Figure 5.56). Similarly, analyses were conducted for each individual ability and decision task (see Appendix L), to investigate which

component predicts better, and how BNT-C perform as compared to all other ability measures.

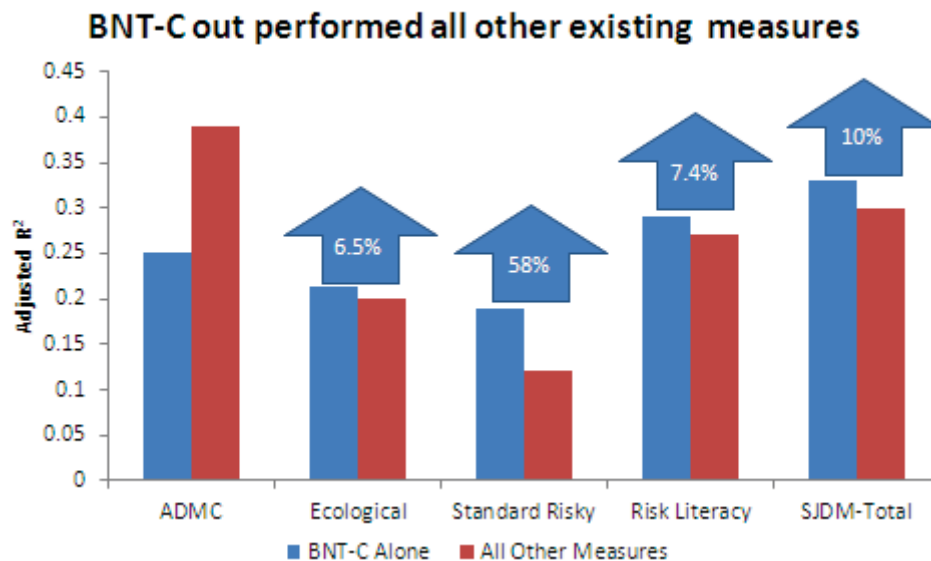


Figure 5.16. Variance explained by BNT-C in proportion to variance explained by all other ability measures combined

Results from all individual hierarchical regressions for each ability and decision task indicated that the four component numeracy skills are differently involved in different types of decisions; (algebra predicted ratio bias; geometry predicted framing and confidence; probability predicted sunk cost, risk perception, NUMi, ADMC, and standard risky choice tasks; and operations predicted decision rules, ecological financial decisions, ecological medical decisions, intelligence, BNT, BNT-S, Schwartz, Lipkus, Weller, CRT, risk literacy, overall SDMS). However, most important components for domain general superior decision making are operations and probability, i.e. BNT-Stats. Overall, BNT-C, and BNT-stats were the strongest predictors for majority of the ability and decision

making tasks (see Table 5-6). Figure 5.67 shows the importance of BNT-Stats (operations and probability combined) and BNT-Engineering (Geometry and Algebra combined) in terms of predicting different types of decisions. The data shown in Figure 5-18 clearly indicates that for risk literacy, and superior decision making skill, operations and probabilities are the two skills that we need most.

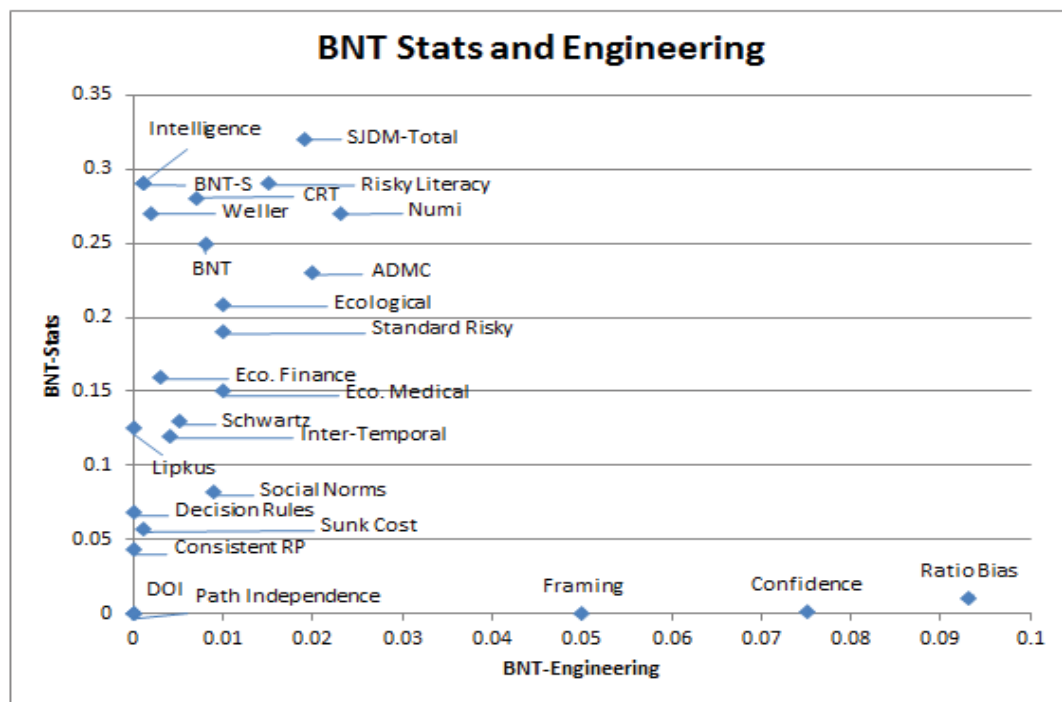


Figure 5.17. BNT-Stats VS BNT-Engineering predicting decision skills; Values indicating R² from hierarchical regression

Table 5-3: R2 from Maximum model, BNT-C only model, and hierarchical models with four components									
Variables	Max Var	BNT-C	% to max	BNT-Stat	BNT-Eng.	Oper.	Probab.	Geom.	Algebr a
Framing	0.013	0.053**	500%	0	0.05*	0	0.01	0.05*	0
Social Norms	0.20**	0.09**	45%	.082**	0.009	0	0.08	0	0
Confidence	0.19**	0.07**	37%	0.001	.075**	0	0	0.065**	0.02
Sunk Cost	0.14*	0.053**	38%	0.057*	0.001	0	0.06*	0	0.012
Decision Rules	0.405**	0.05**	13%	0.068*	0	0.072*	0.006	0	0
Path Indep.	-0.01	0.005	50%	0	0	0	0	0	0
Risk percep.	0.04	0.03*	75%	0.043*	0	0.001	0.05*	0	0
Financial Eco	0.24**	0.15**	63%	.16**	0.003	0.144**	0.02	0	0
Medical Eco	0.24**	0.16**	67%	.15**	0.01	0.124**	.03*	0.01	0
Ration Bias	0.10*	0.094**	94%	0.01	.093**	0	0	0.01	.07**
Intertemporal	0.12*	0.08**	67%	.12**	0.004	0.019	.10**	0	0
Intelligence	0.34**	0.25**	74%	.29**	0.001	.25**	.05**	0.01	0.007
BNT	0.34**	0.24**	71%	.25**	0.008	.25**	.027*	0.01	0.001
BNT-S	0.56**	0.26**	46%	.29**	0.001	.26**	.046**	0.007	0.004
Schwartz	0.23**	0.085**	37%	.13**	0.005	.095**	.03*	0	0.01
Lipkus	0.20**	0.106**	55%	.125**	0	.14**	0.007	0.002	0.003
Weller	0.42**	0.24**	57%	.27**	0.002	0.23**	.05**	0.004	0
NUMi	0.46**	0.28**	61%	.27**	.023`	0.02	.23**	0.045**	0.002
CRT	0.40**	0.27**	68%	.28**	0.007	.23**	.056**	0.002	0.006
ADMC	0.39**	0.25**	65%	.23**	0.02	.05**	.18**	0.005	0.01
Ecological	0.23**	0.213**	107%	0.208**	0.01	0.19**	0.034*	0.01	0.001
Stand Risky	0.20**	0.19**	158%	0.19**	0.01	0.03*	0.15**	0.01	0
Risk literacy	0.33**	0.29**	107%	0.29**	0.015	0.24**	0.056**	0.016	0.001
SDMS- Total	0.37**	0.33**	110%	0.32**	.019`	0.25**	0.075**	0.017	0.002

Overall findings indicate that not only the different component skills are differentially involved in different types of decisions, but range of ability skills assessed also allows for better prediction, that's why BNT-C as total was found to be the best predictor for majority of the ability and decision tasks. One goal of the research was to develop a technology that is valid and reliable, and that predicts abilities, risk literacy, and superior decision making across variety of decision domains. Based on the superior predictive power of the BNT-C (best optimized models), further, I generated regression equations for predicting performance on each of the decision and ability measure.

5.4.5 Model Fitting

Following the regression equations from the optimized best fitted models, I calculated the predicted scores for each decision and ability task at each of the BNT-C quartile level, and interpreted them in terms of percentiles. For example, if someone is at 50th percentile on BNT-C (his score on BNT-C is 9/20), then his/her abilities and decision making skills would look similar to those in Table 5-8. I calculated the predicted scores, their confidence intervals (95%), and percentiles for each measure at BNT-C 25th, 50th and 75th percentile (see 5-Table 5-7, Table 5-8, and Table 5-9).

Table 5-7: *Regression equations of best fitted models for predicting decision making and abilities in college educated samples from BNT-C scores at 25th percentile*

Variables	a (Intercept)	b (slope)	Pred- icted	C I	Percentiles
Framing	3.85	0.02	3.97	3.9-4.06	30 th to 40 th
Social Norms	0.35	0.01	0.428	0.39-.47	30 th to 40 th
Confidence	3.5	-0.08	3.0	2.7-3.3	40 th to 60 th
Sunk Cost	3.7	0.034	3.9	3.8-4.1	40 th to 50 th
Decision Rules	65.0	1.04	71.2	66.7-75.8	30 th to 40 th
Risk Perception	58.4	0.50	61.4	58.8-64.0	40 th to 50 th
Eco Financial	0.42	0.083	0.91	.70-1.1	30 th to 50 th
Ecological Medical	0.51	0.094	1.1	0.85-1.3	30 th to 50 th
Ratio Bias	1.45	0.05	1.8	1.6-1.9	30 th to 50 th
Inter-Temp	5.03	0.13	5.8	5.36-6.25	30 th to 40 th
Intelligence	4.7	0.29	6.4	5.9-7.0	30 th to 40 th
BNT	0.73	0.12	1.4	1.2-1.7	30 th to 40 th
Schwartz	2.16	0.045	2.4	2.3-2.6	30 th to 40 th
Lipkus	8.4	0.08	8.9	8.6-9.1	30 th to 60 th
Weller	4.05	0.15	5.0	4.7-5.2	40 th to 50 th
NUMi	15.29	0.234	16.7	16.3-17.1	30 th to 40 th
CRT	0.19	0.126	0.95	0.73-1.2	30 th to 40 th
BNT_S	2.9	0.165	3.9	3.6-4.2	30 th to 50 th
ADMC	0.597	0.008	0.645	0.63-0.66	30 th to 40 th
ECOLOGICAL STANDARD RISKY	0.46	0.09	0.99	0.82-1.2	30 th to 40 th
RISK LITERACY	0.53	0.01	0.59	0.57-0.62	30 th to 40 th
SDMS-TOTAL	0.73	0.05	1.0	0.95-1.1	30 th to 40 th
	1.33	0.06	1.69	1.6-1.8	30 th to 40 th

Note: Predicted values and CI were calculated using SPSS regression save command

As can be seen, performance on ability and decision tasks is quite consistent with BNT-C quartile levels. At 25th percentile, the performance on other measures is mostly between 30th to 50th percentile; on BNT-C 50th percentile, performance on other

measures is mostly between 50th to 60th percentile; and on BNT-C third quartile, performance is mostly between 70th to 80th percentile (see Table 5.7, 5.8, and 5.9).

Table 5-8: *Regression equations of best fitted models for predicting decision making and abilities in college educated samples from BNT-C scores at 50th percentile*

Variables	a (intercept)	b (slope)	Pred- icted	C I	Percentiles
Framing	3.85	0.02	4.03	3.96-4.1	50 th to 60 th
Social Norms	0.35	0.01	0.47	0.43-0.50	40 th to 50 th
Confidence	3.5	-0.08	2.75	2.5-3.0	50 th to 60 th
Sunk Cost	3.7	0.034	4.04	3.9-4.15	50 th to 60 th
Decision Rules	65.0	1.04	74.3	70.8-77.9	50 th to 60 th
Risk Perception	58.4	0.50	62.9	60.9-64.9	50 th to 60 th
Eco Financial	0.42	0.083	1.16	1.0-1.32	50 th to 60 th
Eco Medical	0.51	0.094	1.35	1.18-1.53	50 th to 60 th
Ratio Bias	1.45	0.05	1.9	1.8-2.05	50 th to 70 th
Inter-Temp	5.03	0.13	6.2	5.8-6.54	50 th to 60 th
Intelligence	4.7	0.29	7.3	6.9-7.7	50 th to 60 th
BNT	0.73	0.12	1.8	1.6-2.00	50 th to 60 th
Schwartz	2.16	0.045	2.6	2.45-2.7	60 th to 70 th
Lipkus	8.4	0.08	9.1	8.9-9.3	60 th to 70 th
Weller	4.05	0.15	5.4	5.2-5.6	50 th to 60 th
NUMi	15.29	0.234	17.4	17.1-17.7	50 th to 60 th
CRT	0.19	0.126	1.33	1.16-1.5	50 th to 60 th
BNT_S	2.9	0.165	4.38	4.15-4.6	50 th to 60 th
ADMC	0.597	0.008	0.67	0.66-0.68	50 th to 60 th
ECOLOGICAL	0.46	0.09	1.26	1.1-1.4	50 th to 60 th
STAND. RISKY	0.53	0.01	0.63	0.61-0.65	50 th to 60 th
RISK LIT.	0.73	0.05	1.2	1.1-1.3	50 th to 60 th
SDMS-TOTAL	1.33	0.06	1.86	1.8-1.94	50 th to 60 th

Note: Predicted values and CI were calculated using SPSS regression save command

Table 5-9: *Regression equations of best fitted models for predicting decision making and abilities in college educated samples from BNT-C scores at 75th percentile*

Variables	a (intercept)	b (slope)	Pred- icted	C I	Percentiles
Framing	3.85	0.02	4.12	4.0-4.2	60 th to 70 th
Social Norms	0.35	0.01	0.52	0.48-0.56	60 th to 70 th
Confidence	3.5	-0.08	2.42	2.1-2.7	70 th to 80 th
Sunk Cost	3.7	0.034	4.2	4.0-4.3	70 th to 80 th
Decision Rules	65.0	1.04	78.5	74.2-82.8	60 th to 80 th
Risk Perception	58.4	0.50	64.9	62.4-67.4	60 th to 80 th
Eco Financial	0.42	0.083	1.5	1.3-1.7	70 th to 80 th
Eco Medical	0.51	0.094	1.7	1.5-1.9	70 th to 80 th
Ratio Bias	1.45	0.05	2.1	2.0-2.3	70 th to 80 th
Inter-Temp	5.03	0.13	6.7	6.3-7.1	70 th to 80 th
Intelligence	4.7	0.29	8.5	8.0-9.0	70 th to 80 th
BNT	0.73	0.12	2.3	2.1-2.5	70 th to 80 th
Schwartz	2.16	0.045	2.8	2.6-2.9	70 th to 80 th
Lipkus	8.4	0.08	9.4	9.2-9.6	70 th to 80 th
Weller	4.05	0.15	6.0	5.7-6.3	70 th to 80 th
Numi	15.29	0.234	18.3	18.0-18.7	70 th to 80 th
CRT	0.19	0.126	1.8	1.6-2.0	70 th to 80 th
BNT_S	2.9	0.165	5.0	4.8-5.3	70 th to 80 th
ADMC	0.597	0.008	0.70	0.69-0.72	70 th to 80 th
Ecological	0.46	0.09	1.6	1.4-1.8	70 th to 80 th
Standard Risky	0.53	0.01	0.67	0.65-0.70	70 th to 80 th
Risk Literacy	0.73	0.05	1.4	1.3-1.5	70 th to 80 th
SDMS-Total	1.33	0.06	2.1	2.0-2.2	70 th to 80 th

Note: Predicted values and CI were calculated using SPSS regression save command

5.4.5.1 Generating profiles

Based on the predicted values and percentiles, I generated some sample profiles for decision making skills and abilities, which is unique deliverable of my thesis. Figure 5.78 presents a sample profile for an individual whose score on BNT-C is at 50th percentile. The model indicates that if someone gets 9 or 10 on BNT-C, his/her performance would be, on average, between 40th to 60th percentile (except NUMi, Lipkus, and Schwartz, due to ceiling effect) on majority of the decision and ability tasks, suggesting an overall achievement on tests (see Figure 5.78).

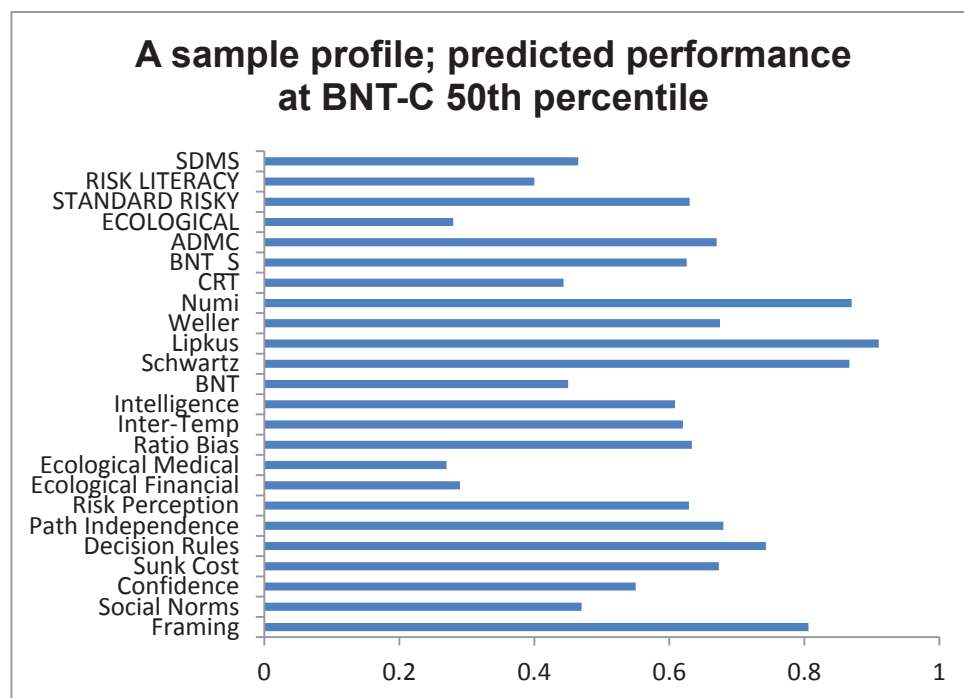


Figure 5.18. Predicted scores for all decision tasks and ability measures (weighted proportion on each measure) from BNT-C at 50th percentile

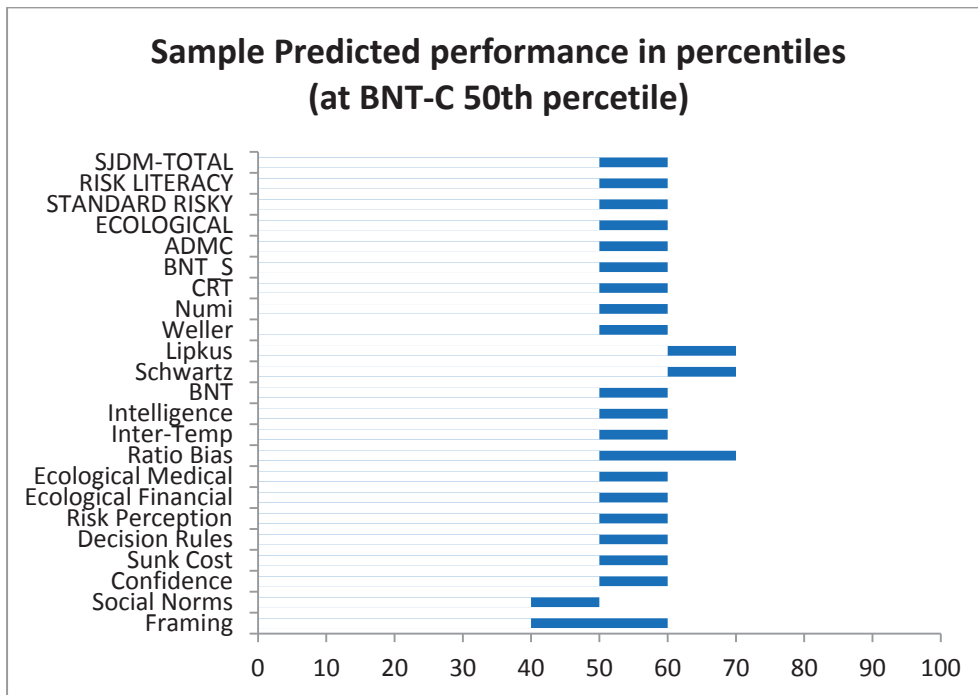


Figure 5.19. Predicted performance (in percentiles) on each ability & decision task

5.4.5.2 Model Recovery

To see the accuracy of the prediction models, I randomly drew a sample at each of the BNT-C quartile. Results indicated that there was no significant difference between actual and predicted values at 25th percentile ($t(46) = -1.0, P = .30$) and 50th percentile ($t(46) = .56, P = .58$) and 75th percentile ($t(46) = -1.6, P = .14$) (see, Figure 5.9, Figure 5.10, and Figure 5-22 for actual and predicted values plotted).

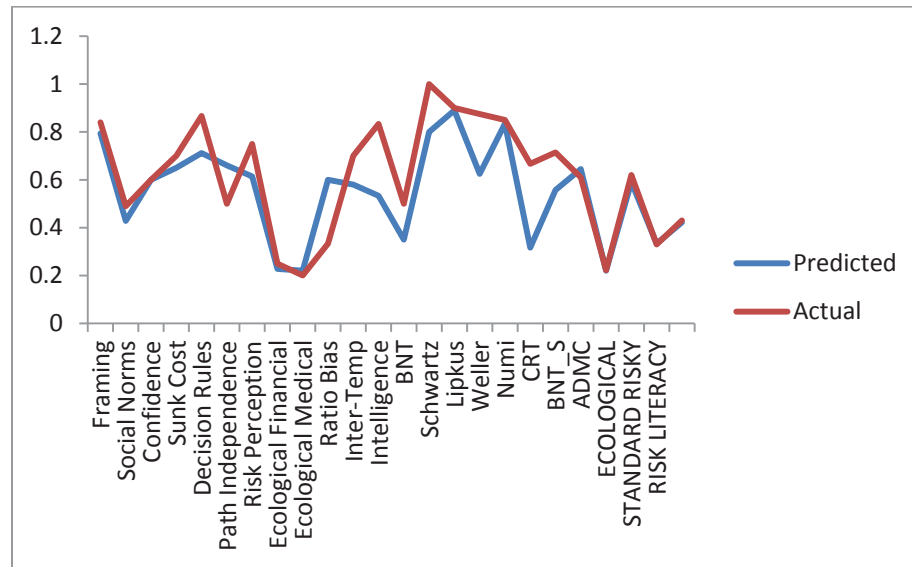


Figure 5.20. Predicted and actual scores for all decision and ability task at BNT-C 25th Percentile

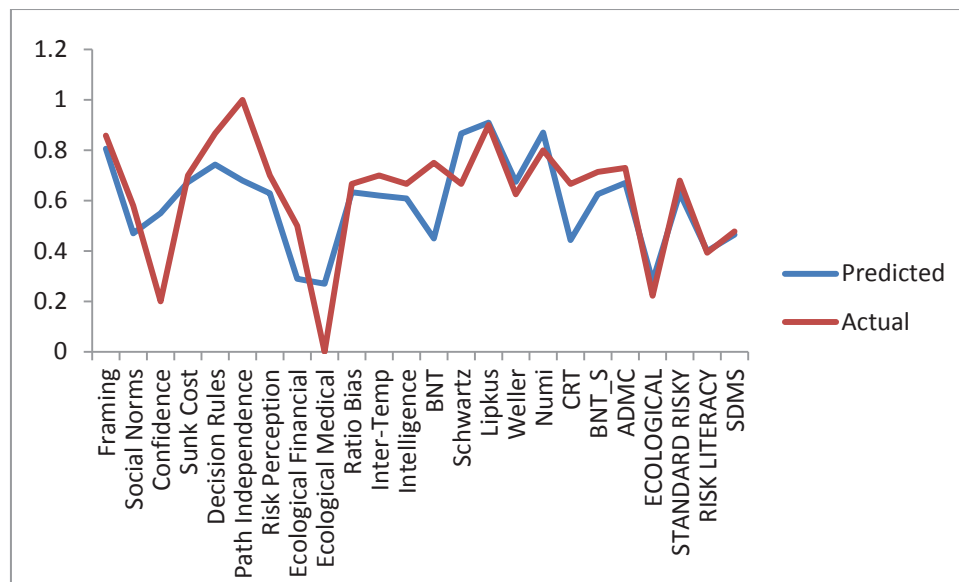


Figure 5.21. Predicted and actual scores for all decision and ability task at BNT-C 50th Percentile.

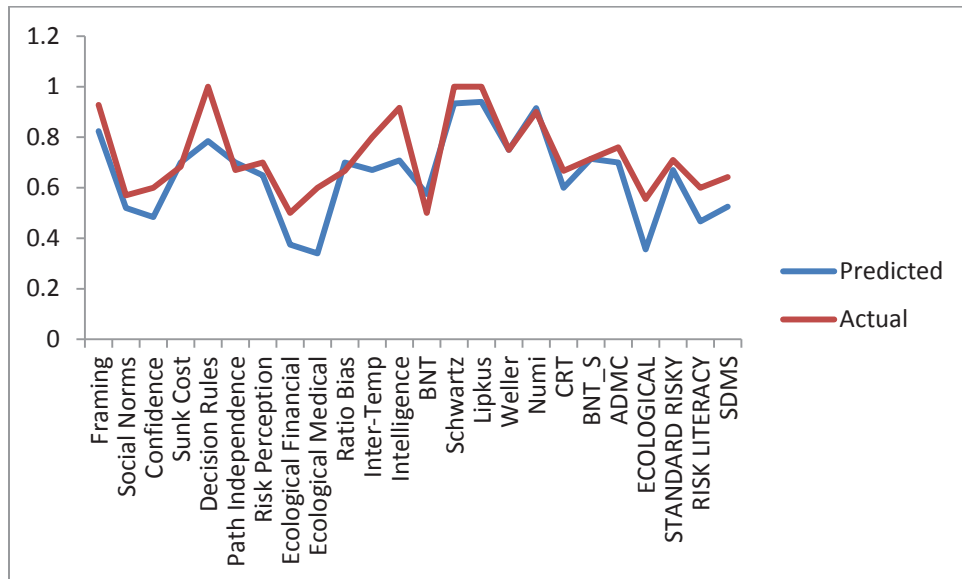


Figure 5.22. Predicted and actual scores for all decision and ability task at BNT-C 75th Percentile.

5.4.6 Summary of the Study 2

The goal of the second study was to systematically assess and explore the links between numeracy its sub-skills and various types of decision making. Component analysis indicated that four components are differently related to different types of decisions; however, operations and probability are the two major components that explained most of the variation in overall superior decision making and abilities. Results from theoretically based hierarchical models and simultaneous or step wise models largely converged, suggesting central influence of statistical numeracy (BNT-Stats). Competitive analysis identified the best fit model; i.e. overall BNT-C total predicted better than four components individually. Overall BNT-C overtook more than maximum variance explained in predicting overall superior decision making skill (SDMS). Based on the optimized best models, regression equations were generated for predicting each ability and decision task. The prediction models/profiles that I generated aligned very close and consistent with the actual data (no significant differences were found between actual and predicted scores), indicating consistency and reliability of the BNT-C.

5.5 Discussion

In study I, I developed an optimized brief component numeracy test for use with college educated samples—i.e., the Berlin Numeracy Components Test (BNT-C). The BNT-C provides rapid and robust assessment of overall adult numeracy and major numeracy sub-skills. Results reveal desirable psychometric performance across the full

range of essential functions (e.g. difficulty, discrimination, and sensitivity). All four subscales provided comparable test information with maximal information produced close to the theoretically ideal theta level of 0 (i.e., the average ability level). Results further indicated considerable reliability (e.g., average cronbach alpha in two studies =.85), content validity (e.g., high correlations with reported SAT scores), convergent validity (e.g., high correlations with other numeracy, ability, and cognitive style measures), and discriminant validity (e.g., uncorrelated with personality subscales). Study II also documented superior psychometric performance of the BNT-C in an out-of-sample cross-validation test. Cognitive and predictive modeling provided the first detailed map of the differential relations of numeracy sub-skills and component decision skills and abilities. Results include the first known evidence that algebra and geometry sub-components are uniquely related to some influential decision skills (e.g., confidence calibration, sunk costs, ratio biases). However, in accord with the leading theoretical account (Cokely et al., 2012), overall decision skill was primarily a function of statistical numeracy—i.e., operations and probability numeracy components. Moreover, when compared to other ability instruments, the BNT-C provided 10% better predictive power than the best linear combination of all other ability measures combined. Without modifying any parameters, when compared to all of the world's best and most well-established numeracy and ability measures, including the Raven's Progressive Matrices test of general fluid intelligence, the BNT-C was consistently the best predictor of general decision making skill across a broad representative range of paradigmatic and ecological decision tasks.

5.5.1 Theoretical Implications

What do the current results reveal about the nature of superior decision making? Additional confirmatory factor analyses (PCA) indicated that as a single factor SDMS explained 22% of the variance in overall decision task performance—i.e., a relatively large effect by most standards. However, results from the optimal fit model indicated that superior decision making skill can be better explained as a function of four essential factors accounting for 56% of the total variance in all decision task performance (see Figure 5-23). To further investigate the structure of overall decision skill, I investigated predictive model recovery of the decision skill factors (e.g., PCA factor estimates) using a step wise regression. Regression accounted for 48% of the reliable variance when all ability measures were included, 80% of which was explained by the BNT-C (see Figure 5-24).



Figure 5.23. Factor structure of decision skill: percentage of total variance; explained and unexplained

In short, numeracy explained almost 40% of the reliable variance across all general decision making tasks. To put numeracy's predictive strength into perspective, the link between the BNT-C and decision making skill is about as strong as estimates of the link between gender and height (i.e., men tend to be tall). It is similar in magnitude to the relationship between temperature and distance from the equator in the United States. It is also about 30 times stronger than the meta-analytic estimate of the effect of ibuprofen on pain reduction (Meyer et al., 2001; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

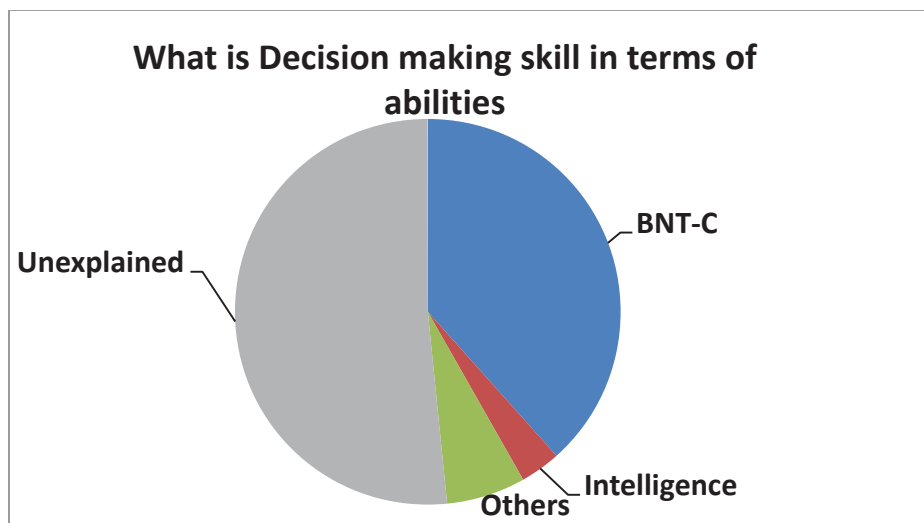


Figure 5.24. Percentage of total variance explained in Overall superior decision making skill by abilities

Why is BNT-C such a strong predictor of decision making skill? Theoretically, the tight link reflects the shared or “identical” cognitive elements present in numeracy and decision making, such as interpreting and inferring probabilities and risks, and regulating one’s own motivation, emotion, and information processing at the

metacognitive level (e.g., encoding, double checking, searching, disconfirming, representing, evaluating, etc.). Analyses further reveal that most of the relation between general decision skill and numeracy is mediated by assessed risk literacy (i.e., 66% of the variance in numeracy was explained by risk literacy; see Figure 5-25). In contrast, as shown previously (Cokely et al., 2012; Låg, et al., 2012; Lindberg, & Friborg, 2013; Schapira et al., 2012; Weller et al., 2013), intelligence was only a modest-to-moderate predictor of decision making skill. This weaker association reflects the fact that fluid intelligence was largely independent of assessed risk literacy and is better explained by full-scale numeracy plus some influence of decision competence.

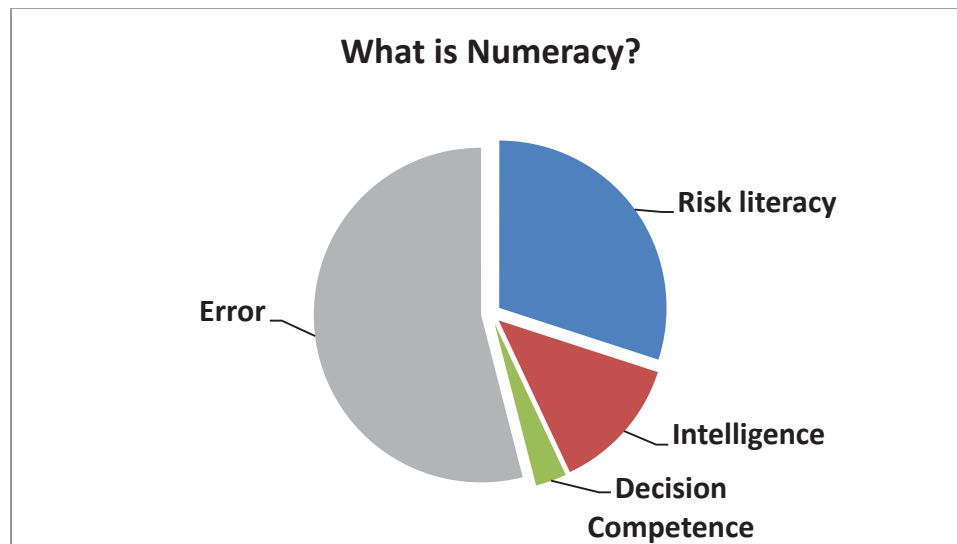


Figure 5.25. Percentage of total variance explained in numeracy by abilities and decision skills

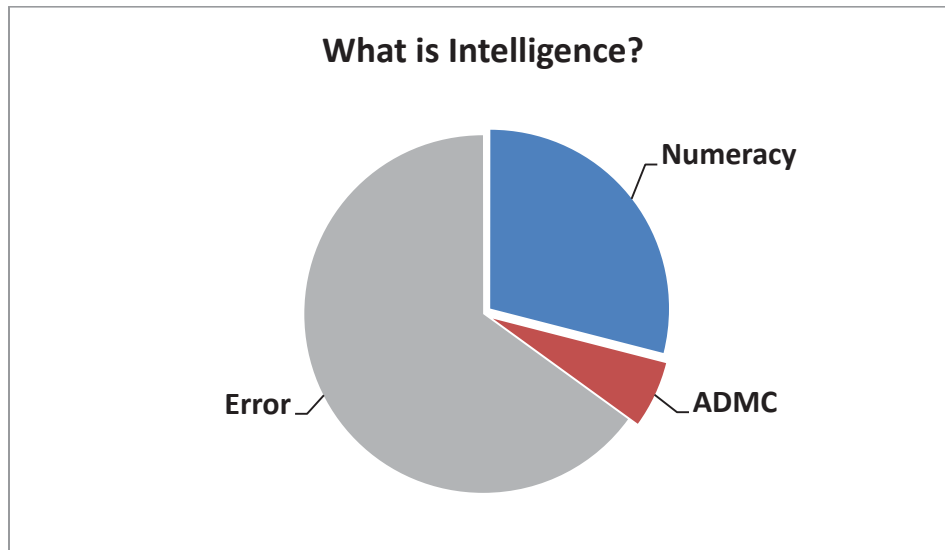


Figure 5.26. Percentage of variance explained in intelligence by abilities and decision skills

5.5.2 Emerging and Future Applications

The development and introduction of this new measurement technology (i.e., the BNT-C) is timely given the emerging need for a better theoretical understanding of the nature of numeracy and its relations to general cognitive abilities, superior decision making, and cognitive strategies (Cokely et al., 2012; Peters, 2012). With this technology, estimates of different types of decision making skills and abilities (profiles) can be provided in about 10% of the time required for comprehensive assessment (i.e., <20 minutes vs 3 hours). Once we are able to measure and understand these differences we can further develop adaptive tutoring systems and provide interactive educational and training systems. Such adaptive tutoring systems are often very effective, particularly in mathematics and sciences where adaptive tutoring systems are routinely used in

thousands of middle and high schools every year (Graesser et al., 2001; Koedinger, Aleven, Heffernan, McLaren & Hockenberry, 2004). Theoretically, these types of intelligent tutoring systems will also be powerful when attempting to train and improve risk literacy and general decision making skill.

Going forward, research is now being conducted in order to tune the BNT-C for the general population. Together with my mentors and colleagues, I am also working to create and validate an adaptive version of BNT-C for a very wide-range of skills. This adaptive modification will provide precise and robust assessment in less than 5 minutes. This kind of time savings would be valuable for many reasons including ease of use in research and clinical settings, and when used for outreach applications (e.g., www.RiskLiteracy.org). Preliminary analyses using decision tree structures were conducted and a beta-version structure of the BNT-C adaptive was developed by careful examining item parameters, correlations, and probability estimates, without any modeling fitting on the criteria. The initial BNT-C adaptive test requires only 2 to 4 questions and provides estimates of overall numeracy as well as rough estimates of component sub-skills (see Figure 5-27). The correlation between BNT-C and beta-version BNT-C adaptive was found to be .83 and the adaptive test was psychometrically similar to the full BNT-C. Preliminary analyses reveal no significant loss of predictive power as compared to the standard version. Nevertheless, the BNT-C can be assessed in a fraction of the time (about 3 minutes or 15% of the total time on average). Preliminary data and development of the BNT-C adaptive component sub-scales version is also presented below (see Figure 5-28).

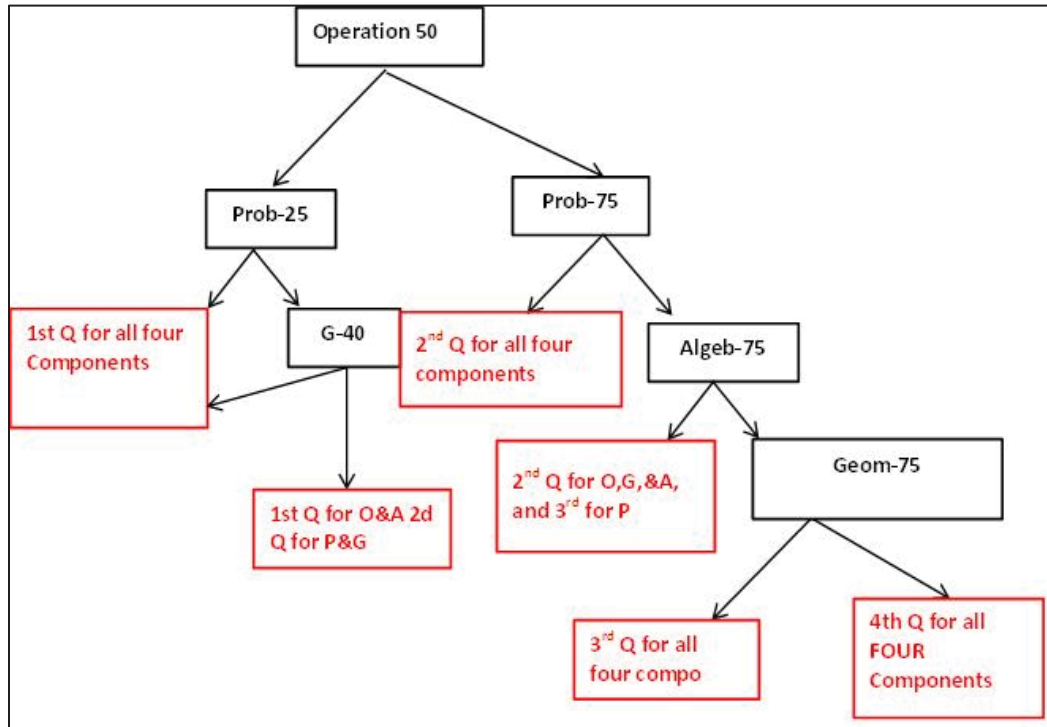


Figure 5.27. BNT-C adaptive structure

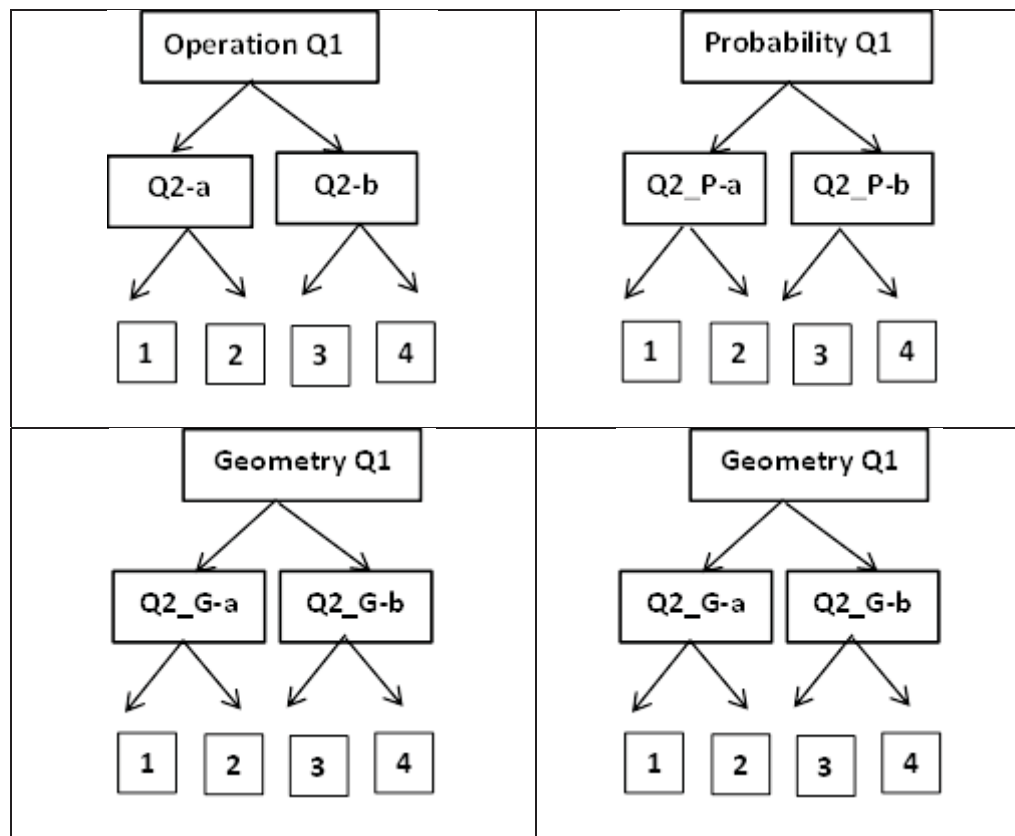


Figure 5.28. Adaptive structure for BNT-C component sub-scales.

5.5.3 Concluding Remarks

Beyond their theoretical contributions, the measurement instruments and cognitive models developed in this dissertation may help inform the design of decision support tools, risk communication efforts, and educational applications (e.g., intelligent tutors). We may also be able to leverage robust links between decision making skills and mathematical skills in order to create exercises for integrated mathematics curricula (e.g., math in context as in the Common Core State Standards) or for integrated critical

thinking and decision skill curricula (e.g., science literacy and risk literacy). More research is needed to critically evaluate the strength of the causal connections between numeracy, risk literacy and superior decision making skill, including longitudinal blind prospective studies in children and adults. Nevertheless, more than a century of research shows us that abilities and skills will only generalize to the extent that similar cognitive elements are present on *training* and *transfer* tasks because transfer requires shared elements (Thorndike & Woodworth, 1901; see also Blume, Ford, Baldwin, Huang, 2010). While many skills are highly domain-specific and unrelated to performance outside a narrow band of expertise (e.g., surgical skill is not related to managerial decision making; Ericsson, Charness, Feltovich, & Hoffman, 2006; Ericsson et al., 2007), numeracy is different because numbers are ubiquitous in our complex and uncertain world. Ultimately, this structure of our natural environment helps explain why numeracy is such a robust predictor of general decision skill. Numbers and mathematical systems are technologies created in part to help people think about uncertainty and to precisely estimate and quantify trade-offs. Numeracy is the language of reasoning and risk, and good decision making is reckoning with risk (Gigerenzer, 1998, 2002, 2012).

References

- Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6), 716-723.
- Alonso, D., & Fernandez-Berrocal, P. (2003). Irrational decisions: attending to numbers rather than ratios. *Personality and Individual Differences*, 35(7), 1537-1547.
- Arkes, H. R. (1991). Costs and Benefits of Judgment Errors - Implications for Debiasing. *Psychological Bulletin*, 110(3), 486-498.
- Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., & Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Education and Counseling*, 38(1), 33-42.
- Baker, S. F. (2010). *Calibration analysis within the cognitive and personality domains: Individual differences in confidence, accuracy, and bias. (Unpublished doctoral dissertation)*. (PhD), University of Southern Queensland, Australia.
- Banks, J., o'Dea, C., & Oldfield, Z. (2010). Cognitive Function, Numeracy and Retirement Saving Trajectories*. *The Economic Journal*, 120(548), F381-F410.
- Baron, J. (1978). Intelligence and general strategies. In G. Underwood (Eds.), *Strategies of information processing* (pp. 403–450). Academic Press.
- Baron, J. (1985). Rationality and intelligence. *New York, NY: Cambridge University Press*.
- Baron, J. (2008). Thinking and Deciding. *New York, NY: Cambridge University Press*.
- Baron, J., Badgio, P., & Gaskins, I. W. (1986). Cognitive style and its

- improvement: A normative approach. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, vol. 3 (pp. 173–220). Hillsdale, NJ: Erlbaum.
- Barton, A., Cokely, E. T., Galesic, M., Koehler, A., & Haas, M. (2009). Comparing risk reductions: On the dynamic interplay of cognitive strategies, numeracy, complexity, and format. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 347–2352). Austin, TX: Cognitive Science Society.
- Beazley, K. (1984). Education in Western Australia: Report of the Committee of Inquiry into Education in Western Australia. *Education Department of Western Australia*.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*, 395-479.
- Blume, B. D., Ford, J. K., Baldwin, T. T., & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36, 1065-1105.
- Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3), 382-398.
- Brewer, S., & Director, I. A. (2011). Adaptation of Mathematics Anxiety Rating Scale-Revised (MARS-R) for Adult Online Students.
- Bruine de Bruin, W., & Bostrom, A. (2013). Assessing what to address in science communication. *Proceedings of the National Academy of Sciences*, 110(Supplement 3), 14062-14068.
- Bruine de Bruin, W., Parker, A. M., & Fischhoff, B. (2007). Individual differences in

- adult decision-making competence. *Journal of Personality and Social Psychology*, 92(5), 938-956.
- Coben, D. (2000). Numeracy, mathematics and adult learning. In I. Gal (Ed.), *Adult Numeracy Development: Theory, research, practice* (pp. 33-50). Cresskill, NJ: Hampton Press.
- Cockcroft, W. H., & Britain, G. (1982). *Mathematics counts*: HM Stationery Office.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment & Decision Making*, 7(1).
- Cokely, E. T., Ghazal, S., Galesic, M., & Garcia- Retamero, R. (2013). How to measure risk comprehension in educated samples. In R. Garcia-Retamero & M. Galesic (Eds.), *Transparent communication of health risks: Overcoming cultural differences* (pp. 29– 52). New York: Springer.
- Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (in press). Measuring numeracy. In B. L. Anderson & J. Schulkin (Eds.), *Numerical reasoning in medical decision making*. Cambridge, UK: Cambridge University Press.
- Cokely, E. T., & Kelley, C. M. (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4(1), 20-33.
- Cokely, E. T., Kelley, C. M., & Gilchrist, A. H. (2006). Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychonomic Bulletin & Review*, 13, 991-997.

- Cokely, E. T., Parpart, P., & Schooler, L. J. (2009). *On the link between cognitive control and heuristic processes*. Paper presented at the Proceedings of the 31st Annual Conference of the Cognitive Science Society, Austin, TX.
- Cornoldi, D. L. C. (1997). Mathematics and metacognition: What is the nature of the relationship? *Mathematical cognition*, 3(2), 121-139.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.
- Crowther, G. (1959). A report of the Central Advisory Council for Education: London.
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: when people behave against their better judgment. *Journal of personality and social psychology*, 66(5), 819.
- Di Girolamo, A., Harrison, G. W., Lau, M. I., & Swarthout, J. T. (2013). Characterizing Financial and Statistical Literacy: Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*: Sage.
- Edwards, A., & Elwyn, G. (2009). *Shared decision-making in health care: Achieving evidence-based patient choice*: Oxford University Press.
- Edwards, W. (1961). Behavioral decision theory. *Annual review of psychology*, 12(1), 473-498.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *J Pers Soc Psychol*, 84(1), 5.

- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1), 98-121.
- Ericsson, K. A., Charness, N., Feltovich, P., & Hoffman, R. R. (2006). *Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological review*, 102(2), 211.
- Ericsson, K. A., Prietula, M. J., & Cokely, E. T. (2007). The making of an expert. *Harvard business review*, 85(7/8), 114.
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Med Decis Making*, 27(5), 672-680.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, 34(10), 906.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25-42.
- Gal, I. (1995). Statistical tools and statistical literacy: The case of the average. *Teaching Statistics*, 17(3), 97-99.
- Gal, I., Van Groenestijn, M., Manly, M., Schmitt, M. J., & Tout, D. (2005). Adult numeracy and its assessment in the ALL survey: A conceptual framework and

- pilot results. *Measuring adult literacy and life skills: New frameworks for assessment*, 137-191.
- Garcia-Retamero, R., & Cokely, E. T. (2011). Effective communication of risks to young adults: Using message framing and visual aids to increase condom use and STD screening. *Journal of Experimental Psychology: Applied*, 17(3), 270.
- Garcia-Retamero, R., & Cokely, E. T. (2013). Communicating health risks with visual aids. *Current Directions in Psychological Science*, 22(5), 392-399.
- Garcia-Retamero, R., & Cokely, E. (in press). The influence of skills, message frame, and visual aids on prevention of sexually transmitted diseases. *Journal of Behavioral Decision Making*.
- Garcia-Retamero, R., Cokely, E.T., & Ghazal, S. (In press). Measuring subjective graph literacy in diverse samples and cultures. *Medical Decision Making*.
- Garcia-Retamero, R., Cokely, E. T., & Ghazal, S. (In preparation). *Comparing risk literacy in 31 countries: New results from the Berlin Numeracy Test*. Manuscript in preparation.
- Garcia-Retamero, R., Wicki, B., Cokely, E. T., & Hanson, B. (2014). Factors Predicting Surgeons' Preferred and Actual Roles in Interactions with Their Patients.
- Garofalo, J., & Lester Jr, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for research in mathematics education*, 163-176.

- Ghazal, S., Cokely, E. T., & Garcia-Retamero, R. (2014). Predicting biases in very highly educated samples: Numeracy and metacognition. *Judgment and Decision Making*, 9(1), 15-34.
- Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. Simon and Schuster.
- Gigerenzer, G. (2012). Risk literacy. In J. Brockman (Ed.), *This will make you smarter: New scientific concepts to improve your thinking* (pp. 259-261). New York: Harper Perennial.
- Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological science in the public interest*, 8(2), 53-96.
- Ginsburg, L., Manly, M., & Schmitt, M. J. (2006). The Components of Numeracy. NCSALL Occasional Paper. *National Center for the Study of Adult Learning and Literacy (NCSALL)*.
- Golbeck, A. L., Ahlers-Schmidt, C. R., Paschal, A. M., & Dismuke, S. E. (2005). A definition and operational framework for health numeracy. *American journal of preventive medicine*, 29(4), 375-376.
- Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in personality*, 37(6), 504-528.
- Guadagnoli, E., & Ward, P. (1998). Patient participation in decision-making. *Social science & medicine*, 47(3), 329-339.

- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American psychologist*, 53(4), 449.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Disposition, skills, structure training, and metacognitive monitoring. *American psychologist*, 53(4), 449.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Educational Measurement: Issues and Practice*, 8(1), 35-41.
- Hertzog, C., & Robinson, A. E. (2005). Metacognition and intelligence. In O. Wilhelm & R. W. Engle, (Eds.), *Handbook of understanding and measuring intelligence* (pp. 101–121). CA: Sage publishers.
- Huff, D. (2010). *How to lie with statistics*: WW Norton & Company.
- Hunt, E., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence*, 36(1), 1-9.
- Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In S. Chaiken & Y. Trope, (Eds.), *Dual process theories in social psychology* (pp. 383–402). New York, US: The Guilford Press.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, 12(5), 852-857.
- Johnston, B., & Tout, D. (1995). Adult numeracy teaching: Making meaning in

- mathematics. *Melbourne: National Staff Development Committee for Vocational Education and Training.*
- Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, 58(9), 697.
- Kahneman, D. (2011). *Thinking, fast and slow*: Macmillan.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.
- Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, 31(4), 331-358.
- Koriat, A. (2012). The self-consistence model of subjective confidence. *Psychological Review*, 119, 80–113.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological review*, 103(3), 490.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6(2), 107.
- Kutner, M., Greenburg, E., Jin, Y., & Paulsen, C. (2006). The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483. *National Center for Education Statistics.*
- Låg, T., Bauger, L., Lindberg, M., & Friborg, O. (2013). The Role of Numeracy and Intelligence in Health-Risk Estimation and Medical Data Interpretation. *Journal of Behavioral Decision Making.*
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual

- Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, 25(4), 361-381. doi: Doi 10.1002/Bdm.752
- Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for non-nested latent class models. *Journal of Educational and Behavioral Statistics*, 22(3), 249-264.
- Lindenskov, L., & Wedege, T. (2001). *Numeracy as an analytical tool in mathematics education and research*: Publication.
- Lindskog, M., Winman, A., & Juslin, P. (2013). Are there rapid feedback effects on Approximate Number System acuity? *Frontiers in human neuroscience*, 7.
- Lipkus, I. M., & Peters, E. (2009). Understanding the role of numeracy in health: proposed theoretical framework and practical insights. *Health Educ Behav*, 36(6), 1065-1081. doi: 10.1177/1090198109341533
- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General Performance on a Numeracy Scale among Highly Educated Samples. *Medical Decision Making*, 21(1), 37-44.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*: Routledge.
- Metcalf, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174-179.
- Mitchum, A. L., & Kelley, C. M. (2010). Solve the problem first: constructive solution strategies can influence the accuracy of retrospective confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3),

699.

- Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 407-423). New York: Guilford
- Moxley, J. H., Anders Ericsson, K., Charness, N., & Krampe, R. T. (2012). The role of intuition and deliberative thinking in experts' superior tactical decision-making. *Cognition*, 124(1), 72-78.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The psychology of learning and motivation*, 26, 125-141.
- Nelson, W., Reyna, V. F., Fagerlin, A., Lipkus, I., & Peters, E. (2008). Clinical implications of numeracy: theory and practice. *Annals of behavioral medicine*, 35(3), 261-274.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2), 175.
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1-18.
- Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, 25(4), 390-401.
- Orrill, C. H. (2001). Building technology-based, learner-centered classrooms: The evolution of a professional development framework. *Educational Technology*

- Research and Development*, 49(1), 15-34.
- Pachur, T., & Galesic, M. (2013). Strategy Selection in Risky Choice: The Impact of Numeracy, Affect, and Cross-Cultural Differences. *Journal of Behavioral Decision Making*, 26(3), 260-271.
- Parker, A. M., & Fischhoff, B. (2005). Decision-making competence: External validation through an individual-differences approach. *Journal of Behavioral Decision Making*, 18(1), 1-27.
- Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences*: Macmillan.
- Peters, E. (2012). Beyond comprehension the role of numeracy in judgments and decisions. *Current Directions in Psychological Science*, 21(1), 31-35.
- Peters, E., Dieckmann, N., Dixon, A., Hibbard, J. H., & Mertz, C. K. (2007). Less is more in presenting quality information to consumers. *Med Care Res Rev*, 64(2), 169-190. doi: 10.1177/10775587070640020301
- Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*, 3(6), 435-448.
- Peters, E., & Levin, I. P. (2008). Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. *Judgment and Decision Making*, 3(6), 435-448.
- Peters, E., Meilleur, L., & Tompkins, M. K. Numeracy and the affordable care act: Opportunities and challenges. *Chapter prepared for the Roundtable on Health*

Literacy, Institute of Medicine.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C., Mazzocco, K., & Dickert, S. (2006).

Numeracy and decision making. *Psychol Sci*, 17(5), 407-413.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological review*, 117(3), 864.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and uncertainty*, 23(1), 5-32.

Reyna, V. F., & Brainerd, C. (1991). Fuzzy-trace theory and children's acquisition of mathematical and scientific concepts. *Learning and Individual Differences*, 3(1), 27-59.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences*, 7(1), 1-75.

Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin*, 135(6), 943-973. doi: 10.1037/a0017327

Riege, A. H., & Teigen, K. H. (2013). Corrigendum to “Additivity neglect in probability estimates: Effects of numeracy and response format” [Organizational Behavior and Human Decision Processes 121 (2013) 41–52]. *Organizational Behavior and Human Decision Processes*, 121(2), 278.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25.

Schapira, M. M., Walker, C. M., Cappaert, K. J., Ganschow, P. S., Fletcher, K. E.,

- McGinley, E. L., . . . Jacobs, E. A. (2012). The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. *Med Decis Making*, 32(6), 851-865.
- Schapira, M. M., Walker, C. M., & Sedivy, S. K. (2009). Evaluating existing measures of health numeracy using item response theory. *Patient Educ Couns*, 75(3), 308-314. doi: 10.1016/j.pec.2009.03.035
- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Handbook of research on mathematics teaching and learning*, 334-370.
- Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, 45(4), 258-266.
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of internal medicine*, 127(11), 966-972.
- Sheridan, S. L., Pignone, M. P., & Lewis, C. L. (2003). A randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *Journal of General Internal Medicine*, 18(11), 884-892.
- Sherrod, P. H. (2003). DTREG predictive modeling software. *Software available at* <http://www.dtreg.com>.

- \Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.
- Stankov, L. (2000). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*(1), 35-51.
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*(4), 961-976. doi: 10.1037/a0012546
- Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. *The Oxford handbook of thinking and reasoning, 343-365*.
- Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences, 23*(05), 701-717.
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *J Pers Soc Psychol, 94*(4), 672.
- Steen, L. A. (1990). *On the shoulders of giants: New approaches to numeracy*: National Academies Press.
- Steen, L. A. (2001). Mathematics and numeracy: Two literacies, one language. *The mathematics educator, 6*(1), 10-16.
- Sternberg, R. J. (1977). Component processes in analogical reasoning. *Psychological review, 84*(4), 353.

- Toplak, M. E., West, R. F., & Stanovich, K. E. (in press a). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking and Reasoning*.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (in press b). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology*.
- Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive psychology*, 63(3), 107-140.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.
- Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*: Springer.
- Vigneau, F., Caissie, A. F., & Bors, D. A. (2005). Eye movement analysis demonstrates strategic influence on intelligence. *Intelligence*, 34, 261–272.
- Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach. *Journal of Behavioral Decision Making*, 26(2), 198-212. doi: 10.1002/bdm.1751
- Wells, H. G. (1903). *Mankind in the Making*: Chapman & Hall, Id.
- Woller-Carter, M. M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). Communicating and Distorting Risks with Graphs: An Eye-Tracking Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*,

56(1), 1723-1727.

Woodworth, R. S., & Thorndike, E. (1901). The influence of improvement in one mental function upon the efficiency of other functions. (I). *Psychological review*, 8(3), 247.

Yamagishi, K. (1997). When a 12.86% mortality is more dangerous than 24.14%: implications for risk communication. *Applied Cognitive Psychology*, 11(6), 495-506.

Zikmund-Fisher, B. J., Smith, D. M., Ubel, P. A., & Fagerlin, A. (2007). Validation of the Subjective Numeracy Scale: effects of low numeracy on comprehension of risk communications and utility elicitations. *Med Decis Making*, 27(5), 663-671. doi: 10.1177/0272989X07303824.

APPENDICES

Appendix A

Item analysis for pilot data; subscale analysis with 20 questions each (n=30)

Table A-6-1. *Item analysis for 1st numeracy component “Number Sense” (20 items)*

Items	Difficulty	Item-Total Correlation
N-1	0.91	0.49
N-2	0.38	0.38
N-3	0.88	0.61
N-4	0.45	0.32
N-5	0.59	0.63
N-6	0.33	0.46
N-7	0.22	0.36
N-8	0.73	0.67
N-9	0.63	0.61
N-10	0.59	0.53
N-11	0.38	0.60
N-12	0.89	0.46
N-13	0.41	0.53
N-14	0.56	0.68
N-15	0.56	0.19
N-16	0.25	0.40
N-17	0.47	0.47
N-18	0.28	-0.28
N-19	0.34	0.35
N-20	0.41	0.33

Table A-2. *Item analysis for 2nd numeracy component “Probability and Statistics” (20 items)*

Items	Difficulty	Item-Total Correlation
P-1	0.63	0.11
P-2	0.41	0.42
P-3	0.59	0.59
P-4	0.53	0.27
P-5	0.5	0.42
P-6	0.81	0.35
P-7	0.53	0.42
P-8	0.22	0.07
P-9	0.56	0.51
P-10	0.47	0.40
P-11	0.19	-0.24
P-12	0.63	0.30
P-13	0.66	0.56
P-14	0.66	0.14
P-15	0.19	-0.16
P-16	0.25	0.01
P-17	0.47	0.37
P-18	0.31	0.18
P-19	0.34	-0.21
P-20	0.03	0.30

Table A-3. *Item analysis for 3rd numeracy component “Geometry” (20 items)*

Items	Difficulty	Item-Total Correlation
G-1	0.66	0.63
G-2	0.38	0.23
G-3	0.5	0.56
G-4	0.16	-0.07
G-5	0.47	0.47
G-6	0.31	0.52
G-7	0.31	0.19
G-8	0.38	0.26
G-9	0.25	0.45
G-10	0.09	-0.28
G-11	0.16	-0.40
G-12	0.19	0.34
G-13	0.22	0.52
G-14	0.41	0.18
G-15	0.47	0.70
G-16	0.47	0.42
G-17	0.34	0.33
G-18	0.28	0.32
G-19	0.28	0.57
G-20	0.31	0.52

Table A-4. *Item analysis for 4th numeracy component “Algebra” (20 items)*

Items	Difficulty	Item-Total Correlation
A-1	0.15	0.57
A-2	0.72	0.21
A-3	0.78	0.32
A-4	0.34	0.26
A-5	0.41	0.31
A-6	0.53	0.44
A-7	0.64	0.59
A-8	0.44	0.51
A-9	0.25	0.52
A-10	0.13	0.08
A-11	0.13	-0.25
A-12	0.53	0.41
A-13	0.56	0.42
A-14	0.28	0.40
A-15	0.34	0.50
A-16	0.06	-0.17
A-17	0.30	0.52
A-18	0.28	0.46
A-19	0.09	0.41
A-20	0.34	0.59

Appendix B

Table B-1. Item Analysis for final comprehensive test of numeracy (20 items) with its four components (5 items each)

	Difficulty	Discrimination	P(x=1 z=0)
N2	0.345274326	1.0788128	0.4079404
N3	-0.902463257	1.2876966	0.7617139
N7	0.033066738	1.9563154	0.4838334
N9	0.987177759	0.9457456	0.2821911
N4	-1.168893689	1.2720540	0.8156119
P1	-0.505423008	1.4457540	0.6749627
P5	-0.589510559	1.6639949	0.7272952
P8	0.271457880	1.0991314	0.4259565
P9	1.258125098	0.8530131	0.2547955
P10	-0.194238916	1.2271351	0.5593088
G1	-1.156569045	1.3746363	0.8305967
G2	-0.147705992	1.6931638	0.5621988
G7	-0.003867466	1.2051337	0.5011652
G8	-0.972052729	0.8827572	0.7022607
G10	-0.357721671	1.4405131	0.6260487
A2	-0.629718368	2.0252405	0.7816540
A5	-1.383396749	1.3264620	0.8623590
A6	0.164776781	1.4497276	0.4405620
A8	0.945923094	1.7099954	0.1655466
A10	0.388807919	2.0597189	0.3098469

Appendix C

Model selection for all four subscales (10 items in each scale)

1) Operations and number sense:

	AIC	BIC	Log.lik	LR	<i>df</i>	<i>P</i>
One parameter	1849	1882	-913			
Two- parameter	1841	1903	-900	25.7	9	.002
Three- parameter	1854	1947	-897	6.9	10	.7

Model with lower AIC value represents better fit to the data

2) Probability

	AIC	BIC	Log.lik	LR	<i>df</i>	<i>P</i>
One parameter	1883	1917	-931			
Two- parameter	1862	1925	-911	38.8	9	.001
Three- parameter	1880	1973	-910	2.6	10	.9

Model with lower AIC value represents better fit to the data

Geometry

	AIC	BIC	Log.lik	LR	<i>df</i>	<i>P</i>
One parameter	2012	2046	-995			
Two- parameter	1990	2053	-975	39.8	9	..001
Three- parameter	2007	2100	-973	3.6	10	.9

Model with lower AIC value represents better fit to the data

3) Algebra

	AIC	BIC	Log.lik	LR	<i>df</i>	<i>P</i>
One parameter	1802	1836	-890			
Two- parameter	1798	1860	-879	22.6	9	.007
Three- parameter	1811	1903	-875	7.1	10	.7

Model with lower AIC value represents better fit to the data

Appendix D

Process of collecting Ecological Risk Literacy Tasks

Sampling the risky scenarios from natural ecology is a very broad topic and challenging indeed. I narrowed down my search to two major domains that are most relevant and important in the real life situations'; i.e., medical and financial decisions. For medical risky decision making, (i.e. to see how people understand risks and benefits of treatments and procedures), I chose diseases that are more prevalent in USA (i.e. heart, cancer, arthritis & obesity) (Source: CDC/NHS, National Vital Statistics System) and searched the most commonly visited/used websites (CDC, American Heart Association, and American Cancer Society) for each of the disease. The majority of the information found on these websites and in medical brochures was qualitative and descriptive in nature, with very few probability or proportions terms/quantities (see Table below for sample summary of my search in medical decisions). Clinical trial data was also sampled that was available on these websites for people so they can use the information to understand risks and benefits of certain treatments and procedures.

Summary table of brochures in Health decision making From the top three websites; CDC, American Heart Association, & American Cancer Association	
Heart diseases	
Women and Cholesterol	Qualitative - (increase or decrease levels; good or bad; therapy preventer not prevent
Heart Disease Behavior	Qualitative - Tobacco, diet, physical activity, obesity, and alcohol effects heart disease
Heart Disease Conditions	Qualitative -when and how blood cholesterol, blood pressure and diabetes lead to heart attack
Heart Disease Prevention	Qualitative - change your lifestyle and treat it properly with regular checks and medication
High Blood Pressure Behavior	Qualitative and quantitative - blood pressure low and high quantities/levels
Blood Pressure & diet	Qualitative - appropriate amounts of sodium and potassium, what are right diets
Blood Pressure Conditions	Qualitative - conditions such as prehypertension and diabetes that effect high blood pressure
High Blood Pressure Facts	Qualitative & Quantitative - high blood pressure, heart disease by age, gender, race and ethnicity
Blood Pressure & Heredity	Qualitative - factors that are unable to change that affect high blood pressure
Blood Pressure- Ethnicity	Qualitative - theories about why African Americans are at a higher risk
Blood Pressure and Women	Qualitative - high blood pressure during child bearing years
Your Risk for High Blood Pressure	Quantitative and qualitative - all risk factors
Arthritis CDC	
Data and Statistics National Data	Qualitative & Quantitative: %ages and data of women and men that have arthritis – body weight
Basics - Fibromyalgia	Qualitative - back ground - prevalence - mortality - ambulatory - hospitalization - costs
Basics - General	Qualitative - back ground - prevalence - mortality - ambulatory - hospitalization - costs
Basics - Gout	Qualitative - background - prevalence - incidence - hospitalization - costs
Osteoarthritis	Qualitative - back ground - prevalence - mortality - ambulatory – hospitalization- cost - incidence
Basics - Rheumatoid	Qualitative - back ground - treatment - risk

	factors- prevalence - incidence- mortality
Basics - Risk Factors	Qualitative - non modifiable and modifiable risk factors leading to Arthritis
Data and Statistics	Qualitative - prevalence with age, race, gender & overweight- limitations-costs-mortality
Diabetes and Obesity	
Your Risk for Diabetes	Qualitative: indicating low and high risk people
Diabetes and Obesity	Qualitative and quantitative
Lifetime Risks	Qualitative
Long Term Trends in Diabetes	Qualitative and quantitative
Maps Diabetes Trends	Maps and charts showing trends in obesity, make sense of prevalence and future
Diabetes Obesity Trends	Quantitative: showing preset past and future number with obesity
High Blood Pressure	How blood pressure and diabetes are related
Blood Cholesterol & Lipids	Qualitative - Race - Good and bad cholesterol
National Diabetes Fact Sheet	Quantitative - Diagnosed and undiagnosed cases, prevalence as per age, gender, race, mortality...
* Quantitative means %ages and numbers of victims/at risk. NO data state probabilities or proportions to estimate level of risk	

Financial decision scenarios were selected using the aforementioned systematic approach. Retirement, student loans, and credit card decisions were highlighted and searched due to their importance. I also collected scenarios on consumer and environmental risks and decisions. Because the test is long, and not validated yet, I will use one scenario from medical decisions, and one from financial decisions to be included as criterion for study two (See the scenarios in appendix F)

Appendix E

Material used in study one

Comprehensive Test of Numeracy (20 items) with four components

Part I; Operations and number sense

1) If a sack of dried dog food feeds 4 dogs or 5 puppies for one week, then 5 sacks of the food will feed 15 puppies and how many dogs? **8 dogs**

2) The odometer of a new automobile functions improperly and registers only 2 miles for every 3 miles driven. If the odometer indicates 48 miles, how many miles has the automobile actually been driven?

- a) 144
- b) 72**
- c) 64
- d) 32
- e) 24

3) Helpers are needed to prepare for the fete. Each helper can make either 2 large cakes per hour or 35 small cakes per hour. The kitchen is available for 3 hours and 20 large cakes and 700 small cakes are needed. How many helpers are required?

- a) 10**
- b) 15
- c) 20
- d) 25
- e) 30

4) If United States imports increased 20 percent and exports decreased 10 percent during a certain year, the ratio of imports to exports at the end of the year was how many times the ratio at the beginning of the year?

- a) 12/11
- b) 4/3**
- c) 11/8
- d) 3/2
- e) 2

5) n and p are integers greater than 1

$5n$ is the square of a number

$75np$ is the cube of a number.

The smallest value for $n + p$ is

a) 14

b) 18

c) 20

d) 30

e) 50

Part II; Probability and statistics

1) Imagine that you are throwing 6 dice up in the air. What is the probability that all of them would land on even numbers?

a) $1/432$

b) $3/216$

c) $1/64$

d) $3/6$

e) $1/21$

2) Imagine you are drawing a picture, and are missing 2 spots you want to color. There are 7 colors to choose from. What's the probability that both spots end up colored orange?

a) $1/49$

b) $2/49$

c) $1/7$

d) $2/7$

e) $6/7$

3) Imagine you are throwing 8 dice up in the air. What's the probability that half will land on an even number, while the other half land on 1?

a) $1/10368$

b) $1/20736$

c) $1/432$

d) $1/1728$

e) $1/6$

4) Phil is holding 4 cards in his hand: 8 of clubs, 5 of hearts, king of hearts, and ace of diamonds. If he places them on a table in random order, what is the probability that the first and last cards will both be hearts?

a) $\frac{1}{2}$

b) $\frac{1}{3}$

c) $\frac{1}{4}$

d) $1/6$

e) $1/8$

5) n is an integer chosen at random from the set $\{5, 7, 9, 11\}$

p is chosen at random from the set $\{2, 6, 10, 14, 18\}$

What is the probability that $n + p = 23$?

a) 0.1

b) 0.2

c) 0.25

d) 0.3

e) 0.4

Part III; Geometry and shapes

1) The slope of a line through points $P(1, 1)$ and $Q(k, 7)$ is $3/2$. What is the value of k ?

a) 4

b) 5

c) 6

d) 7

e) 8

2) $A, B, C,$ and D are points on a line, with D the midpoint of BC . The lengths of AB, AC and

BC are 10, 2, and 12, respectively. What is the length of AD ?

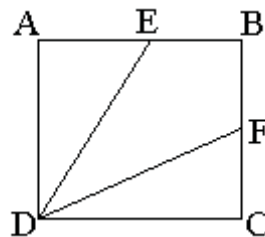
a) 2

b) 4

c) 6

d) 10

e) 12



3) $ABCD$ is a square of side 3, and E and F are the mid points of sides AB and BC respectively.

What is the area of the quadrilateral $EBFD$?

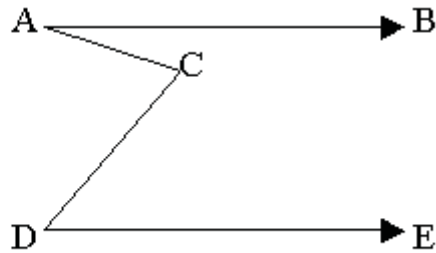
a) 2.25

b) 3

c) 4

d) 4.5

e) 6

*(figure not to scale)*

4) AB and DE are parallel. Angle $BAC = 30^\circ$, angle $CDE = 50^\circ$. What is the measure of angle ACD?

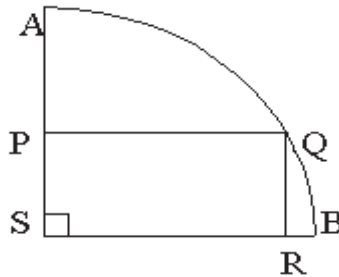
a) 100

b) 90

c) 80

d) 70

e) cannot be determined from the information



5) ASB is a quarter circle. PQRS is a rectangle with sides $PQ = 8$ and $PS = 6$. What is the length of the arc AQB?

f) 5π g) 10π

h) 25

i) 14

j) 28

Part IV; Pattern functions and Algebra:

1) If the equation of a line p in the coordinate plane is $y = 3x + 2$, what is the equation of line q which is a reflection of line p in the x-axis?

a) $y = -3x + 2$ b) $y = -3x - 2$

- c) $y = 3x - 2$
- d) $y = -1/3x - 5$
- e) $y = -1/3x + 5$

2) $(3x + 2)(2x - 5) = ax^2 + kx + n$.

What is the value of $a - n + k$?

- a) 5
- b) 8
- c) 9
- d) 10
- e) 11

3) If $f(x) = x^2 - 3$, where x is an integer, which of the following could be a value of $f(x)$?

- I 6
- II 0
- III -6

- a) I only
- b) I and II only
- c) II and III only
- d) I and III only
- e) I, II and III

4) Six years ago Anita was P times as old as Ben was. If Anita is now 17 years old, how old is

Ben now in terms of P ?

- a) $11/P + 6$
- b) $P/11 + 6$
- c) $17 - P/6$
- d) $17/P$
- e) $11.5P$

5) If x / y is an integer, which of the following statements must be true?

- a) both x and y are integers
- b) x is an integer
- c) either x or y is negative
- d) y / x is an integer
- e) $x = ny$ where n is an integer

Demographics:

Age -----

Gender-----

Major -----

SAT/ACT scores -----

SAT math scores -----

Time in completing the test-----

Appendix F

Material Used in Study 2

a) Adult Decision Making Competence (ADMC)	162
b) Decision Outcome Inventory (DOI)	206
c) Ecological Risky Decision Tasks	211
d) Paradigmatic risky decisions and Intertemporal choice questions	140
1) Intertemporal Choice Tasks	
3) Risk Preferences (Lotteries)	
5) Ratio Bias/Denominator Neglect	
e) Cognitive Reflection Test (CRT) (Fredrick, 2005)	162
f) Berlin Numeracy Test (Cokely et al., 2012)	143
g) Schwartz three items numeracy test (Schwartz et al., 1997)	144
h) Lipkus 8 items numeracy test (Lipkus et al., 2001)	144
i) Abbreviated numeracy scale (Weller et al., 2012)	145
j) Numeracy Understanding in Medicine Instrument (NUMi) (Schapira et al., 2012)	145
k) Mathematics Anxiety Rating Scale Revised (MARS-R)	151
l) Personality short test	152
m) Raven's Advance Progressive Matrices	153
n) Berlin Numeracy components Test (BNT-C)	

a) Adult Decision Making Competence (ADMC)**Instructions:**

Each of the following problems presents a choice between two options. Each problem is presented with a scale ranging from 1 (representing one option) through 6 (representing the other option). For each item, please circle the number on the scale that best reflects your relative preference between the two options.

Problem 1

Imagine that recent evidence has shown that a pesticide is threatening the lives of 1,200 endangered animals. Two response options have been suggested:

If Option A is used, 600 animals will be saved for sure.

If Option B is used, there is a 75% chance that 800 animals will be saved, and a 25% chance that no animals will be saved.

Which option do you recommend to use?

1	2	3	4	5	6
Definitely would					Definitely would
choose A					choose B

Problem 2

Because of changes in tax laws, you may get back as much as \$1200 in income tax. Your accountant has been exploring alternative ways to take advantage of this situation. He has developed two plans:

If Plan A is adopted, you will get back \$400 of the possible \$1200.

If Plan B is adopted, you have a 33% chance of getting back all \$1200, and a 67% chance of getting back no money.

Which plan would you use?

1	2	3	4	5	6
Definitely would					Definitely would
choose A					choose B

Problem 3

Imagine that in one particular state it is projected that 1000 students will drop out of school during the next year. Two programs have been proposed to address this problem, but only one can be implemented. Based on other states' experiences with the programs, estimates of the outcomes that can be expected from each program can be made. Assume for purposes of this decision that these estimates of the outcomes are accurate and are as follows:

If Program A is adopted, 400 of the 1000 students will stay in school.

If Program B is adopted, there is a 40% chance that all 1000 students will stay in school and 60% chance that none of the 1000 students will stay in school.

Which program would you favor for implementation?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 4

Imagine that the U.S. is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a 33% chance that 600 people will be saved, and a 67% chance that no people will be saved.

Which program do you recommend to use?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 5

Imagine that your doctor tells you that you have a cancer that must be treated. Your choices are as follows:

Surgery: Of 100 people having surgery, 90 live through the operation, and 34 are alive at the end of five years.

Radiation therapy: Of 100 people having radiation therapy, all live through the treatment, and 22 are alive at the end of five years.

Which treatment would you choose?

1	2	3	4	5	6
Definitely would choose surgery				Definitely would choose radiation	

Problem 6

Imagine that your client has \$6,000 invested in the stock market. A downturn in the economy is occurring. You have two investment strategies that you can recommend under the existing circumstances to preserve your client's capital.

If strategy A is followed, \$2,000 of your client's investment will be saved.

If strategy B is followed, there is a 33% chance that the entire \$6,000 will be saved, and a 67% chance that none of the principal will be saved.

Which of these two strategies would you favor?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 7

Imagine a hospital is treating 32 injured soldiers, who are all expected to lose one leg. There are two doctors that can help the soldiers, but only one can be hired:

If Doctor A is hired, 20 soldiers will keep both legs.

If Doctor B is hired, there is a 63% chance that all soldiers keep both legs and a 37% chance that nobody will save both legs.

Which doctor do you recommend?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 4

As R&D manager, one of your project teams has come to you requesting an additional \$100,000 in funds for a project you instituted several months ago. The project is already behind schedule and over budget, but the team still believes it can be successfully completed. You currently have \$500,000 remaining in your budget unallocated, but which must carry you for the rest of the fiscal year. Lowering the balance by an additional \$100,000 might jeopardize flexibility to respond to other opportunities.

Evaluating the situation, you believe there is a fair chance the project will not succeed, in which case the additional funding would be lost; if successful, however, the money would be well spent. You also noticed that of the projects undertaken by this team, 30 of the last 50 have been successful.

What is the likelihood you would fund the request?

1	2	3	4	5	6
Very unlikely					Very likely

Problem 5

Suppose a student got 90% correct in the mid-term exam and 70% correct in the final-term exam, what would be your evaluations of this student's performance?

1	2	3	4	5	6
Very poor					Very good

Problem 6

Imagine that a woman parked illegally. After talking to her, you believe that there is a 20% chance that she did not know she parked illegally.

With this in mind, how much of a fine do you believe this woman deserves?

1	2	3	4	5	6
Minimum fine					Maximum fine

Problem 7

Imagine that a new technique has been developed to treat a particular kind of cancer. This technique has a 50% chance of success, and is available at the local hospital.

A member of your immediate family is a patient at the local hospital with this kind of cancer. Would you encourage him or her to undergo treatment using this technique?

1. Do you think it is sometimes OK ...
... to steal under certain circumstances?
Yes No
2. Do you think it is sometimes OK ...
... to smoke cigarettes?
Yes No
3. Do you think it is sometimes OK ...
... to commit a crime which could put you in jail?
Yes No
4. Do you think it is sometimes OK ...
... to keep things you find in the street?
Yes No
5. Do you think it is sometimes OK ...
... to experiment with marijuana?
Yes No
6. Do you think it is sometimes OK ...
... to use your fists to resolve a conflict?
Yes No
7. Do you think it is sometimes OK ...
... to drink and drive?
Yes No
8. Do you think it is sometimes OK ...
... to yell and argue to solve a conflict?
Yes No
9. Do you think it is sometimes OK ...
... not to hold the door open for people?
Yes No
10. Do you think it is sometimes OK ...
... not to tell the police when you witness a crime?
Yes No
11. Do you think it is sometimes OK ...
... not to give directions to someone who is lost?
Yes No
12. Do you think it is sometimes OK ...

- ... not to be on time for appointments?
Yes No
13. Do you think it is sometimes OK ...
... not to return something you borrowed?
Yes No
14. Do you think it is sometimes OK ...
... not to keep secrets that a friend told you?
Yes No
15. Do you think it is sometimes OK ...
... not to return phone calls right away?
Yes No
16. Do you think it is sometimes OK ...
... not to spend time with friends in need?
Yes No

Instructions:

This survey presents true/false questions about various aspects of everyday life. Please indicate, for each statement, whether you believe it to be true or false, by circling the “true” or “false”. You may think that some items do not have a clear-cut answer. For those items, please try to give the answer that would be true in general, or in most cases.

Please read through the following examples to find out more about this survey.

Example 1:**Pittsburgh's hockey team is the Bruins.**

We want you to do two things:

First, answer the question. In this example, you might think “*No, it's the Penguins. So the statement is FALSE.*” Then you would circle ‘False’.

Pittsburgh's hockey team is the Bruins.

This statement is [True / False].

Second, think about how sure you are of your answer. Give a number from 50% to 100%. In other words, what is the percent chance that you are right? Circle one of the numbers on the scale.

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

If your answer is a total guess, circle 50%. This means that there is a 50% chance that you are right, and a 50% chance that you are wrong. If you are absolutely sure, circle 100%. If you aren't sure, then circle a number in between, to show how sure you are.

In this example, you might think "*I'm absolutely sure it's false, so 100%.*" So you would circle 100%.

Pittsburgh's hockey team is the Bruins.

This statement is [True / False].

50%

60%

70%

80%

90%

100%

just guessing

absolutely sure

Please read the examples below. They show answers given by other people. Read them closely, and make sure you understand their answers.

Example 2:

Thanksgiving Day is on the fourth Thursday of November.

- *Yes, I think that's when Thanksgiving is. I would say TRUE.*
- *I'm pretty sure, but it might be on the third Thursday of November, so 80%.*

Your answer would look like this:

Thanksgiving Day is on the fourth Thursday of November.

This statement is [True / False].

50%

60%

70%

80%

90%

100%

just guessing

absolutely sure

Example 3:

Amman is the capital of Jordan.

- *I really don't know, so I'll just take a guess. I'll say, uh, TRUE.*
- *I'm guessing, so 50%.*

Your answer would look like this:

Amman is the capital of Jordan.

This statement is [True / False].

50%

60%

70%

80%

90%

100%

just guessing

absolutely sure Example 4:

- *Oh yes it does! I think it's one of the rivers. So that's FALSE.*
- *I'm almost positive that's false, so I'll say 90%.*

This statement is [True / False].

This statement is [True/ False].

50% 60% 70% 80% 90% 100%
just guessing absolutely sure

4. After a fight with your partner, you should not focus on who was to blame.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

5. There is no way to improve your memory.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

6. The grace period on your credit card is the amount of time you do not have to pay interest on outstanding payments.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

7. Red wine stains are easier to remove than beer stains.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

8. Muscles do not burn calories when you are at rest.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

9. Alcohol causes dehydration.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

10. Problems with in-laws contribute to more than 30% of divorces.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

11. Homosexual couples are not legally allowed to adopt.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

12. A promotion means that you will get a more satisfying job.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

13. IRS forms are available on-line.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

14. Procrastination is worse when you work in a cluttered environment.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

15. A venture capital fund invests in new businesses by providing startup capital.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

16. It is wise to handle all negotiations yourself, even if your opponent uses a lawyer.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

17. Carbohydrates are fattening no matter how much you eat of them.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

18. Young people face few stereotypes when looking for a job.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

19. It can be instructive for children to see their parents resolve a fight.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

20. There are nonprofit organizations that help people with debt counseling.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

21. Assertive behavior makes your brain experience an increase in pleasure.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

22. Credit card companies can offer lower payments if you can come up with a lump sum settlement.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

23. Contracting a sexually transmitted disease is not an automatic sign that your partner has had an affair.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

24. Some sexually transmitted diseases can cause infertility.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

25. Self-employed people pay the same amount of taxes as people who work for an employer.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

26. When buying a new home, there is little need to have it inspected before you buy it.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

27. Creating a routine is an important step in getting unpleasant work done.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

28. Once you have experienced an event, your memory of it can not be changed.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

29. Meditation slows the heart rate.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

30. If you get into an auto accident, let the other person take the lead in handling the details.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

31. There is no way you can negotiate a lower rate with a credit card company.

This statement is [True / False].

50%	60%	70%	80%	90%	100%
just guessing					absolutely sure

The following examples use the table above. Please read each carefully.

Example 1. Chris selects the DVD player with the highest rating in Programming Options.

Which **one** of the presented DVD player would Chris prefer? A

Example 2. Chris only wants a DVD player with a sound quality that is rated higher than 4.

Which **one** of the presented DVD player would Chris prefer? none

Example 3. Chris only wants the best in Picture Quality.

Which **two** of the presented DVD players would Chris prefer? A, and

B

The following questions are about other people choosing between DVD players, like the ones above. **Please read each question carefully, because they ask for different answers.** For each question, think about how each person makes their choice, then pick the DVD they choose. But be careful, because the DVD players will change from question to question.

Very Low 1	Low 2	Medium 3	High 4	Very High 5
---------------	----------	-------------	-----------	----------------

Question 1:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	5	4	2	1	\$369
	B	5	5	3	3	\$369
	C	5	2	4	4	\$369
	D	1	5	5	3	\$369
	E	4	5	1	1	\$369

Brian selects the DVD player with the highest number of ratings greater than “Medium”

Which **one** of the presented DVD players would Brian prefer? _____

Question 2:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	2	5	5	5	\$369
	B	5	4	4	5	\$369

C	5	3	2	5	\$369
D	3	5	2	2	\$369
E	4	4	4	5	\$369

Sally first selects the DVD players with the best Sound Quality. From the selected DVD players, she then selects the best on Picture Quality. Then, if there is still more than one left to choose from, she selects the one best on Programming Options.

Which **one** of the presented DVD players would Sally prefer? _____

Very Low	Low	Medium	High	Very High
1	2	3	4	5

Question 3:

		Features				Price
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	
DVD	A	3	1	2	5	\$369
	B	5	5	3	2	\$369
	C	4	3	3	3	\$369
	D	5	5	5	4	\$369
	E	2	5	4	4	\$369

Pat doesn't want to read through the entire table. He decides to read the table row by row until he finds the very first DVD player that has no ratings below "Medium." He will just choose that DVD player.

Which **one** of the presented DVD players would Pat prefer? _____

Question 4:

		Features				Price
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	
DVD	A	3	5	5	1	\$369
	B	1	2	1	2	\$369
	C	5	5	4	4	\$369
	D	5	3	4	2	\$369
	E	4	5	2	2	\$369

LaToya only wants a DVD player that got a “Very High” rating on Reliability of Brand. Which **one** of the presented DVD players LaToya prefer? _____

Very Low	Low	Medium	High	Very
High 1	2	3	4	5

Question 5:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	5	5	5	3	\$369
	B	3	5	4	5	\$369
	C	5	2	2	4	\$369
	D	5	1	2	5	\$369
	E	4	2	4	5	\$369

From the DVD players with the best available Picture Quality, Tricia selects the DVD players with the lowest number of ratings below “Medium.” If there is more than one DVD player left to choose from, she then picks the one that has the best rating on “Reliability of Brand.”

Which **one** of the presented DVD players would Tricia prefer? _____

Question 6:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	3	1	5	2	\$369
	B	1	2	1	2	\$369
	C	5	4	3	1	\$369
	D	4	2	3	3	\$369
	E	4	4	2	4	\$369

Lisa wants the DVD player with the highest average rating across features.

Which **one** of the presented DVD players would Lisa prefer? _____

Question 7:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	5	3	5	5	\$369
	B	2	5	4	1	\$369
	C	4	5	2	3	\$369
	D	3	5	3	1	\$369
	E	3	5	3	4	\$369

Andy wants the DVD player with the highest average rating he can get while still making sure to keep the best rating on Sound Quality.

Which **one** of the presented DVD players would Andy prefer? _____

Question 8:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	5	4	5	3	\$369
	B	5	4	1	2	\$369
	C	3	3	5	5	\$369
	D	5	5	1	2	\$369
	E	3	5	1	3	\$369

Shane wants no DVD players that score below “Medium” on Picture Quality, no DVD players that score below “Medium” on Sound Quality, and no DVD players that score “Very Low” on any other feature.

Which **two** of the presented DVD players would Shane prefer? _____ and _____

Very Low	Low	Medium	High	Very High
1	2 3	4	5	

Question 9:

		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	2	1	5	2	\$369
	B	1	5	4	2	\$369
	C	5	3	1	1	\$369
	D	5	4	5	4	\$369
	E	3	3	3	3	\$369

Tyrone wants a DVD player that either has a “Very High” rating for Programming Options, or one that scores at least “Medium” on every feature.

Which **three** of the presented DVD players would Tyrone prefer?

_____, _____, and _____

Question 10:

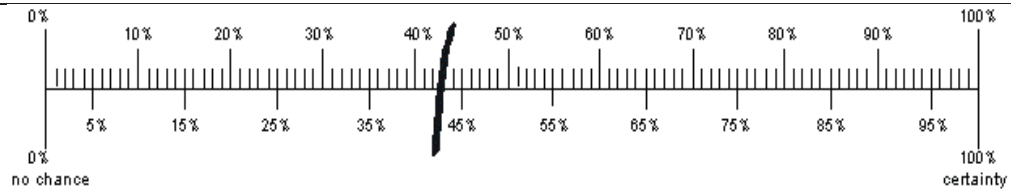
		Features				
		Picture Quality	Sound Quality	Program Options	Reliability of Brand	Price
DVD	A	2	1	5	4	\$369
	B	4	5	1	3	\$369
	C	1	3	5	5	\$369
	D	4	2	5	4	\$369
	E	5	5	1	3	\$369

Julie wants the best Reliability of Brand, but is willing to give up one point on Reliability of Brand for each increase of at least two points in the rating of Picture Quality. She isn’t concerned about the other features.

Which **three** of the presented DVD players would Julie prefer? _____, _____, and _____

Instructions:

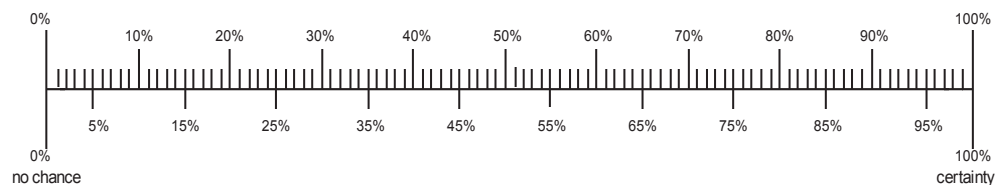
Each of these questions asks for your best guess at the chance that something will happen in the future. They use the “probability” scale that you see below. To answer each question, please put a mark on the scale at one specific tick mark, as follows:



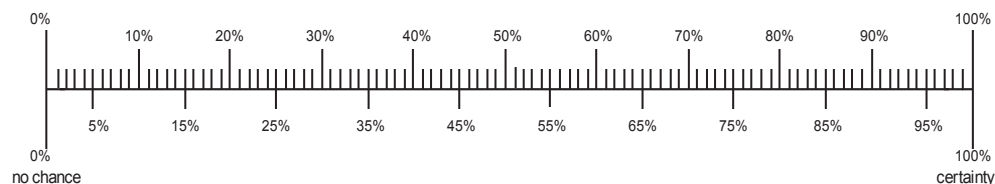
If you think that something has no chance of happening, mark it as having a 0% chance. If you think that something is certain to happen, mark it as having a 100% chance.

Just to make sure that you are comfortable with the scale, please answer the following practice questions.

What is the probability that you will eat pizza during the next year?



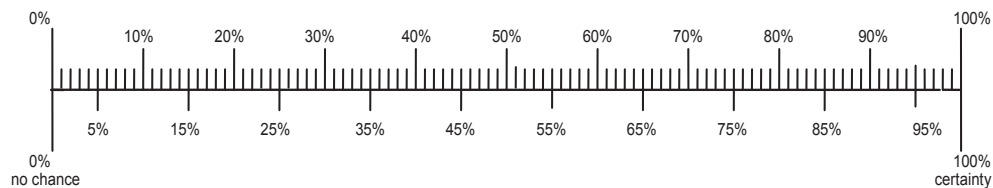
What is the probability that you will get the flu during the next year?



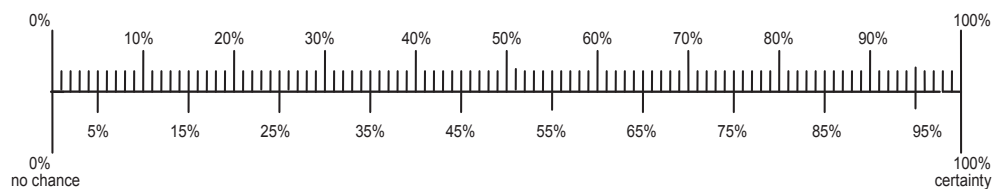
That is the end of the practice. If you have any questions, please ask them now.

A. The following questions ask about events that may happen some time during *the next year*.

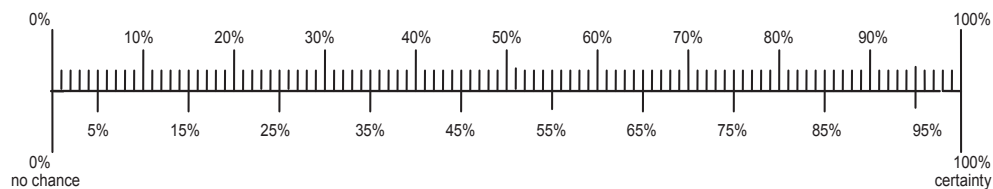
1. What is the probability that you will get into a car accident while driving during the next year?



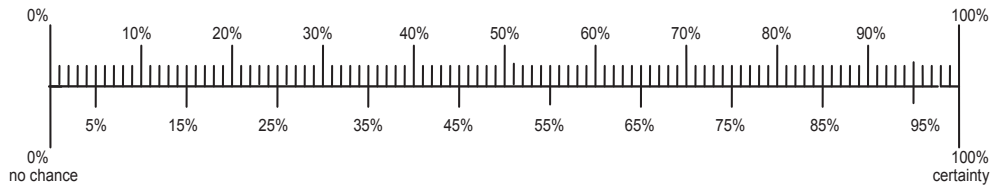
2. What is the probability that you will have a cavity filled during the next year?



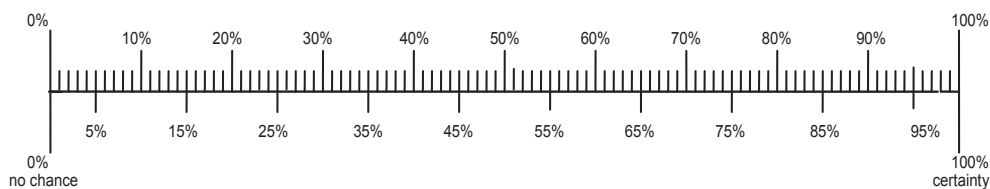
3. What is the probability that you will die (from any cause -- crime, illness, accident, and so on) during the next year?



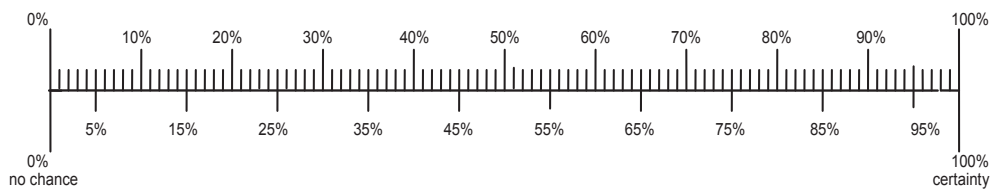
4. What is the probability that someone will steal something from you during the next year?



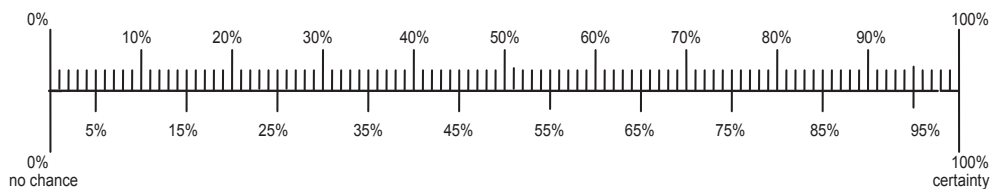
5. What is the probability that you will move your permanent address to another state some time during the next year?



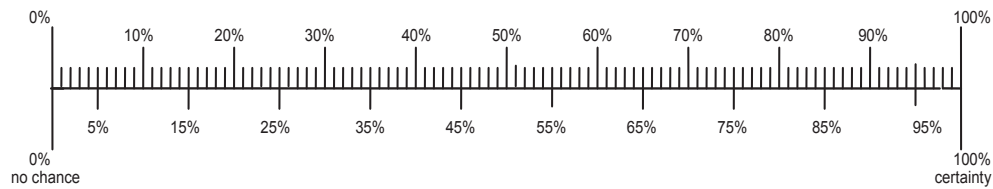
6. What is the probability that you will die in a terrorist attack during the next year?



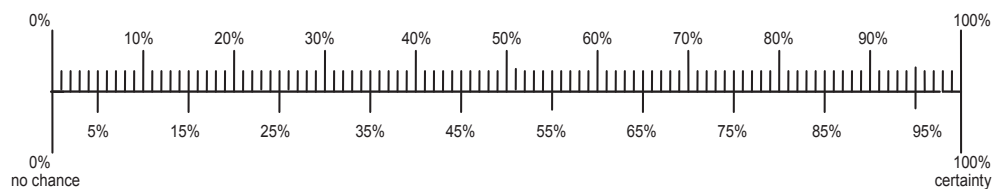
7. What is the probability that someone will break into your home and steal something from you during the next year?



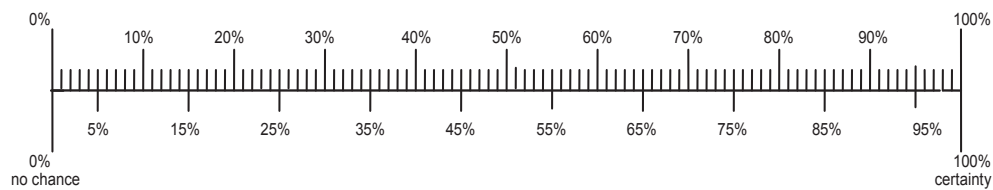
8. What is the probability that you will keep your permanent address in the same state during the next year?



9. What is the probability that you will visit a dentist, for any reason, during the next year?

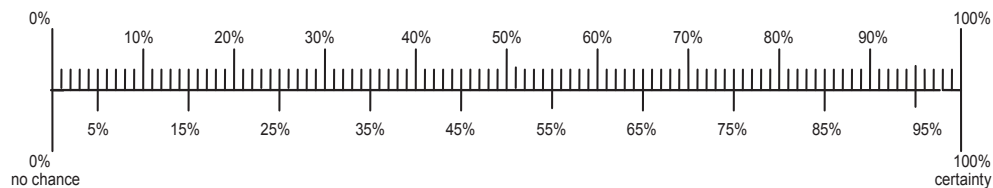


10. What is the probability that your driving will be accident-free during the next year?

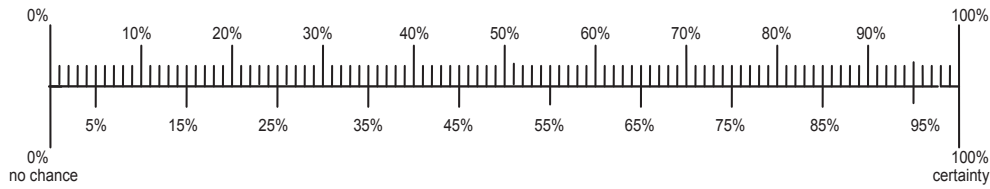


B. The following questions ask about events that may happen some time during *the next 5 years*.

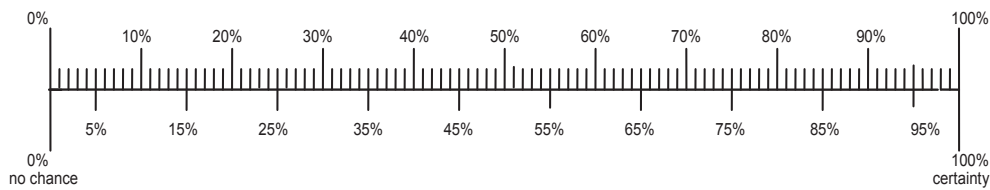
1. What is the probability that you will get into a car accident while driving during the next 5 years?



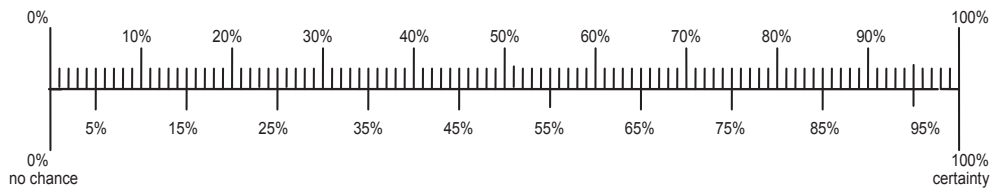
2. What is the probability that you will have a cavity filled during the next 5 years?



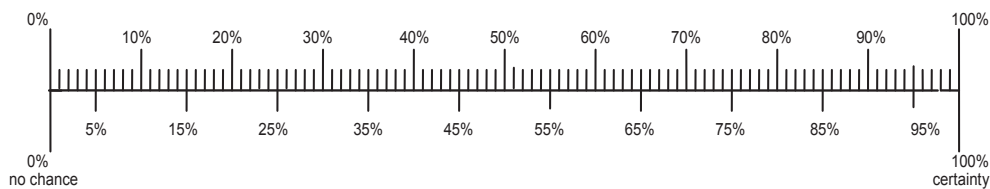
3. What is the probability that you will die (from any cause -- crime, illness, accident, and so on) during the next 5 years?



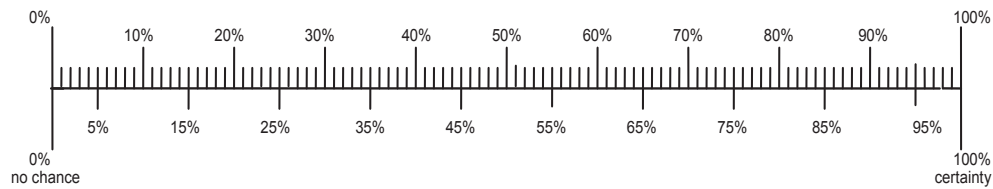
4. What is the probability that someone will steal something from you during the next 5 years?



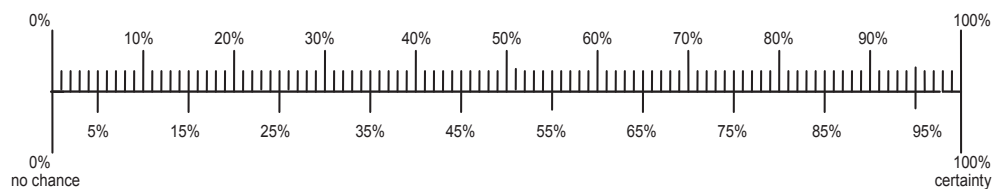
5. What is the probability that you will move your permanent address to another state some time during the next 5 years?



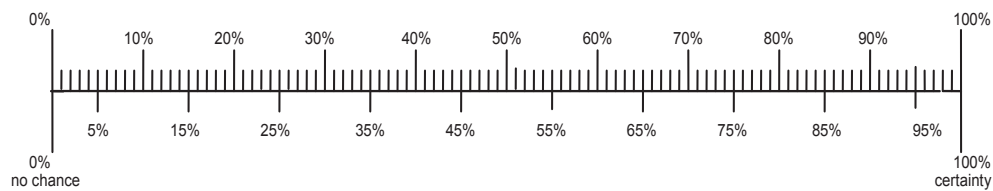
6. What is the probability that you will die in a terrorist attack during the next 5 years?



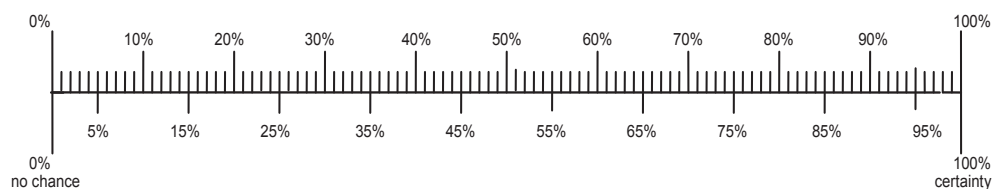
7. What is the probability that someone will break into your home and steal something from you during the next 5 years?



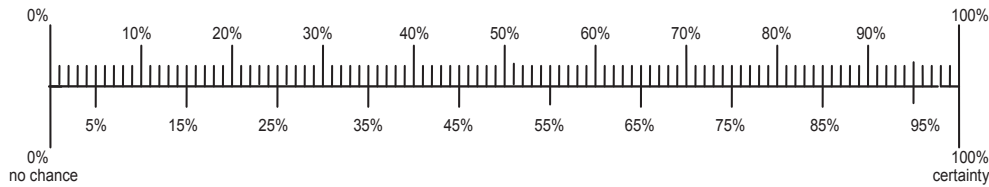
8. What is the probability that you will keep your permanent address in the same state during the next 5 years?



9. What is the probability that you will visit a dentist, for any reason, during the next 5 years?



10. What is the probability that your driving will be accident-free during the next 5 years?

**Instructions:**

In each of the following problems, choose between flipping a coin and a sure thing. Or, if they both seem the same to you, choose "Doesn't Matter."

1. Which do you like best, (1), (2), or (3)?

(1) (2)

Flip a Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$50** for sure

Doesn't Matter to Me

2. Which do you like best, (1), (2), or (3)?

(1) (2)

Flip a Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$60** for sure

Doesn't Matter to Me

3. Which do you like best, (1), (2), or (3)?

(1) (2)

Flip a Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$40** for sure

Doesn't Matter to Me

The next questions are about losses.

4. Which do you like best, (1), (2), or (3)?

(1) (2)

Flip a Coin

If Heads, lose **\$100**

If Tails, lose **\$0**

(3)

Sure Loss

Lose **\$50** for sure

Doesn't Matter to Me

5. Which do you like best, (1), (2), or (3)?

(1) (2)

Flip a Coin

If Heads, lose **\$100**

If Tails, lose **\$0**

(3)

Sure Loss

Lose **\$60** for sure

Doesn't Matter to Me

6. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip a Coin*If Heads, lose **\$100**If Tails, lose **\$0**

(3)

*Sure Loss*Lose **\$40** for sure*Doesn't Matter to Me*

In each of the next questions, choose between flipping two coins and flipping one coin.

7. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, win **\$100**Otherwise, win **\$0**

(3)

*Flip One Coin*If Heads, win **\$50**If Tails, win **\$0***Doesn't Matter to Me*

8. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, win **\$100**Otherwise, win **\$0**

(3)

*Flip One Coin*If Heads, win **\$60**If Tails, win **\$0***Doesn't Matter to Me*

9. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, win **\$100**Otherwise, win **\$0**

(3)

*Flip One Coin*If Heads, win **\$40**If Tails, win **\$0***Doesn't Matter to Me*

The next questions are about losses.

10. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, lose **\$100**Otherwise, lose **\$0**

(3)

*Flip One Coin*If Heads, lose **\$50**If Tails, lose **\$0***Doesn't Matter to Me*

11. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, lose **\$100**Otherwise, lose **\$0**

(3)

*Flip One Coin*If Heads, lose **\$60**If Tails, lose **\$0***Doesn't Matter to Me*

12. Which do you like best, (1), (2), or (3)?

(1) (2)

*Flip Two Coins*If Two Heads, lose **\$100**Otherwise, lose **\$0**

(3)

*Flip One Coin*If Heads, lose **\$40**If Tails, lose **\$0***Doesn't Matter to Me*

Questions 13 - 18 are just like 7 - 12 above, but one coin was already flipped. It came up heads, so you now have the following choices:

13. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$50** for sure

Doesn't Matter to Me

14. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$60** for sure

Doesn't Matter to Me

15. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, win **\$100**

If Tails, win **\$0**

(3)

Sure Win

Win **\$40** for sure

Doesn't Matter to Me

The next questions are about losses.

16. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, lose **\$100**

If Tails, lose **\$0**

(3)

Sure Loss

Lose **\$50** for sure

Doesn't Matter to Me

17. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, lose **\$100**

If Tails, lose **\$0**

(3)

Sure Loss

Lose **\$60** for sure

Doesn't Matter to Me

18. If you had already flipped once and it came up heads, which do you like best, (1), (2), or (3)?

(1) (2)

Flip Second Coin

If Heads, lose **\$100**

If Tails, lose **\$0**

(3)

Sure Loss

Lose **\$40** for sure

Doesn't Matter to Me

In each of the next questions, a coin will be flipped to see if you get a choice or not. Without knowing the result of the first flip, what would you choose in each of the following situations?

19. First Flip:

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, win **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, win **\$100**

Win **\$50** for sure

If Tails, win **\$0**

20. First Flip:

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, win **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, win **\$100**

Win **\$60** for sure

If Tails, win **\$0**

21. First Flip:

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, win **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, win **\$100**

Win **\$40** for sure

If Tails, win **\$0**

The next questions are about losses.

22. First Flip:

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, lose **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, lose **\$100**

Lose **\$50** for sure

If Tails, lose **\$0**

23. First Flip: _____

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, lose **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, lose **\$100**

Lose **\$60** for sure

If Tails, lose **\$0**

24. First Flip: _____

Flip a Coin

If Heads, get the **Choice** below

If Tails, **don't** get the Choice below, lose **\$0**

Choice: Before the first flip, which do you like best, (1), (2), or (3)?

(1) (2)

(3)

Flip a Coin

Sure Win

Doesn't Matter to Me

If Heads, lose **\$100**

Lose **\$40** for sure

If Tails, lose **\$0**

Instructions:

Each of the following problems presents a choice between two options. Each problem is presented with a scale ranging from 1 (representing one option) through 6 (representing the other option). For each item, please circle the number on the scale that best reflects your relative preference between the two options.

Problem 1

You are buying a gold ring on layaway for someone special. It costs \$200 and you have already paid \$100 on it, so you owe another \$100. One day, you see in the paper that a new jewelry store is selling the same ring for only \$90 as a special sale, and you can pay for it using layaway. The new store is across the street from the old one. If you decide to get the ring from the new store, you will not be able to get your money back from the old store, but you would save \$10 overall.

Would you be more likely to continue paying at the old store or buy from the new store?

1	2	3	4	5	6
Most likely to continue paying at the old store			Most likely to buy from the new store		

Problem 2

You enjoy playing tennis, but you really love bowling. You just became a member of a tennis club, and of a bowling club, both at the same time. The membership to your tennis club costs \$200 per year and the membership to your bowling club \$50 per year. During the first week of both memberships, you develop an elbow injury. It is painful to play either tennis or bowling. Your doctor tells you that the pain will continue for about a year.

Would you be more likely to play tennis or bowling in the next six months?

1	2	3	4	5	6
Most likely to play tennis			Most likely to play bowling		

Problem 3

You have been looking forward to this year's Halloween party. You have the right cape, the right wig, and the right hat. All week, you have been trying to perfect the outfit by cutting out a large number of tiny stars to glue to the cape and the hat, and you still need to glue them on. On the day of Halloween, you decide that the outfit looks better without all these stars you have worked so hard on.

Would you be more likely to wear the stars or go without?

1	2	3	4	5	6
Most likely to wear stars			Most likely to not wear stars		

Problem 4

After a large meal at a restaurant, you order a big dessert with chocolate and ice cream.

After a few bites you find you are full and you would rather not eat any more of it.

Would you be more likely to eat more or to stop eating it?

1	2	3	4	5	6
Most likely to eat more			Most likely to stop eating		

Problem 5

You are in a hotel room for one night and you have paid \$6.95 to watch a movie on pay TV. Then you discover that there is a movie you would much rather like to see on one of the free cable TV channels. You only have time to watch one of the two movies.

Would you be more likely to watch the movie on pay TV or on the free cable channel?

1	2	3	4	5	6
Most likely to watch pay TV			Most likely to watch free cable		

Problem 6

You have been asked to give a toast at your friend's wedding. You have worked for hours on this one story about you and your friend taking drivers' education, but you still have some work to do on it. Then you realize that you could finish writing the speech faster if you start over and tell the funnier story about the dance lessons you took together.

Would you be more likely to finish the toast about driving or rewrite it to be about dancing?

1	2	3	4	5	6
Most likely to write about driving				Most likely to write about dancing	

Problem 7

You decide to learn to play a musical instrument. After you buy an expensive cello, you find you are no longer interested. Your neighbor is moving and you are excited that she is leaving you her old guitar, for free. You'd like to learn how to play it.

Would you be more likely to practice the cello or the guitar?

1	2	3	4	5	6
Most likely to play cello				Most likely to play guitar	

Problem 8

You and your friend are at a movie theater together. Both you and your friend are getting bored with the storyline. You'd hate to waste the money spent on the ticket, but you both feel that you would have a better time at the coffee shop next door. You could sneak out without other people noticing.

Would you be more likely to stay or to leave?

1	2	3	4	5	6
Most likely to stay				Most likely to leave	

Problem 9

You and your friend have driven halfway to a resort. Both you and your friend feel sick.

You both feel that you both would have a much better weekend at home. Your friend says it is "too bad" you already drove halfway, because you both would much rather spend the time at home. You agree.

Would you be more likely to drive on or turn back?

1	2	3	4	5	6
Most likely to drive on				Most likely to turn back	

Problem 10

You are painting your bedroom with a sponge pattern in your favorite color. It takes a long time to do. After you finish two of the four walls, you realize you would have preferred the solid color instead of the sponge pattern. You have enough paint left over to redo the entire room in the solid color. It would take you the same amount of time as finishing the sponge pattern on the two walls you have left.

Would you be more likely to finish the sponge pattern or to redo the room in the solid color?

1	2	3	4	5	6
Most likely to finish sponge pattern				Most likely to redo with a solid color	

Instructions: Each of the following problems presents a choice between two options. Each problem is presented with a scale ranging from 1 (representing one option) through 6 (representing the other option). For each item, please circle the number on the scale that best reflects your relative preference between the two options.

Problem 1

Imagine a hospital is treating 32 injured soldiers, who are all expected to lose one leg.

There are two doctors that can help the soldiers, but only one can be hired:

If Doctor A is hired, 12 soldiers will lose one leg.

If Doctor B is hired, there is a 63% chance that nobody loses a leg and a 37% chance that all lose a leg.

Which doctor do you recommend?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 2

Imagine that the U.S. is preparing for the outbreak of an unusual disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 400 people will die.

If Program B is adopted, there is a 33% chance that nobody will die, and a 67% chance that 600 people will die.

Which program do you recommend to use?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 3

Imagine that your client has \$6,000 invested in the stock market. A downturn in the economy is occurring. You have two investment strategies that you can recommend under the existing circumstances to preserve your client's capital.

If strategy A is followed, \$4,000 of your client's investment will be lost.

If strategy B is followed, there is a 33% chance that the nothing will be lost, and a 67% chance that \$6,000 will be lost.

Which of these two strategies would you favor?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 4

Because of changes in tax laws, you may get back as much as \$1200 in income tax. Your accountant has been exploring alternative ways to take advantage of this situation. He has developed two plans:

If Plan A is adopted, you will lose \$800 of the possible \$1200.

If Plan B is adopted, you have a 33% chance of losing none of the money, and a 67% chance of losing all \$1200.

Which plan would you use?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 5

Imagine that recent evidence has shown that a pesticide is threatening the lives of 1,200 endangered animals. Two response options have been suggested:

If Option A is used, 600 animals will be lost for sure.

If Option B is used, there is a 75% chance that 400 animals will be lost, and a 25% chance that 1,200 animals will be lost.
Which option do you recommend to use?

1	2	3	4	5	6
Definitely would choose A				Definitely would choose B	

Problem 6

Imagine that your doctor tells you that you have a cancer that must be treated. Your choices are as follows:

Surgery: Of 100 people having surgery, 10 die because of the operation, and 66 die by the end of five years.

Radiation therapy: Of 100 people having radiation therapy, none die during the treatment, and 78 die by the end of five years.

Which treatment would you choose?

1	2	3	4	5	6
Definitely would choose surgery				Definitely would choose radiation	

Problem 7

Imagine that in one particular state it is projected that 1000 students will drop out of school during the next year. Two programs have been proposed to address this problem, but only one can be implemented. Based on other states' experiences with the programs, estimates of the outcomes that can be expected from each program can be made. Assume for purposes of this decision that these estimates of the outcomes are accurate and are as follows:

If Program A is adopted, 600 of the 1000 students will drop out of school.

If Program B is adopted, there is a 40% chance that none of the 1000 students will drop out of school and 60% chance that all 1000 students will drop out of school.

Which program would you favor for implementation?

1	2	3	4	5	6
---	---	---	---	---	---

Definitely would
choose A

Definitely would
choose B

Instructions: Each of the following problems ask you to rate your judgment of a product or a situation. Each problem is presented with a scale ranging from 1 (representing the worst rating) through 6 (representing the best rating). For each problem, please circle the number on the scale that best reflects your judgment.

Problem 1

As R&D manager, one of your project teams has come to you requesting an additional \$100,000 in funds for a project you instituted several months ago. The project is already behind schedule and over budget, but the team still believes it can be successfully completed. You currently have \$500,000 remaining in your budget unallocated, but which must carry you for the rest of the fiscal year. Lowering the balance by an additional \$100,000 might jeopardize flexibility to respond to other opportunities.

Evaluating the situation, you believe there is a fair chance the project will not succeed, in which case the additional funding would be lost; if successful, however, the money would be well spent. You also noticed that of the projects undertaken by this team, 20 of the last 50 have been unsuccessful.

What is the likelihood you would fund the request?

1	2	3	4	5	6
Very unlikely					Very likely

Problem 2

Imagine that a woman parked illegally. After talking to her, you believe that there is an 80% chance that she knew she parked illegally.

With this in mind, how much of a fine do you believe this woman deserves?

1	2	3	4	5	6
Minimum fine					Maximum fine

Problem 3

Considering the results of the survey, how would you rate the incidence of cheating at your university?

1	2	3	4	5	6
Very low					Very high

Imagine that a new technique has been developed to treat a particular kind of cancer.

A member of your immediate family is a patient at the local hospital with this kind of cancer. How likely are you to encourage him or her to undergo treatment using this technique?

1	2	3	4	5	6
Definitely no					Definitely yes

Imagine the following situation. You are entertaining a special friend by inviting them for dinner. You are making your favorite lasagna dish with ground beef. Your roommate goes to the grocery store and purchases a package of ground beef for you. The label says 20% fat ground beef.

1	2	3	4	5	6
Very low					Very high

Imagine that a type of condom has a 5% failure rate. That is, if you have sex with someone who has the AIDS virus, there is a 5% chance that this type of condom will fail to prevent you from being exposed to the AIDS virus.

1 2 3 4 5 6

Definitely no

Definitely yes

Suppose a student got 10% incorrect in the mid-term exam and 30% incorrect in the final-term exam, what would be your evaluations of this student's performance?

Instructions: The following problems ask *out of 100 people your age*, how many would say that it is sometimes OK to do different things. For each question, please circle a number between 0 (meaning *no one* thinks that it is sometimes OK) and 100 (meaning *everyone* thinks that it is sometimes OK).

6. Out of 100 people your age, how many would say it is sometimes OK ...

- ... to use your fists to resolve a conflict?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
7. Out of 100 people your age, how many would say it is sometimes OK ...
... to drink and drive?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
8. Out of 100 people your age, how many would say it is sometimes OK ...
... to yell and argue to solve a conflict?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
9. Out of 100 people your age, how many would say it is sometimes OK ...
... not to hold the door open for people?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
10. Out of 100 people your age, how many would say it is sometimes OK ...
... not to tell the police when you witness a crime?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
11. Out of 100 people your age, how many would say it is sometimes OK ...
... not to give directions to someone who is lost?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
12. Out of 100 people your age, how many would say it is sometimes OK ...
... not to be on time for appointments?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
13. Out of 100 people your age, how many would say it is sometimes OK ...
... not to return something you borrowed?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone
14. Out of 100 people your age, how many would say it is sometimes OK ...
... not to keep secrets that a friend told you?
- | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|-----|
| 0 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|----|----|----|----|----|----|----|----|----|-----|
- No one
Everyone

15. Out of 100 people your age, how many would say it is sometimes OK ...
... not to return phone calls right away?

0 10 20 30 40 50 60 70 80 90 100

No one

Everyone

16. Out of 100 people your age, how many would say it is sometimes OK ...
... not to spend time with friends in need?

0 10 20 30 40 50 60 70 80 90 100

No one

Everyone

b) Decision Outcome Inventory (DOI)

Instructions: The following questions ask whether different events have happened to you *in the last 10 years*. Please indicate “yes” or “no” for each.

In the last 10 years, have you ever...

☐ Yes ☐ No Rented a movie

☐ Yes ☐ No Returned a movie you rented without having watched it at
all

☐ Yes ☐ No Bought new clothes or shoes

☐ Yes ☐ No Bought new clothes or shoes you never wore

☐ Yes ☐ No Gone shopping for food or groceries

☐ Yes ☐ No Threw out food or groceries you had bought, because they
went bad

☐ Yes ☐ No Done your own laundry

☐ Yes ☐ No Ruined your clothes because you didn't follow the laundry
instructions on the label

☐ Yes ☐ No Been enrolled in any kind of school

☐ Yes ☐ No Been suspended from school for at least one day for any
reason

☐ Yes ☐ No Had any kind of job

☐ Yes ☐ No Quit a job after a week

☐ Yes ☐ No Had a driver's license

☐ Yes ☐ No Had your driver's license taken away from you by the
police

☐ Yes ☐ No Driven a car

☐ Yes ☐ No Been accused of causing a car accident while driving

☐ Yes ☐ No Gotten more than 5 parking tickets

☐ Yes ☐ No Gotten more than 5 speeding tickets

☐ Yes ☐ No Gotten lost or gone the wrong way for more than 10
minutes while driving

☐ Yes ☐ No Locked your keys in the car

☐ Yes ☐ No Bought any kind of car

☐ Yes ☐ No Had to spend at least \$500 to fix a car you had owned for
less than half a year

☐ Yes ☐ No Taken a trip by airplane

☐ Yes ☐ No Missed a flight

☐ Yes ☐ No Taken the train or the bus

☐ Yes ☐ No Taken the wrong train or bus

☐ Yes ☐ No Had any form of ID (driver's license, passport, birth certificate)

☐ Yes ☐ No Had your ID replaced because you lost it

☐ Yes ☐ No Lived in a rented apartment or other rental property

☐ Yes ☐ No Been kicked out of an apartment or rental property before the lease ran out

☐ Yes ☐ No Carried a key to your home

☐ Yes ☐ No Had the key to your home replaced because you lost it

☐ Yes ☐ No Locked yourself out of your home

☐ Yes ☐ No Been responsible for electricity, cable, gas or water payments

☐ Yes ☐ No Had your electricity, cable, gas or water shut off because you didn't pay on time

☐ Yes ☐ No Been responsible for a mortgage or loan

☐ Yes ☐ No Foreclosed a mortgage or loan

☐ Yes ☐ No Been responsible for rent or mortgage payments

☐ Yes ☐ No Paid a rent or mortgage payment at least 2 weeks too late

In the last 10 years, have you ever...

☐ Yes ☐ No Used checks

☐ Yes ☐ No Had a check bounce

☐ Yes ☐ No Had a credit card

☐ Yes ☐ No Had more than \$5000 in credit card debt

☐ Yes ☐ No Invested in the stock market

☐ Yes ☐ No Lost more than \$1000 on a stock-market investment

☐ Yes ☐ No Been to a bar, restaurant, or hotel

☐ Yes ☐ No Been kicked out of a bar, restaurant, or hotel by someone
who works there

☐ Yes ☐ No Loaned more than \$50 to someone

☐ Yes ☐ No Loaned more than \$50 to someone and never got it back

☐ Yes ☐ No Had a romantic relationship that lasted for at least 1 year

☐ Yes ☐ No Cheated on your romantic partner of 1 year by having
sex with someone else

☐ Yes ☐ No Been married

☐ Yes ☐ No Been divorced

☐ Yes ☐ No Had sex

☐ Yes ☐ No Been diagnosed with an STD

☐ Yes ☐ No Had an unplanned pregnancy (or got someone pregnant,
unplanned)

☐ Yes ☐ No Had sex with a condom

☐ Yes ☐ No Had a condom break, tear, or slip off

☐ Yes ☐ No Had an alcoholic drink

☐ Yes ☐ No Consumed so much alcohol you vomited

☐ Yes ☐ No Received a DUI for drunk driving

☐ Yes ☐ No Been out in the sun

☐ Yes ☐ No Got blisters from sun burn

☐ Yes ☐ No Been in a jail cell overnight for any reason

☐ Yes ☐ No Been in a public fight or screaming argument

☐ Yes ☐ No Declared bankruptcy

☐ Yes ☐ No Forgotten a birthday of someone close to you and did not realize until the next day or later.

☐ Yes ☐ No Been diagnosed with type 2 diabetes

☐ Yes ☐ No Broke a bone because you fell, slipped, or misstepped

c) Ecological risky decisions

i) *Financial decisions*

Debt When Loan Enters Repayment	Standard (10 years)		Graduated (10 years)			Extended-Fixed (25 years)		Extended-Graduated (25 years)		
	Payment	Total Paid	Minimum Payment	Maximum Payment	Total Paid	Payment	Total Paid	Minimum Payment	Maximum Payment	Total Paid
\$10,000	\$115	\$13,810	\$66	\$109	\$14,860	-	-	-	-	-
\$20,000	\$230	\$27,619	\$133	\$398	\$29,720	-	-	-	-	-
\$30,000	\$345	\$41,429	\$199	\$598	\$44,580	-	-	-	-	-
\$40,000	\$460	\$55,239	\$266	\$797	\$59,439	\$278	\$83,289	\$227	\$397	\$90,207
\$50,000	\$575	\$69,048	\$332	\$996	\$74,300	\$347	\$104,111	\$283	\$496	\$112,762
\$60,000	\$690	\$82,858	\$398	\$1,195	\$89,160	\$416	\$124,933	\$340	\$595	\$135,314
\$70,000	\$806	\$96,667	\$465	\$1,393	\$104,020	\$486	\$145,755	\$397	\$694	\$157,865
\$80,000	\$921	\$110,477	\$531	\$1,593	\$118,880	\$555	\$166,577	\$453	\$793	\$180,427
\$90,000	\$1,036	\$124,287	\$597	\$1,791	\$133,740	\$625	\$187,399	\$510	\$892	\$202,980
\$100,000	\$1,151	\$138,096	\$664	\$1,991	\$148,600	\$694	\$208,222	\$567	\$992	\$225,531

Imagine that you take out a \$50,000 federal student loan to help pay for college. You are offered four possible repayment plans. The table below provides examples of the monthly repayments for each plan. Note: For the Graduated (10 years) plan, you would start by paying the minimum amount; the payment amount then increases every two years up to the maximum amount.

Look at the table carefully and answer the following questions

1. What is the total amount of interest payable on the Extended-Fixed (25 years) plan?
(**\$54,111**)
2. Which option has the minimum interest payment (normatively superior)? (**Standard 10 years**)
3. Assume someone has borrowed 50,000 for his studies, and he hopes to get a good job after his graduation in 5 years (when he will be able to pay more toward his debt). In this case which option he should choose?
 - a) Standard
 - b) Graduated**
 - c) Extended fixed
 - d) Extended graduated
4. What is the total interest paid in percentage if you have borrowed \$50,000 and returned \$69,048?
(**38.1%**)

ii) Medical decisions

- 1) Cervical cancer is very rare. 4 out of 100,000 women are affected by this cancer. The human papillomavirus (HPV) vaccine is federally approved and is being promoted as a method that helps reduce cervical cancer. Research studies suggest that the vaccine is 90 percent effective in preventing transmission of certain virus types. This conclusion is based on the results from a large international medical trial of 18,525 women aged 15-25, sponsored by the drug's manufacturer. 23 cases of the HPV virus were detected in the medical trial. Two of these cases were among the 9,258 women receiving the HPV vaccine, and 21 were among the 9,267 controls, who received a hepatitis A vaccine. The mean follow-up time was 14.8 months.

<http://www.cancer.gov/clinicaltrials/results/summary/2007/hpv-vaccine0707>

- i. To what extent is the conclusion that “the vaccine was 90% effective” correct? Please provide an answer in using the following 7 point scale

- a. Completely confident that conclusion is right 1 2 3 4 5 6 7 Completely confident that conclusion is wrong
- ii. What is the relative effectiveness of the vaccine? (express your answer in percentage) about 90%
- iii. What is the absolute effectiveness of the vaccine? (express your answer in percentage) Less than 1% (.2%)
- iv. What are the chances that a woman gets cervical cancer after getting vaccinated? (express your answer in percentage) very low, Less than 1%
- v. What did the women receive in control group? (Hepatitis A Vaccine)

d) Paradigmatic risky decisions and Intertemporal choice

Items taken from Fredrick (2005); items that significantly differentiated between low and high ability groups

1) Intertemporal Tasks: (Fredrick, 2005; Read, 2001)

- i) \$3400 this month or \$3800 next month
- ii) \$100 now or \$140 next year
- iii) \$100 now or \$1100 in 10 years
- iv) \$9 now or \$100 in 10 years
- v) \$40 immediately or \$1000 in 10 years
- vi) \$100 now or \$20 every year for 7 years
- vii) \$400 now or \$100 every year for 10 years
- viii) \$500 in Eight months or \$1060 in sixteen months
- ix) \$500 now or \$2400 in 2 years
- x) \$1000 in sixteen month or \$2400 in two years
- xi) Smallest amount in 4 days preferred to \$170 in 2 months \$-----?

2) Risk preferences (lotteries) (Cokely et al., 2009; Fredrick, 2005; Kahneman & Tversky, 1979; Pachur & Galesic, 2012; Peters & Levin, 2008)

- i) Lose \$50 or 50% chance to lose \$400 (50 vs 200 lose)
- ii) Lose \$120 or 5% chance to lose \$1600 (120 vs 80 lose)
- iii) Lose \$200 or 1% chance to lose \$3000 (200 vs 30 lose)
- iv) Lose \$275 or 20% chance to lose \$900 (275 vs 180 lose)
- v) Lose \$400 or 70% chance to lose \$480 (400 vs 336 lose)
- vi) Gain \$50 or 50% chance to win \$400 (50 vs 200 gain)
- vii) Gain \$120 or 5% chance to win \$1600 (120 vs 80 gain)
- viii) Gain \$200 or 1% chance to win \$3000 (200 vs 30 gain)
- ix) Gain \$275 or 20% chance to win \$900 (275 vs 180 gain)
- x) Gain \$400 or 70% chance to win \$480 (400 vs 336 gain)
- xi) \$100 for sure or a 75% chance of \$200
- xii) Lose \$100 for sure or 60% chance to lose \$250
- xiii) \$500 for sure or a 15% chance of \$1,000,000
- xiv) Lose \$100 for sure or a 3% chance to lose \$7000

- xv) 25% chance to win \$6,000 or 25% chance to win \$4000 and 25% chance to win \$2000
- xvi) 33% chance to win \$2500 and 67% chance of winning nothing
Or
34% chance of winning \$2400 and 66% chance of winning nothing
- xvii) 15% chance to lose \$20 and 85% to lose nothing
Or
10% chance to lose \$25 and a 90% chance of losing nothing
- xviii) Choose Medicine A: with probability of 15% the medication leads to fever as a side effect, and 85% no side effects occur.
Or
Choose Medicine B: with probability of 10% the medication leads to insomnia as a side effect with the probability of no side effects occur.
- xix) 3,600 animals that are endangered by the fire. Two programs have been proposed to protect the animals. Choose one program

Program A: 2,040 animals will perish.
Or
Program B: 1/3 probability that no animals will perish, and a 2/3 probability that 3,600 animals will perish.
- xx) A severe drought is foreseen to hit the South of Spain this summer. The drought will cause the destruction of 24,000 acres of crops. Two programs of water supply have been proposed: choose one program

Program A: 5,600 acres of crops will be saved.

OR

Program B: 1/3 probability that 24,000 acres of crops will be saved, and a 2/3 probability that none will be saved.

3) Ratio Bias/ Denominator neglect (Alonso & Fernandez-Berrocal, 2003; Denes-Raj & Epstein, 1994; Okan et al., 2012; Yamagishi, 1997)

- i) (From Okan et al., 2012)
With the new drug BENOFRENO, the risk of death from a heart attack reduced for people with high cholesterol. A study with 900 with high cholesterol showed that 80 of the 800 people who have not taken the drug

deceased after a heart attack, compared with 16 of the 100 people who have taken the drug.

1. How beneficial was the Benofreno?

Not beneficial 1 2 3 4 5 6 7 very beneficial

2. How confident are you about your decision?

Not sure 1 2 3 4 5 6 7 very sure

- ii) (From Denes-Raj & Epstein, 1994. Though they actually manipulated this) Suppose you have a chance to win \$5 by drawing a Red ball from either of the two bags. Bag A contains 1 red ball out of 9, and bag B contains 10 red balls out of 100. Indicate the bag from which you wish to draw a ball?

Bag A

Bag B

- iii) (Taken from Alonso & Fernandez-Berrocal, 2003)

Imagine that you have finished your studies and you need to find a job. You are looking through the newspaper and you read an advert from a company that is looking for people like you. This company offers two types of job positions: Type P and Type Q. Both are of the same category and you like them equally. Therefore, you quickly go to the company to present your application to work in either of them. Once there, they tell you that you cannot request both at the same time, you have to opt for one of them: P or Q.

For the Type P job, 2 people are needed and only 10 candidates are admitted (one of them would be you). For the Type Q job, 10 people are needed and only 100 candidates are admitted (one of them would be you).

What job type would you choose? :

___ Type P; ___ Type Q; ___ No preference

What job type do you believe most people would choose?

___ Type P; ___ Type Q; ___ No preference

What job type do you believe a completely logical person would choose?

___ Type P; ___ Type Q; ___ No preference

- iV) (Taken from Yamagishi, 1997)

Cancer causes deaths; below are two situations that presents cancer risk statistics, please rate how risky they appear to you on a scale provided

a) Cancer kills 1286 people out of 10,000

No risk at all 1 2 3 4 5 6 7

Maximum possible risk

b) Cancer kills 24.14 people out of 100

No risk at all 1 2 3 4 5 6 7

Maximum possible risk

e) Cognitive Reflection Test (Fredrick, 2005)

- 1) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
- 2) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?
- 3) A bat and ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?

f) Berlin Numeracy Test (Cokely et al., 2012)

- 1) Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in a choir 100 are men. Out of the 500 inhabitants that are not in a choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.

(25%)

- 2) Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?

(30)

- 3) Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On

average, out of these 70 throws how many times would the die show the number

6? **(20)**

4) In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?

(50)

g) Schwartz Three Items Numeracy Test (Schwartz, 1997)

- 1) Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips?
- 2) In the Big Bucks Lottery, the chance of winning a \$10 prize is 1%. What is your best guess about how many people would win a \$10 prize if 1,000 people each buy a single ticket to Big Bucks?
- 3) In ACME Publishing Sweepstakes, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME Publishing Sweepstakes win a car?

h) Lipkus 8 items numeracy test (Lipkus, 2001)

- 4) Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up even (2, 4, or 6)?
- 5) Which of the following numbers represents the biggest risk of getting a disease? 1 in 100, 1 in 1000, 1 in 10
- 6) Which of the following represents the biggest risk of getting a disease? 1%, 10%, 5%

- 7) If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?
- 8) If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 1000?
- 9) If the chance of getting a disease is 20 out of 100, this would be the same as having a ____% chance of getting the disease.
- 10) If Person A's risk of getting a disease is 1% in ten years, and Person B's risk is double that of A's, what is B's risk?
- 11) If Person A's chance of getting a disease is 1 in 100 in ten years, and Person B's risk is double that of A, what is B's risk?
- 12) The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?

i) Abbreviated Numeracy Scale: A Rasch Analysis Approach (Weller et al., 2012)

Include total 8 items (2 CRT, 3 Schwartz and 2 Lipkus (8 & 9) Items) and one following item.

- 8) Suppose you have a close friend who has a lump in her breast and must have a mammography . . . The table below summarizes all of this information. Imagine that your friends tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

j) Numeracy Understanding in Medicine Instrument (NUMi; Schapira et al., 2012)

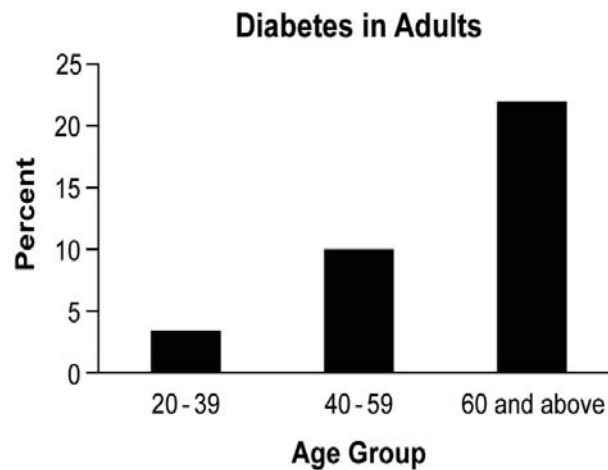
1. James has diabetes. His goal is to have his blood sugar between 80 and 150 in the morning. Which of the following blood sugar readings is within his goal?
 - a. 55
 - b. 140
 - c. 165
 - d. 180
2. Nathan has a pain rating of 5 on a pain scale of 1 (no pain) to 10 (worst possible pain). One day later Nathan still has pain but not as much. Now, what pain rating might Nathan give?
 - a. 3
 - b. 5
 - c. 7
 - d. 9
3. Natasha started taking a new medicine that may cause the side effects listed below. Which side effect is Natasha least likely to have?
 - a. Dizziness 1 in 5 people
 - b. Nausea 1 in 10 people
 - c. Stomach pain 1 in 100 people
 - d. Allergic reaction 1 in 200 people
4. Frank has a test done to look for blockages in the arteries of his heart. The doctor said that the greater the percent (%) blockage in the artery, the greater the risk of a heart attack. Which percent (%) blockage is most likely to cause a heart attack?
 - a. 33%
 - b. 50%
 - c. 75%
 - d. 98%

5. The doctor told Maria not to take more than 3 grams (g) of Tylenol a day. Each Tylenol pill is 500 milligrams (mg). What is the greatest number of pills that Maria can take in one day?
 - a. 3 pills
 - b. 6 pills
 - c. 8 pills
 - d. 12 pills
6. A medical study will randomly assign people so they are equally likely to get medicine A or medicine B. If there are 300 people in the study, about how many are expected to get medicine A?
 - a. 100 people
 - b. 150 people
 - c. 200 people
 - d. 250 people
7. Older age and smoking both increase the risk of a heart attack over time. David is now 50 years old and smokes. His risk of a heart attack in the next 10 years is 10%. If he continues to smoke which of the following could be his risk of a heart attack over the next 20 years?
 - a. 5%
 - b. 10%
 - c. 30%
 - d. 100%
8. James starts a new blood pressure medicine. The chance of a serious side effect is 0.5%. If 1000 people take this medicine, about how many would be expected to have a serious side effect?
 - a. 1 person
 - b. 5 people
 - c. 50 people
 - d. 500 people
9. The PSA (Prostate Specific Antigen) is a blood test that can be used to screen for prostate cancer. However, 30% of men who have an abnormal test result will turn out not to have cancer. John has an abnormal test result. What is the chance that John has prostate cancer?
 - a. 0%
 - b. 30%
 - c. 70%
 - d. 100%

10. Rebecca is treated for stage 2 breast cancer. The chance that the cancer will come back is 10% over 10 years. If Rebecca takes a new medicine, this chance will decrease by 30%. If 100 women like Rebecca take this medicine, how many are now expected to have breast cancer come back within 10 years?
- 3 out of 100 women
 - 7 out of 100 women
 - 10 out of 100 women
 - 30 out of 100 women
11. A study found that chemotherapy decreased the risk of dying from colon cancer by about 30%. The study was 95% sure that the actual benefit was between 10% and 50%. Which of the following is not in the expected range of benefit?
- 11% decrease in risk
 - 30% decrease in risk
 - 45% decrease in risk
 - 95% decrease in risk
12. A study in arthritis patients found that Medicine A decreased arthritis pain 10% more often than Medicine B. The difference was not statistically significant. Which of the following best describes these results?
- Medicine A and Medicine B work equally well
 - Medicine A is proven to be better than Medicine B
 - Medicine B is proven to be better than Medicine A
13. A study found that a new diabetes medicine controlled blood sugar 8% more often than the old medicine. This difference was statistically significant ($p < 0.05$). The probability that this finding is due to chance alone is less than:
- 1 in 5
 - 1 in 10
 - 1 in 15
 - 1 in 20
14. In general, the results of randomized controlled trial will be more reliable if a larger number of people are in the study.
- True
 - False
15. A study was done of health habits. A group of people took a survey every few years for 20 years. The study found that people who exercised 3 times a week or more lived an average of 2 years longer than those who did not. What does this study show?
- Exercise was the cause of living a longer life
 - There is a relationship between exercise and living a longer life

16. According to the graph below, what percent (%) of adults in the 40-59 year old age group have diabetes?

- a. 5%
- b. 10%
- c. 15%
- d. 20%



17. John has a fever. The doctor tells him to come to the hospital if his temperature is above 102.5 F. Otherwise, John should take Tylenol and rest. John's temperature is shown in the picture below. What should John do?

- a. Take Tylenol and rest
- b. Go to the hospital

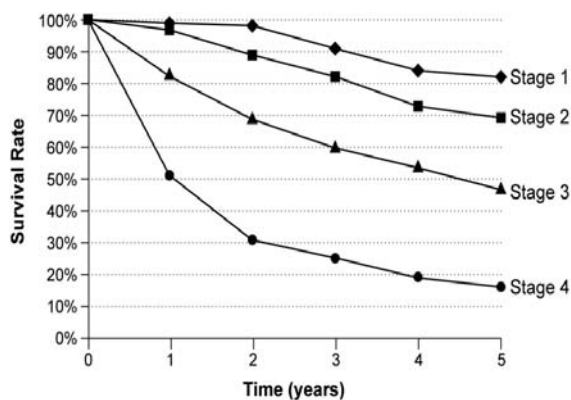


18. Mary has 2 cups of food whose nutrition label is below. How many calories are in the 2 cups of food?

- a. 140 calories
- b. 280 calories
- c. 560 calories
- d. 680 calories

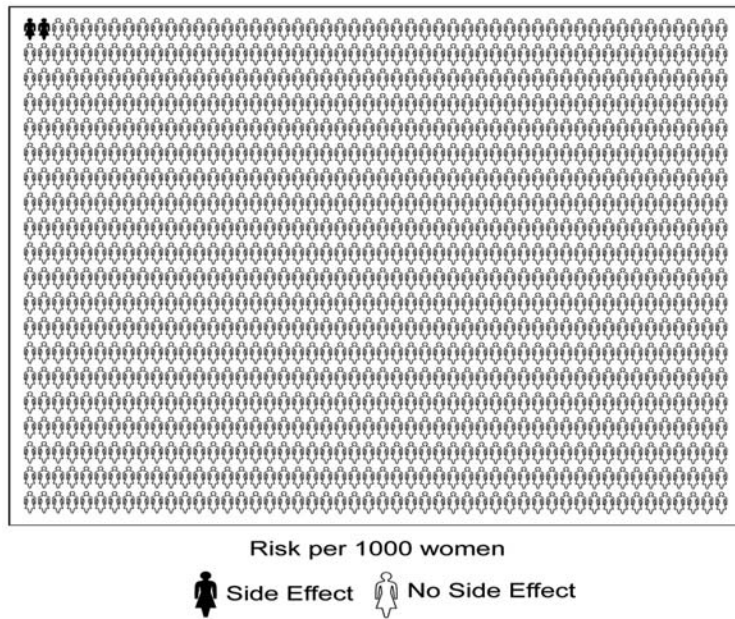
Nutrition Facts	
Serving Size 1 cup (228g)	
Servings per Container 2	
Amount Per Serving	
Calories 280	Calories from Fat 120
% Daily Value*	
Total Fat 13g	20%
Saturated Fat 5g	25%
Trans Fat 2g	
Cholesterol 2mg	10%
Sodium 660 mg	28%
Total Carbohydrate 31g	10%
Dietary Fiber 3g	
Sugars 5g	
Protein 5g	
Vitamin A 4%	Vitamin C 2%
Calcium 15%	Iron 4%
*Percent Daily Values are based on a 2,000-calorie diet. Your Daily values may be higher or lower depending on your calorie needs	

19. Andrea has stage 2 breast cancer. According to the graph below, what is Andrea's chance of surviving 3 years after her diagnosis?
- 56%
 - 82%
 - 92%
 - 100%



20. Carol is taking a new medicine. The risk of a side effect is very small. According to the picture below, what is her risk of having a side effect?
- 0.0002
 - 0.002

- c. 0.02
d. 0.20



k) Mathematics Anxiety Rating Scale Revised (MARS_R) (Brewer, 2011)

Please review the following items and state how they apply to you. The scale is 1 through 5, 1 means no anxiety and 5 means high anxiety.

- 1) Looking through the pages in a math book

No anxiety 1 2 3 4 5 High

Anxiety

- 2) Logging into a math class

No anxiety 1 2 3 4 5 High

Anxiety

- 3) Having to use the tables in the back of a math book

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 4) Down loading a math textbook

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 5) Thinking about tomorrow's upcoming math test

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 6) Watching a video where the teacher works an algebraic equation on the board

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 7) Taking a quiz in a math course

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 8) Logging in to an online tutoring program to work on homework

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 9) Reading and interpreting graphs and charts

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 10) Getting ready to study for a math test

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

- 11) Receiving a homework assignment of many difficult problems due the next class meeting

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

12) Listening to a recorded lecture for a math class

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

13) Waiting to get a math test returned in which you expected to do well

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

14) Working on an abstract mathematical problem

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

15) Completing a graded assignment in math class

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

16) Taking a final examination in math class

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

17) Listening to another student explain a math formula

No anxiety	1	2	3	4	5	High
------------	---	---	---	---	---	------

Anxiety

I) Ten-Item Personality Inventory-(TIPI) (Gosling, Rentfrow, and Swann, 2003)

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree

with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

Disagree Strongly	1
Disagree moderately	2
Disagree a little	3
Neither agree nor disagree	4
Agree a little	5
Agree moderately	6
Agree strongly	7

I see myself as:

1. _____ Extraverted, enthusiastic.
2. _____ Critical, quarrelsome.
3. _____ Dependable, self-disciplined.
4. _____ Anxious easily upset.
5. _____ Open to new experiences, complex.

6. _____ Reserved, quiet.
7. _____ Sympathetic, warm.
8. _____ Disorganized, careless.
9. _____ Calm, emotionally stable.
10. _____ Conventional, uncreative

**m) Raven's Advance Progressive Matrices__ a 12 items test of Fluid Intelligence
(Bors and Stokes, 1998)**

Please follow the link below for an example.

<http://unipark.de/uc/Cokely/9911/>

Appendix G

Table G-1. *Descriptive statistics for all variables (Predictors, criterion and control)*
Study 2

	N	Min.	Max	Mean	SD	Skew	SE
BNT1	124	0	4	1.90	1.1	.34	.217
BNT_S	124	1.00	7.00	4.50	1.48	-.29	.217
BNT_C	124	2.00	19.00	9.78	4.61	.28	.217
BNT- Engineering	124	.00	10.00	4.80	2.48	.24	.217
BNT-Stats	124	.00	10.00	4.97	2.59	.06	.217
Schwartz	124	0	3	2.60	.68	-1.6	.217
Lipkus	124	5	10	9.15	1.09	-1.5	.217
Weller	124	2.00	8.00	5.51	1.38	-.64	.217
NUMI	124	11.00	20.00	17.57	2.01	-1.06	.217
CRT	124	0	3	1.43	1.11	-.01	.217
Operations	124	.00	5.00	2.27	1.28	.30	.217

Probability	124	.00	5.00	2.70	1.72	.03	.217
Geometry	124	.00	5.00	2.40	1.36	.32	.217
Algebra	124	.00	5.00	2.40	1.51	.11	.217
Financial Literacy	124	0	12	6.78	2.70	-.34	.217
Math Anxiety	123	20.00	80.00	44.52	12.18	.24	.218
Memory	124	.00	6.00	3.44	1.52	-.28	.217
Ravens	124	.00	12.00	7.55	2.66	-.66	.217
Framing Risk	120	1.57	5.00	3.83	.63	-.87	.221
Framing Attribute	120	3.14	4.86	4.26	.37	-.19	.221
Framing Total	120	2.71	4.86	4.04	.38	-.59	.221
Social Norms	119	-.18	.80	.47	.19	-.77	.222
Overconfidenc e	121	1.00	5.00	2.69	1.43	.45	.220
Sunk Cost	120	2.30	5.80	4.06	.63	.23	.221

Applying Decision Rules	124	.00	100.00	75.16	20.31	-1.7	.217
Path Independence	119	.25	1.00	.68	.17	-.04	.222
Risk Perception	120	35.00	90.00	63.20	11.31	-.04	.221
DOI	118	.20	.80	.57	.09	-.65	.223
Ratio Bias	124	.00	3.00	1.96	.76	-.50	.217
Lotteries	124	5	14	9.36	1.82	.06	.217
Financial Ecological	124	0	4	1.23	.97	.54	.217
Medical Ecological	124	0	5	1.43	1.05	.30	.217
Ecological Total	124	.00	4.00	1.32	.86	.39	.217
ADMC	118	.42	.85	.67	.07	-.56	.223
Standard Risky	124	.34	.90	.63	.11	-.19	.217
Risk Literacy	124	.46	2.55	1.22	.43	.42	.217

SDMS_T	118	.98	3.23	1.90	.47	.35	.223
--------	-----	-----	------	------	-----	-----	------

Appendix H

Theoretical Modeling (Hierarchical regression models)

Hierarchical regression analysis following theoretical models, for each
Decision and Ability Task (values showing adjusted R²s)

Decision Task	Model	β	R	AR²	R²	F
					Change	Change
Framing	Model 1		0.24	0.05	0.05	7.0**
	Geometry	0.24**				
	Model 2		0.25	0.04	0.004	0.49
	Geometry	0.20*				
	Algebra	0.07				
	Model 3		0.25	0.04	0.00	0.006
	Geometry	0.20*				
	Algebra	0.07				
	Operations	0.01				
	Model 4		0.27	0.04	0.01	1.8
	Geometry	0.17				
	Algebra	0.02				
	Operations	0.02				
	Probability	0.14				
Social Norms	Model 1		0.30	0.08	0.08	11.3**
	Probability	0.30**				
	Model 2		0.31	0.08	0.01	0.70
	Probability	0.26**				
	Operations	0.08				
	Model 3		0.32	0.08	0.01	1.1
	Probability	0.23*				
	Operations	0.05				
	Geometry	0.11				
	Model 4		0.32	0.07	0.00	0.26
	Probability	0.21`				
	Operations	0.03				
	Geometry	0.09				
	Algebra	0.06				
Overconfidence	Model 1		0.25	0.06	0.06	8.2**
	Geometry	- 0.25**				

	Model 2		0.29	0.07	0.02	2.4
	Geometry	-0.18				
	Algebra	-0.16				
	Model 3		0.29	0.06	0.00	0.001
	Geometry	-0.18				
	Algebra	-0.16				
	Operations	0.00				
	Model 4		0.29	0.05	0.00	0.22
	Geometry	-0.17				
	Algebra	-0.14				
	Operations	0.01				
	Probability	-0.05				
Sunk Cost	Model 1		0.26	0.06	0.06	8.4**
	Probability					
		0.26**				
	Model 2		0.26	0.05	0.00	0.32
	Probability	0.23*				
	Operations	0.06				
	Model 3		0.26	0.05	0.00	0.1
	Probability	0.24*				
	Operations	0.07				
	Geometry	-0.03				
	Model 4		0.29	0.05	0.00	1.5
	Probability	0.20`				
	Operations	0.02				
	Geometry	-0.06				
	Algebra	0.15				
Applying Decision Rules	Model 1		0.28	0.07	0.07	10.6**
	Operations					
		0.28**				
	Model 2		0.29	0.07	0.01	0.8
	Operations	0.24*				
	Probability	0.09				
	Model 3		0.30	0.07	0.01	0.9
	Operations					
		0.27**				

	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.31	0.06	0.00	0.2
	Operations	0.26				
	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
Path Independence	Model 1		0.05	-0.01	0.002	0.27
	Probability	0.05				
	Model 2		0.11	-0.00	0.01	1.2
	Probability	-0.01				
	Operations	0.11				
	Model 3		0.14	-0.01	0.01	0.9
	Probability	-0.04				
	Operations	0.09				
	Geometry	0.10				
	Model 4		0.14	-0.01	0.00	0.04
	Probability	-0.04				
	Operations	0.07				
	Geometry	0.10				
	Algebra	0.03				
Risk Perception	Model 1		0.23	0.05	0.05	6.9
	Probability	0.05				
	Model 2		0.24	0.04	0.00	0.14
	Probability	-0.01				
	Operations	0.11				
	Model 3		0.24	0.03	0.00	0.13
	Probability	-0.04				
	Operations	0.09				
	Geometry	0.10				
	Model 4		0.24	0.03	0.00	0.19
	Probability	-0.04				
	Operations	0.07				
	Geometry	0.10				
	Algebra	0.03				
Ecological Financial	Model 1		0.39	0.14	0.14	21.7**
	Operations					
		0.39**				

	Model 2		0.42	0.16	0.02	3.3
	Operations	0.31*				
		*				
	Probability	0.17				
	Model 3		0.42	0.16	0.01	0.9
	Operations					
		0.28**				
	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations	0.29				
	Probability	0.15				
	Geometry	0.10				
	Algebra	-0.03				
Ecological Medical	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations	0.24*				
	Probability	0.09				
	Model 3		0.41	0.15	0.01	1.7
	Operations					
		0.27**				
	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.42	0.15	0.00	0.6
	Operations	0.26				
	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
Ratio Bias	Model 1		0.28	0.07	0.07	10.2*
	Algebra					
		0.28**				
	Model 2		0.31	0.08	0.016	2.13
	Algebra	0.21 *				
	Geometry	0.15				
	Model 3		0.32	0.08	0.01	0.9
	Algebra					
		0.17**				
	Geometry	0.12				

	Operations	0.10				
	Model 4		0.32	0.07	0.00	0.43
	Algebra	0.15				
	Geometry	0.11				
	Operations	0.08				
	Probability	0.07				
Inter-temporal	Model 1		0.39	0.14	0.14	21.7**
	Operations					
		0.39**				
	Model 2		0.42	0.16	0.02	3.3
	Operations	0.31*				
		*				
	Probability	0.17				
	Model 3		0.42	0.16	0.01	0.9
	Operations					
		0.28**				
	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations	0.29				
	Probability	0.15				
	Geometry	0.10				
	Algebra	-0.03				
Intelligence	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations	0.24*				
	Probability	0.09				
	Model 3		0.41	0.15	0.01	1.7
	Operations					
		0.27**				
	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.42	0.15	0.00	0.6
	Operations	0.26				
	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
BNT	Model 1		0.39	0.14	0.14	21.7**

	Operations				
		0.39**			
	Model 2		0.42	0.16	0.02
	Operations	0.31*			3.3
		*			
	Probability	0.17			
	Model 3		0.42	0.16	0.01
	Operations				0.9
		0.28**			
	Probability	0.14			
	Geometry	0.09			
	Model 4		0.42	0.15	0.00
	Operations	0.29			0.06
	Probability	0.15			
	Geometry	0.10			
	Algebra	-0.03			
BNT-S	Model 1		0.36	0.12	0.12
					18.4**
	Operations				
		0.28**			
	Model 2		0.40	0.15	0.03
	Operations	0.24*			4.0*
	Probability	0.09			
	Model 3		0.41	0.15	0.01
	Operations				1.7
		0.27**			
	Probability	0.12			
	Geometry	-0.10			
	Model 4		0.42	0.15	0.00
	Operations	0.26			0.6
	Probability	0.11			
	Geometry	-0.11			
	Algebra	0.04			
Schwartz	Model 1		0.39	0.14	0.14
					21.7**
	Operations				
		0.39**			
	Model 2		0.42	0.16	0.02
	Operations	0.31*			3.3
		*			
	Probability	0.17			
	Model 3		0.42	0.16	0.01
	Operations				0.9

		0.28**				
	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations	0.29				
	Probability	0.15				
	Geometry	0.10				
	Algebra	-0.03				
Lipkus	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations	0.24*				
	Probability	0.09				
	Model 3		0.41	0.15	0.01	1.7
	Operations					
		0.27**				
	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.42	0.15	0.00	0.6
	Operations	0.26				
	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
Weller	Model 1		0.39	0.14	0.14	21.7**
	Operations					
		0.39**				
	Model 2		0.42	0.16	0.02	3.3
	Operations	0.31*				
	Probability	0.17				
	Model 3		0.42	0.16	0.01	0.9
	Operations					
		0.28**				
	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations	0.29				
	Probability	0.15				
	Geometry	0.10				

	Algebra	-0.03				
NUMi	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations	0.24*				
	Probability	0.09				
	Model 3		0.41	0.15	0.01	1.7
	Operations					
		0.27**				
	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.42	0.15	0.00	0.6
	Operations	0.26				
CRT	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
	Model 1		0.39	0.14	0.14	21.7**
	Operations					
		0.39**				
	Model 2		0.42	0.16	0.02	3.3
	Operations	0.31*				
		*				
	Probability	0.17				
	Model 3		0.42	0.16	0.01	0.9
	Operations					
		0.28**				
ADMC	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations	0.29				
	Probability	0.15				
	Geometry	0.10				
	Algebra	-0.03				
	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations	0.24*				
	Probability	0.09				

	Model 3	0.41	0.15	0.01	1.7
	Operations				
		0.27**			
	Probability	0.12			
	Geometry	-0.10			
	Model 4	0.42	0.15	0.00	0.6
	Operations	0.26			
	Probability	0.11			
	Geometry	-0.11			
	Algebra	0.04			
Ecological Total	Model 1	0.39	0.14	0.14	21.7**
	Operations				
		0.39**			
	Model 2	0.42	0.16	0.02	3.3
	Operations	0.31*			
		*			
	Probability	0.17			
	Model 3	0.42	0.16	0.01	0.9
	Operations				
		0.28**			
	Probability	0.14			
	Geometry	0.09			
	Model 4	0.42	0.15	0.00	0.06
	Operations	0.29			
	Probability	0.15			
	Geometry	0.10			
	Algebra	-0.03			
Standard Risky	Model 1	0.36	0.12	0.12	18.4**
	Operations				
		0.28**			
	Model 2	0.40	0.15	0.03	4.0*
	Operations	0.24*			
	Probability	0.09			
	Model 3	0.41	0.15	0.01	1.7
	Operations				
		0.27**			
	Probability	0.12			
	Geometry	-0.10			
	Model 4	0.42	0.15	0.00	0.6
	Operations	0.26			

	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				
Risk Literacy	Model 1		0.39	0.14	0.14	21.7**
	Operations					
		0.39**				
	Model 2		0.42	0.16	0.02	3.3
	Operations					
		0.31*				
		*				
	Probability	0.17				
	Model 3		0.42	0.16	0.01	0.9
	Operations					
		0.28**				
	Probability	0.14				
	Geometry	0.09				
	Model 4		0.42	0.15	0.00	0.06
	Operations					
		0.29				
	Probability	0.15				
	Geometry	0.10				
	Algebra	-0.03				
SDMS-Total	Model 1		0.36	0.12	0.12	18.4**
	Operations					
		0.28**				
	Model 2		0.40	0.15	0.03	4.0*
	Operations					
		0.24*				
	Probability	0.09				
	Model 3		0.41	0.15	0.01	1.7
	Operations					
		0.27**				
	Probability	0.12				
	Geometry	-0.10				
	Model 4		0.42	0.15	0.00	0.6
	Operations					
		0.26				
	Probability	0.11				
	Geometry	-0.11				
	Algebra	0.04				

Appendix I

Step wise regression for all ability and decision tasks

Table I-1. *Step wise regression analysis following all ability tasks*

Numeracy	Model	β	R	R ²	R ² Change	F Change
BNT	Model 1		0.51	0.26	0.26	41.7**
	Operations	0.51**				
	Model 2		0.53	0.28	0.03	4.5*
	Operations	0.42**				
Lipkus	Probability	0.19*				
	Model 1		0.39	0.15	0.15	21.4**
	Operation	0.39**				
Schwartz	Model 1		0.32	0.10	0.10	
	Probability	0.32**				13.9**
	Model 2		0.37	0.13	0.03	
	Probability	0.23*				4.5*
	Operations	0.20*				
BNT-S	Model 1		0.52	0.27	0.27	45.0**
	Operations	0.52**				
	Model 2		0.56	0.32	0.05	8.1*
	Operations	0.40**				
Weller	Probability	0.24*				
	Model 1		0.49	0.24	0.24	
	Operations	0.49**				37.6**
	Model 2		0.54	0.29	0.05	8.6*
	Operations	0.37**				
NUMi	Probability	0.26*				
	Model 1		0.48	0.23	0.23	37.3**
	probability	0.48**				
	Model 2		0.53	0.28	0.05	7.6*
	Probability	0.38*				
	Geometry	0.24*				

CRT	Model 1		0.49	0.24	0.24	
	Operations	0.49**				38.5**
	Model 2		0.54	0.30	0.06	
	Operations	0.36**				9.7*
Intelligence	Probability	0.27*				
	Model 1		0.49	0.24	0.24	
	Operations	0.49**				38.5**
	Model 2		0.54	0.30	0.06	
	Operations	0.36**				9.7*
	Probability	0.27*				

** Correlation is significant at 0.01 level; *Correlation significant at 0.05

Analysis of Skewness

Three measures Lipkus, Schwartz, and NUMi were significantly negatively skewed (skewness > +/-1). To see any effect of skewness, I log transformed the three measures and run the step wise regression again. All three measures almost followed the same pattern.

Table I-2. *Step wise regression when measures were log transformed for skewness*

Numeracy	Model	β	R	R ²	R ² Change	F Change
Schwartz	Model 1		0.32	0.10	0.10	14.1**
	Operations	0.32**				
	Model 2		0.37	0.14	0.04	4.9*
	Operations	0.22*				
	Probability	0.21*				
Lipkus	Model 1		0.41	0.17	0.17	24.5**
	Operation	0.41**				
NUMi	Model 1		0.53	0.29	0.29	48.8**
	Probability	0.53**				
	Model 2		0.58	0.34	0.06	10.1*
	Probability	0.41**				
	Operations	0.27*				

Table I-3. *Step wise regression for all decision tasks*

Numeracy	Model	β	R	R ²	R ² Change	F Change
Framing	Model 1		0.51	0.26	0.26	41.7**
	Operations	0.51**				
	Model 2		0.53	0.28	0.03	4.5*
	Operations	0.42**				
	Probability	0.19*				
Social Norms	Model 1			0.15	0.15	21.4**
	Operation	0.39**	0.39			
Confidence	Model 1			0.10	0.10	3.9**
	Probability	0.32**	0.32			
	Probability	0.23*				
	Operations	0.20*				
Sunk Cost	Model 1		0.52	0.27	0.27	45.0**
	Operations	0.52**				
	Model 2		0.56	0.32	0.05	8.1*
	Operations	0.40**				
	Probability	0.24*				
Decision Rules	Model 1		0.49	0.24	0.24	37.6**
	Operations	0.49**				
	Model 2		0.54	0.29	0.05	8.6*
	Operations	0.37**				
	Probability	0.26*				
Path Independ	Model 1		0.48	0.23	0.23	37.3**
	probability	0.48**				
	Model 2		0.53	0.28	0.05	7.6*
	Probability	0.38*				
	Geometry	0.24*				
Risk Perception	Model 1		0.49	0.24	0.24	38.5**
	Operations	0.49**				
	Model 2		0.54	0.30	0.06	9.7*
	Operations	0.36**				
	Probability	0.27*				

Ecolog Financial	Model 1		0.26	0.26	41.7**
			0.51		
	Operations	0.51**			
	Model 2		0.28	0.03	4.5*
			0.53		
	Operations	0.42**			
	Probability	0.19*			
Ecolog Medical	Model 1		0.15	0.15	21.4**
			0.39		
	Operation	0.39**			
Ratio Bias	Model 1		0.10	0.10	3.9**
			0.32		
	Probability	0.32**			
	Probability	0.23*			
	Operations	0.20*			
Lotteries	Model 1		0.27	0.27	45.0**
			0.52		
	Operations	0.52**			
	Model 2		0.32	0.05	8.1*
			0.56		
	Operations	0.40**			
	Probability	0.24*			
Intertemporal	Model 1		0.24	0.24	37.6**
			0.49		
	Operations	0.49**			
	Model 2		0.54	0.05	8.6*
			0.29		
	Operations	0.37**			
	Probability	0.26*			
ADMC	Model 1		0.23	0.23	37.3*
			0.48		*
	probability	0.48**			
	Model 2		0.28	0.05	7.6*
			0.53		
	Probability	0.38*			
	Geometry	0.24*			
Ecological Total	Model 1		0.24	0.24	38.5*
			0.49		*
	Operations	0.49**			
	Model 2		0.54	0.06	9.7*
			0.30		

	Operations	0.36**				
	Probability	0.27*				
Standard Risky	Model 1			0.24	0.24	37.6**
			0.49			
	Operations	0.49**				
			0.54		0 .05	8.6*
	Model 2			0.29		
	Operations	0.37**				
	Probability	0.26*				
Risk Literacy	Model 1			0.23	0.23	37.3*
			0.48			*
	probability	0.48**				
	Model 2			0.28	0.05	7.6*
			0.53			
	Probability	0.38*				
	Geometry	0.24*				
SDMS-Total	Model 1			0.24	0.24	38.5*
			0.49			*
	Operations	0.49**				
	Model 2		0.54		0 .06	9.7*
			0.30			
	Operations	0.36**				
	Probability	0.27*				

Appendix J

Simultaneous regression models: models of best fit

Table J-1. *Model of best fit (simultaneous regression models) all four components predicting performance on all **ability** tasks.*

	Operations	Probability	Geometry	Algebra	R	R ²
BNT	0.38**	0.15	0.15	-0.03	0.54	0.30
BNT-A	0.35**	0.17	0.12	-0.03	0.51	0.26
BNT-S	0.40**	0.24*	0.12	-0.09	0.57	0.33
CRT	0.32**	0.23*	0.02	0.1	0.55	0.30
Schwartz	0.25*	0.27*	0.02	-0.14	0.38	0.15
Lipkus	0.35**	0.10	0.07	-0.08	0.40	0.16
Weller	0.35**	0.24*	0.08	-0.02	0.54	0.29
NUMI	0.15	0.31**	0.18*	0.05	0.55	0.30
Intelligence	0.38**	0.25**	0.15	-0.11	0.57	0.33

** Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

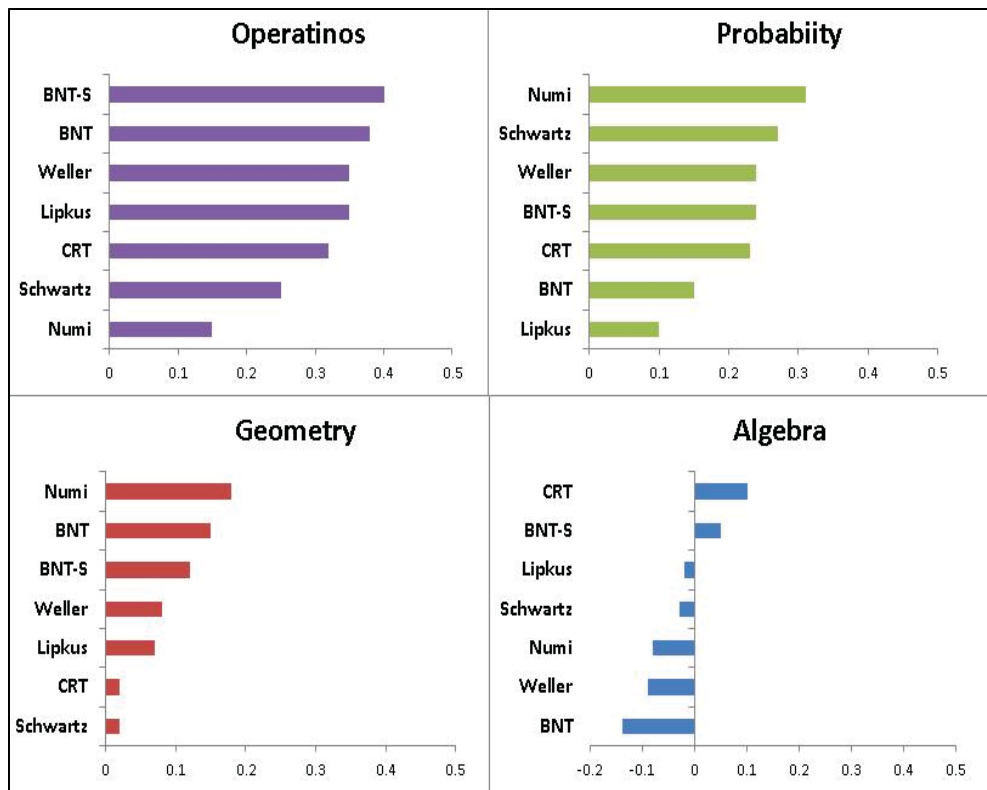


Figure J-1. Standardized beta coefficients (using multiple simultaneous regression) showing relationship of existing numeracy measures with numeracy components

Table J-2. *Model of best fit (simultaneous regression models) all four components predicting performance on all **decision tasks**.*

	Operations	Probability	Geometry	Algebra	R	R2
Intertemporal	.20*	.29**	.01	-.12	.37	.14
Lotteries	.01	.08	.10	.03	.15	.02
Ratio Bias	.08	.07	.11	.15	.32	.10
Framing	-.02	.15	.17	.02	.27	.08
Framing_Risk	-.02	.07	.17	.14	.30	.09
Framing_Attribute	-.02	.19	.06	-.19	.19	.04
Social Norms	.03	.21*	.09	.06	.32	.10
Consistency Risk					.24	.06
Perception	.03	.22*	-.05	.05		
Sunk Cost	.02	.20*	-.07	.15	.29	.08
App Decision					.31	.09
Rules	.26*	.11	-.11	.04		
Over Confidence	.01	-.05	-.17	-.14	.29	.09
Path					.14	.02
Independence	.08	-.04	.10	.03		
DOI	.01	.01	.10	.04	.13	.02
Eco Financial	.29*	.15	.10	-.03	.42	.18
Eco Medical Dec.	.21*	.13	.11	.09	.42	.18
ADMC Total	.18	.25*	.08	.12	.51	.26
Standard Risky	.18	.25*	.12	.03	.46	.21
Risk Literacy	.32**	.21*	.14	.04	.57	.32
SDMS Total	.30**	.25*	.14	.07	.59	.35

** Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

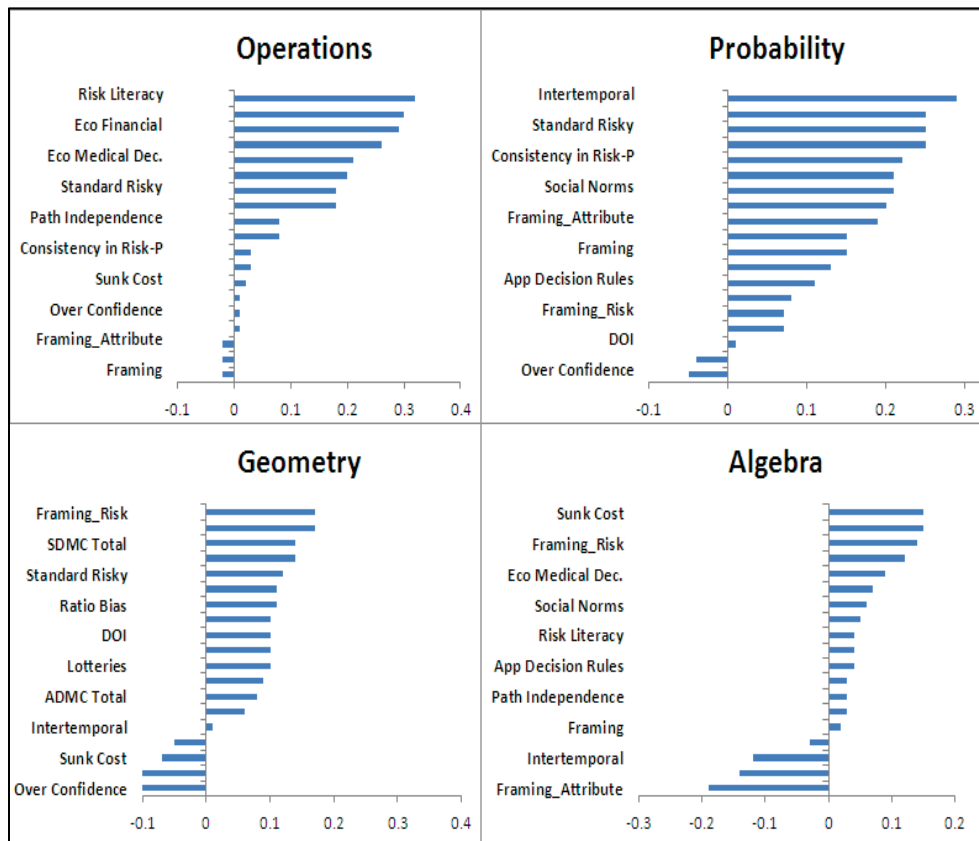


Figure J-2. Graph showing standardized beta coefficients for each of the decision task Predicted by four components entered in to model simultaneously

Table J-3. *Model of best fit (simultaneous regression models) Components along with intelligence predicting performance on all **decision tasks** (Standardized betas)*

	Opera	Prob	Geo	Algebra	Intel	R	R2
Intertemporal	.21*	.30**	.02	-.12	-.05	.37	.14
Lotteries	-.002	.07	.10	-.02	.04	.15	.02
Ratio Bias	.09	.08	.11	.14	-.02	.32	.10
Framing	-.04	.14	.16	.03	.05	.28	.08
Framing Risk	-.06	.04	.15	.12	.12	.32	.10
Framing					-.10	.21	.04
Attribute	.02	.21	.08	-.20			
Social Norms	-.03	.17	.06	.07	.18	.36	.13
Consistency					.07	.25	.06
in Risk-P	.01	.20	-.06	.06			
Sunk Cost	-.07	.14	-.12	.17	.26*	.36	.13
App Decision					.29**	.39	.15
Rules	.15	.03	-.15	.08			
Over					-.03	.29	.09
Confidence	.03	-.05	-.16	-.15			
Path					.10	.17	.03
Independence	.04	-.07	.08	.03			
DOI	-.05	-.04	.06	.05	.19	.20	.04
Eco Financial	.25*	.13	.08	-.02	.10	.43	.19
Eco Medical					.002	.42	.18
Dec.	.21*	.13	.11	.09			
ADMC Total	.07	.18	.03	.14	.31**	.57	.32
Risk Literacy					-.02	.46	.21
Total	.19	.25*	.12	.03			
SDMS Total	.26*	.22*	.12	.07	.11	.60	.36

** Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

Hierarchical linear regression to see if adding intelligence significantly improves the model or not: I entered the four components on first step and then intelligence at the second step; Intelligence brought significant change in three out of 18 decision tasks namely Sunk cost, Applying decision rules, and ADMC with 4%, 6%, and 7% improvement in the model respectively. BNT-C was the strongest predictor for most of the decision tasks (predicted 12 out of 18 decision tasks) even after controlling for intelligence.

Table J-4. *Model of best fit: Simultaneous regression when Operations and probability were entered into the model with combined scores on geometry and algebra (Standardized betas)*

	Operations	Probability	Geo+Algeb	R	R2
Intertemporal	.19	.29**	-.09	.36	.13
Lotteries	.01	.08	.07	.13	.02
Ratio Bias	.08	.07	.22`	.32	.10
Framing	-.04	.14	.18	.26	.07
Framing_Risk	-.02	.07	.27*	.30	.09
Framing_Attribute	-.04	.18	-.11	.15	.02
Social Norms	.03	.21*	.14	.32	.10
Consistency in Risk-P	.04	.22*	-.002	.24	.06
Sunk Cost	.03	.21*	.06	.27	.07
App Decision				.30	.09
Rules	.26*	.11	-.06		
Over Confidence	.02	-.05	-.27*	.29	.09
Path				.14	.02
Independence	.07	-.05	.11		
DOI	.01	.01	.12	.13	.02
Eco Financial	.28**	.15	.07	.42	.18
Eco Medical Dec.	.21*	.13	.17	.42	.18
ADMC Total	.19	.25*	.18	.51	.26
Standard Risky	.17	.24*	.13	.45	.21
Risk Literacy	.31**	.21*	.15	.56	.32
SDMS Total	.29**	.25*	.18	.59	.35

Note: Only framing and overconfidence were the two variables that were significantly predicted by geometry and algebra once I combined them, all other results remained the same.

Table J-5. *BNT-C vs Intelligence: Simultaneously regression models; (standardized beta coefficients)*

	BNT-C	Intel	R	R2
Intertemporal	.29**	.02	.30	.09
Lotteries	.10	.05	.14	.02
Ratio Bias	.33**	-.02	.32	.10
Framing	.22*	.04	.25	.06
Framing Risk	.24*	.09	.30	.09
Framing Attribute	.05	-.06	.06	.003
Social Norms	.22*	.17	.34	.12
Consistency in Risk-P	.17	.07	.21	.04
Sunk Cost	.13	.22*	.31	.10
App Decision Rules	.08	.30**	.35	.12
Over Confidence	-.27*	-.01	.27	.07
Path Independence	.06	.10	.15	.02
DOI	.02	.18	.19	.04
Eco Financial	.32**	.14	.41	.17
Eco Medical Dec.	.40**	.03	.41	.17
ADMC Total	.34**	.30**	.56	.32
Risk Literacy Total	.44**	.01	.44	.20
SDMS Total	.51**	.14	.59	.35

BNT-C predicted overall total decision making competence 4 times more than intelligence

Table J-6. *BNT-C vs all other ability measures predicting decision making*
Standardized beta coefficients from simultaneous regression

	BNT-						
	C	BNT	Schwartz	Lipkus	CRT	Weller	NUMI
Intertemporal	.22*	.04	.51**	-.28	.11	-.09	.02
Lotteries	-.01	.23*	-.21	.14	-.03	.16	-.05
Ratio Bias	.38**	-.25*	.23	-.14	.02	.07	.03
Framing	.15	.03	.19	-.08	.02	-.07	.14
Framing_Risk	.23*	.09	.08	.02	.06	-.18	.08
Framing_Attribute	-.09	-.08	.25	-.19	-.06	.17	.16
Social Norms	.14	.03	.14	-.16	-.09	.13	.29**
Consistency in							
Risk-P	.22*	.07	.11	.11	.07	-.20	-.12
Sunk Cost	.20	.31**	.13	-.15	-.02	-.004	.14
App Decision							
Rules	.06	-.04	-.03	-.29	-.22	.54*	.26*
Over Confidence	-.21	-.07	.15	-.24	-.15	.29	-.12
Path							
Independence	.01	.01	.05	-.06	.07	-.05	.18
DOI	.03	.01	.07	-.13	.43*	-.38	.15
Eco Financial	.23*	-.02	-.13	.04	-.11	.41	.12
Eco Medical Dec.	.31**	-.06	.002	.03	.22	-.09	.09
ADMC_Total	.24*	.19*	.16	-.19	.02	.08	.27**
Standard Risky	.40**	-.08	.38*	-.22	.04	.05	-.06
Risk Literacy	.39**	-.06	.03	-.03	.06	.20	.11
SDMS Total	.38**	-.02	.04	-.05	.06	.20	.15

** Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

Table J-7. *BNT-C vs all other ability measures predicting decision making along with INTELLIGENCE; Standardized beta coefficients from simultaneous regression*

	BNT- C	Schw	Lipkus	CRT	Weller	NUMI	Int.	R	R2
Intertemporal	.04	.24*	.51**	-.29	.11	-.09	.02	-.04	.43 .19
Lotteries	.23*	-.01	-.21	.14	-.03	.16	-.05	-.01	.29 .08
Ratio Bias	-.25*	.39**	.23	-.14	-.01	.07	.03	-.02	.41 .17
Framing	.03	.15	.19	-.08	.02	-.07	.14	.004	.30 .09
Framing Risk	.08	.20	.06	.01	.05	-.18	.08	.08	.31 .10
Framing Attribu	-.08	-.05	.27	-.19	-.04	.16	.16	-.13	.29 .08
Social Norms	.01	.09	.12	-.18	-.12	.14	.30**	.14	.43 .19
Consist. Risk Per	.07	.21	.11	.11	.07	-.20	-.12	.05	.25 .06
Sunk Cost	.29**	.13	.10	-.16	-.05	.002	-.14	.20`	.42 .18
App Decis Rules	-.05	-.03	-.06	-.29	-.26	.53*	.26*	.28**	.45 .21
Over Confidence	-.07	-.21	.15	-.24	-.15	.29	-.12	-.01	.35 .12
Path Independ	.01	-.02	.04	-.06	.06	-.05	.18	.09	.21 .04
DOI	-.02	-.05	.04	-.14	.39*	-.38	.15	.21`	.37 .14
Eco Financial	-.03	.19	-.13	.04	-.11	.41	.12	.07	.49 .24
Eco Medical Dec.	-.06	.32**	.002	.03	.22	-.09	.09	-.01	.45 .20
ADMC_Total	.16	.16	.13	-.21	-.05	.10	.28**	.24*	.64 .41
Standard Risky	-.08	.41**	.39*	-.22	.05	.05	-.06	-.04	.52 .27
Risk Literacy	-.06	.38**	.03	-.03	.05	.20	.11	.03	.60 .36
SDMS Total	-.02	.36**	.03	-.05	.04	.21	.15	.05	.64 .41

** Correlation is significant at the 0.01 level; *Correlation is significant at the 0.05 level

Appendix K

Composite Analysis

BNT vs other numeracy measures predicting decision making (Step wise regression)

Table K-1. *Step wise regression for each Decision Task with all numeracy measures*

Decision Task	Model	β	R	R ²	R ² change	F Change
Inter-temporal	Model 1		0.32	0.10	0.10	14.2**
	Schwartz	0.32**				
	Model 2		0.39	0.15	0.05	6.4*
	Schwartz	.26**				
	BNT-C	.22*				
	Model 3		0.42	0.18	0.03	4.6*
	Schwartz	0.47**				
	BNT-C	0.25				
Lotteries	Lipkus	-0.28*				
	Model		0.25	0.06	0.06	8.4**
Ratio Bias	BNT	0.25**				
	Model		0.32	0.10	0.10	13.7**
	BNT-C	0.32**				
	Model 2		0.37	0.14	0.04	4.9*
	BNT-C	0.43**				
Framing	BNT	-0.22*				
	Model		0.25	0.06	0.06	7.6**
Framing Risk	BNT-C	0.25**				
	Model		0.29	0.08	0.08	10.4**
Framing Attribute	BNT-C	0.29**				
	Model		0.21	0.04	0.04	5.2*
Social Norms	Schwartz	0.21*				
	Model		0.38	0.14	0.24	19.5**
Consistency in RP	NUMI	0.38**				
	Model		0.20	0.04	0.04	5.0*
	BNT-C	0.20*				

Sunk Cost	Model		0.33	0.11	0.11	14.7**
	BNT	0.33**				
App Decision Rules	Model		0.32	0.10	0.10	13.5**
	NUMI	0.32**				
Over Confidence	Model		0.27	0.07	0.07	9.6**
	BNT-C	- 0.27**				
Path Independence	Model		0.19	0.04	0.04	4.2*
	NUMI	0.19*				
DOI	Model		0.18	0.03	0.03	4.0*
	CRT	0.18*				
	Model 2		0.30	0.09	0.06	7.0**
	CRT	0.48**				
	Weller	- 0.38**				
Eco. Financial	Model1		0.42	0.18	0.18	25.9**
	Weller	0.42**				
	Model2		0.47	0.22	0.04	6.8*
	Weller	0.30**				
	BNT-C	0.24*				
Eco Medical	Model		0.41	0.17	0.17	24.8**
	BNT-C	0.41**				
ADMC	Model 1		0.50	0.25	0.25	39.2**
	BNT-C	0.50**				
	Model 2		0.57	0.32	0.07	11.6**
	BNT-C	0.34**				
	NUMI	0.31**				
	Model 3		0.60	0.35	0.03	5.7*
	BNT-C	0.26**				
	NUMI	0.28**				
	BNT	0.21*				
Risk Literacy	Model 1		0.44	0.20	0.20	29.8**
	BNT-C	0.44**				
	Model 2		0.50	0.25	0.05	8.7**
	BNT-C	0.37**				
	Schwartz	0.24**				

SDMS Total	Model 1		0.58	0.34	0.34	59.11**
	BNT-C	0.58**				
	Model 2		0.63	0.40	0.06	10.9**
	BNT-C	0.44**				
	Weller	0.28**				

When Intelligence was also included in the modelTable K-2. *Step wise regression for each Decision Task with all numeracy measures*

Decision Task	Model	β	R	R²	R² change	F Change
Inter-temporal	Model 1		0.32	0.10	0.10	14.2**
	Schwartz	0.32**				
	Model 2		0.39	0.15	0.05	6.4*
	Schwartz	.26**				
	BNT-C	.22*				
	Model 3		0.42	0.18	0.03	4.6*
	Schwartz	0.47**				
	BNT-C	0.25				
Lotteries	Model		0.25	0.06	0.06	8.4**
	BNT	0.25**				
Ratio Bias	Model 1		0.32	0.10	0.10	13.7**
	BNT-C	0.32**				
	Model 2		0.37	0.14	0.04	4.9*
	BNT-C	0.43**				
	BNT	-0.22*				
Framing	Model		0.25	0.06	0.06	7.6**
	BNT-C	0.25**				
Framing Risk	Model		0.29	0.08	0.08	10.4**
	BNT-C	0.29**				
Framing Attribute	Model		0.21	0.04	0.04	5.2*
	Schwartz	0.21*				
Social Norms	Model		0.38	0.14	0.24	19.5**
	NUMI	0.38**				
Consistency in RP	Model		0.20	0.04	0.04	5.0*
	BNT-C	0.20*				
Sunk Cost	Model		0.33	0.11	0.11	14.7**
	BNT	0.33**				
App Decision Rules	Model 1		0.32	0.10	0.10	13.5**
	Intelligence	0.34**	0.34	0.12	0.12	16.3**
	Model 2		0.40	0.16	0.04	6.2*

	Intelligence	0.26**				
	NUMI	0.22*				
Over Confidence	Model		0.27	0.07	0.07	9.6**
	BNT-C	-				
		0.27**				
Path Independence	Model		0.19	0.04	0.04	4.2*
	NUMI	0.19*				
DOI	Model 1		0.19	0.04	0.04	4.3*
	Intelligence	0.19*				
	Model 2		0.27	0.07	0.04	4.4*
	Intelligence	0.27**				
	Lipkus	-0.20*				
	Model 3		0.32	0.10	0.03	4.2*
	Intelligence	.19`				
	Lipkus	-.27**				
	CRT	.22*				
Eco. Financial	Model1		0.42	0.18	0.18	25.9**
	Weller	0.42**				
	Model2		0.47	0.22	0.04	6.8*
	Weller	0.30**				
	BNT-C	0.24*				
Eco Medical	Model		0.41	0.17	0.17	24.8**
	BNT-C	0.41**				
ADMC Total	Model 1		0.50	0.25	0.25	39.2**
	BNT-C	0.50**				
	Model 2		0.57	0.32	0.07	11.6**
	BNT-C	0.34**				
	NUMI	0.31**				
	Model 3		0.61	0.37	0.05	9.5**
	BNT-C	0.21**				
	NUMI	0.28**				
	Intelligence	0.27**				

Standard Risky	Model 1		0.44	0.20	0.20	29.8**
	BNT-C	0.44**				
	Model 2		0.50	0.25	0.05	8.7**
	BNT-C	0.37**				
	Schwartz	0.24**				
Risk Literacy	Model 1		0.54	0.30	0.30	51.4**
	BNT-C	0.54**				
	Model 2		0.59	0.35	0.05	9.4**
	BNT-C	0.42**				
	Weller	0.26**				
SDMS Total	Model 1		0.58	0.34	0.34	59.11**
	BNT-C	0.58**				
	Model 2		0.63	0.40	0.06	10.9**
	BNT-C	0.44**				
	Weller	0.28**				

Almost for all decision tasks structure remained the same, except for DOI, Applying decision rules, and Total superior decision making, where intelligence was a significant predictor in the model.

Appendix L

Competitive analysis: For each decision and ability tasks

(Model comparisons)

Competitive analysis

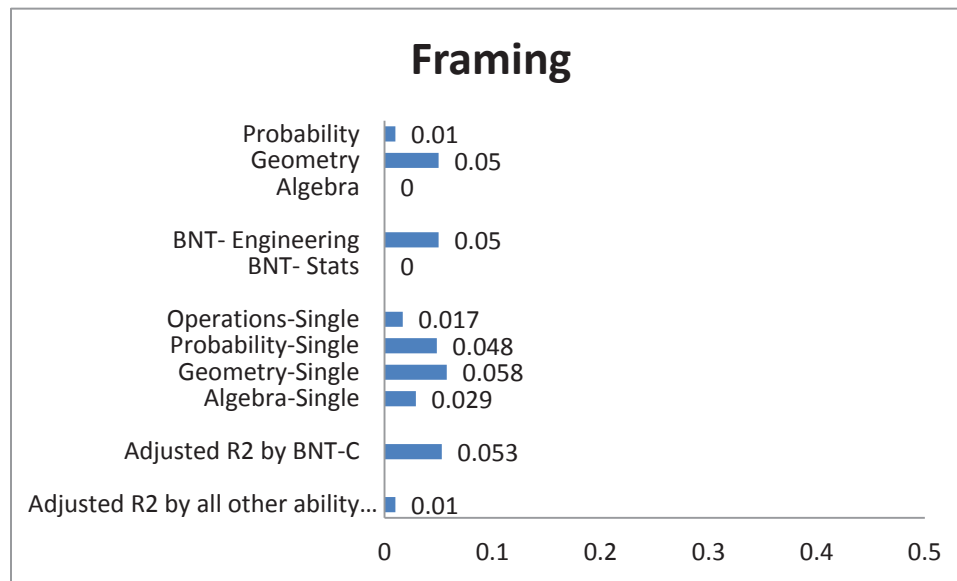


Figure L-1. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical and 2 simultaneous regression analysis; Hierarchical; one with 4 components and second with BNT-Stats, BNT-Engineering. Simultaneous; one with all other ability measures combined for maximum variance, and one with BNT-C as a single predictor

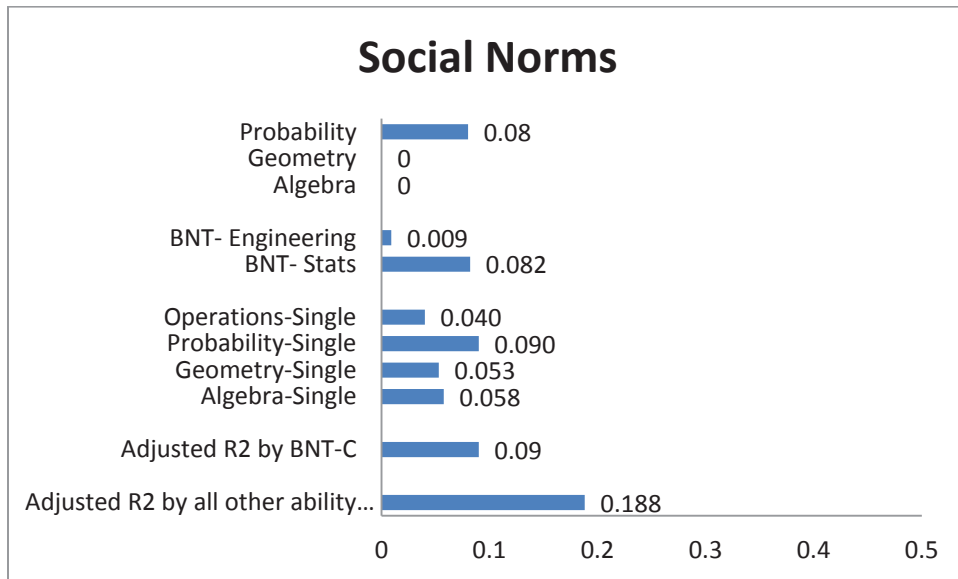


Figure L-2. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

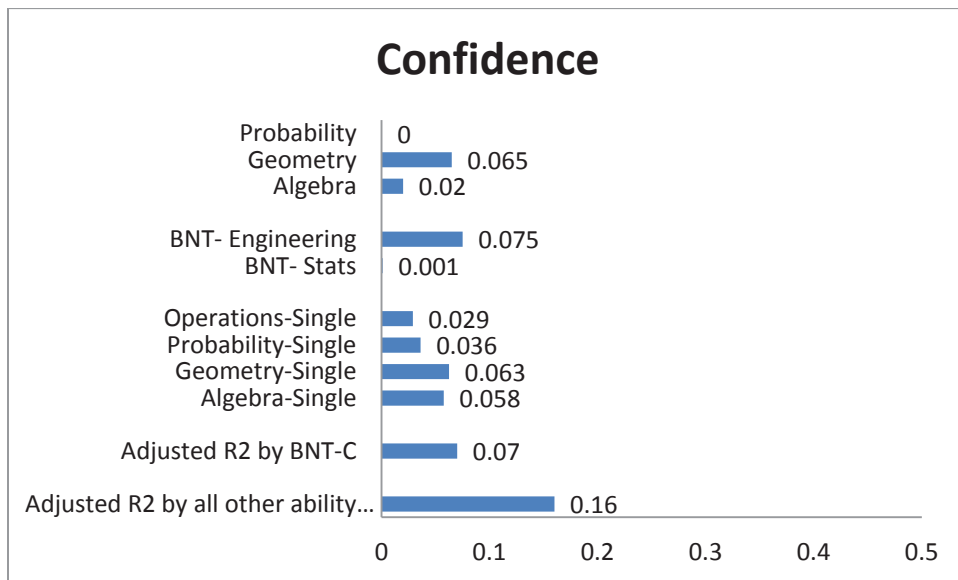


Figure L-3. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

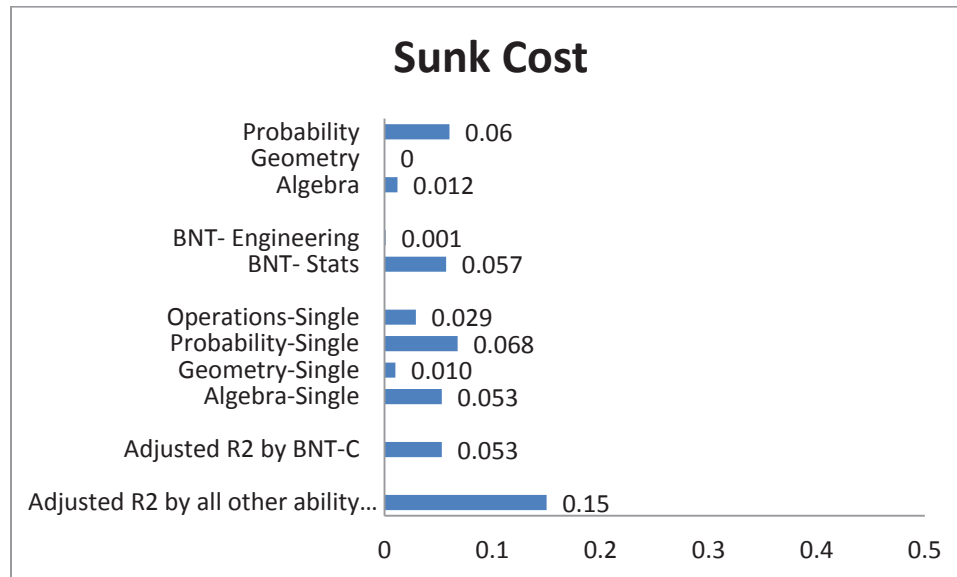


Figure L-4. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

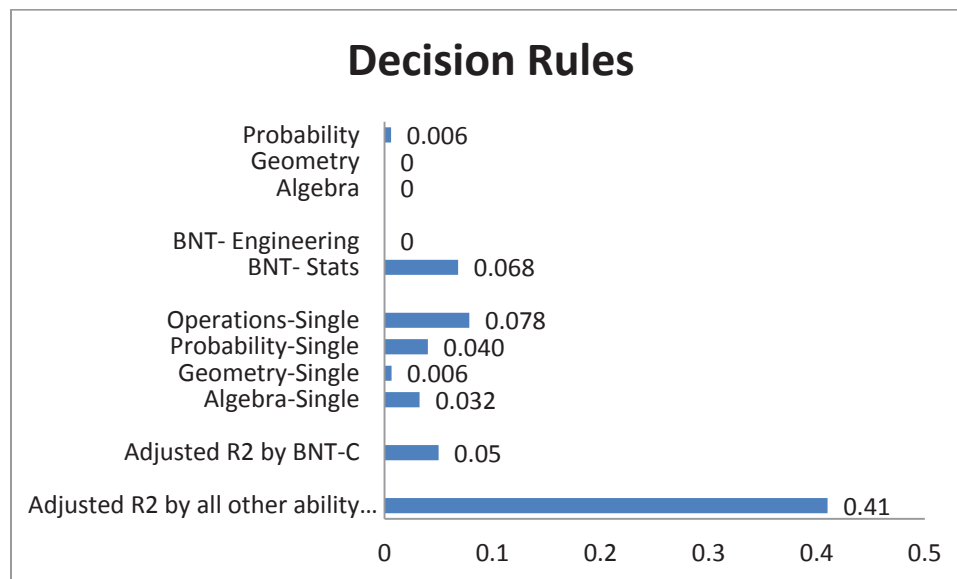


Figure L-5. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

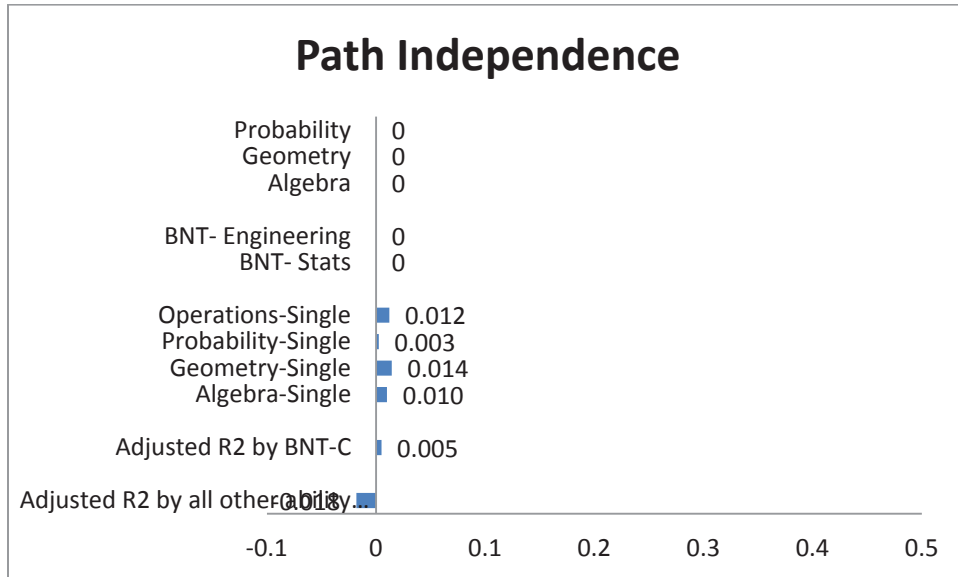


Figure L-6. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

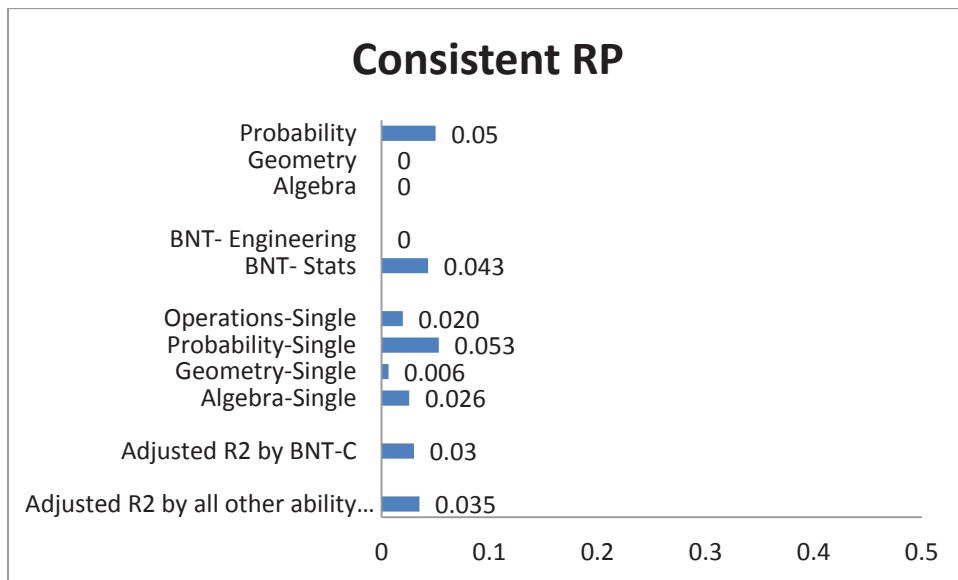


Figure L-7. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

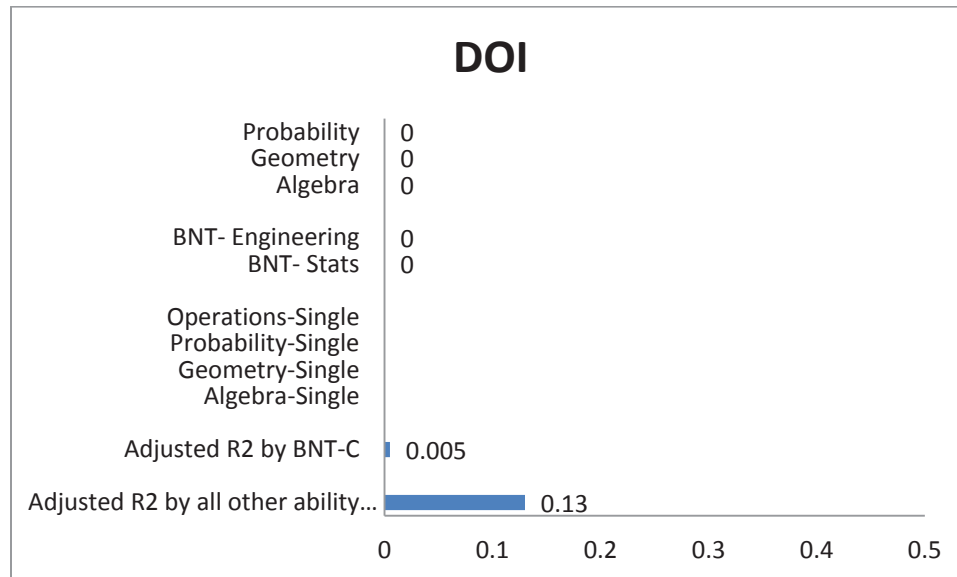


Figure L-8. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

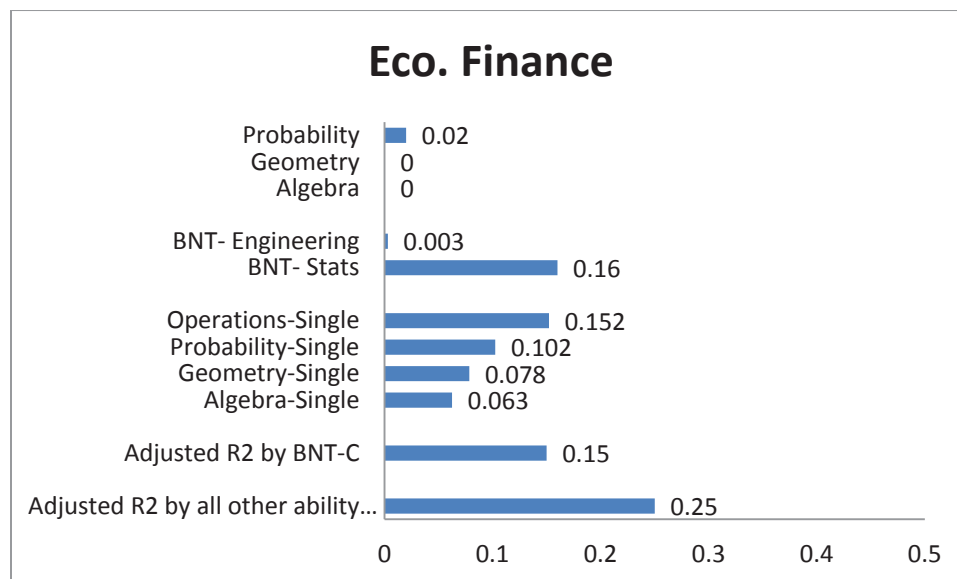


Figure L-9. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

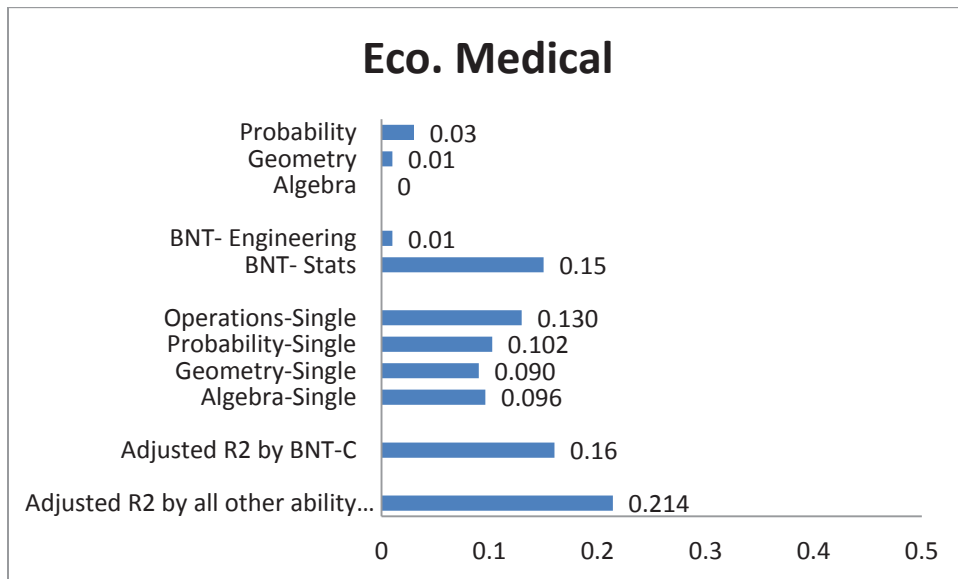


Figure L-10. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

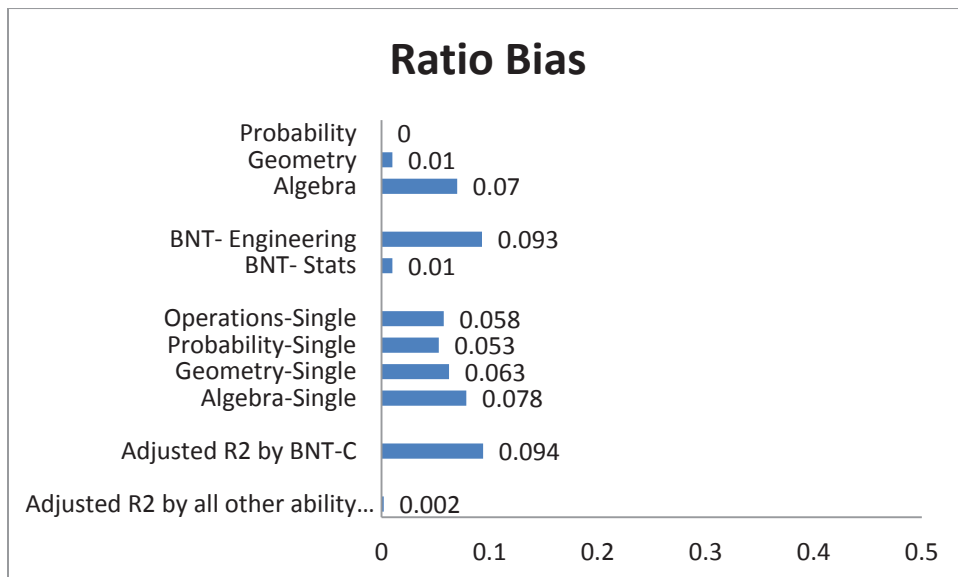


Figure L-11. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

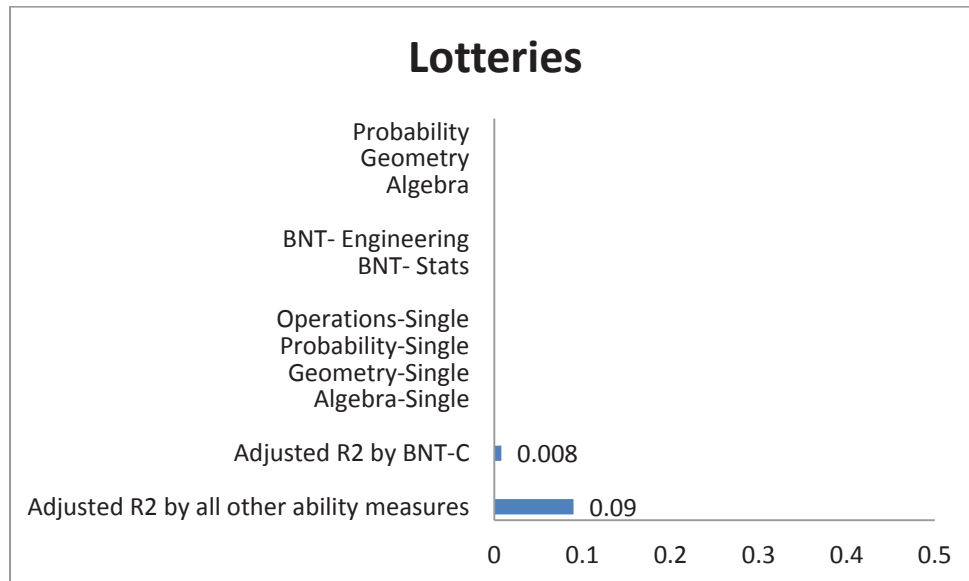


Figure L-12. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

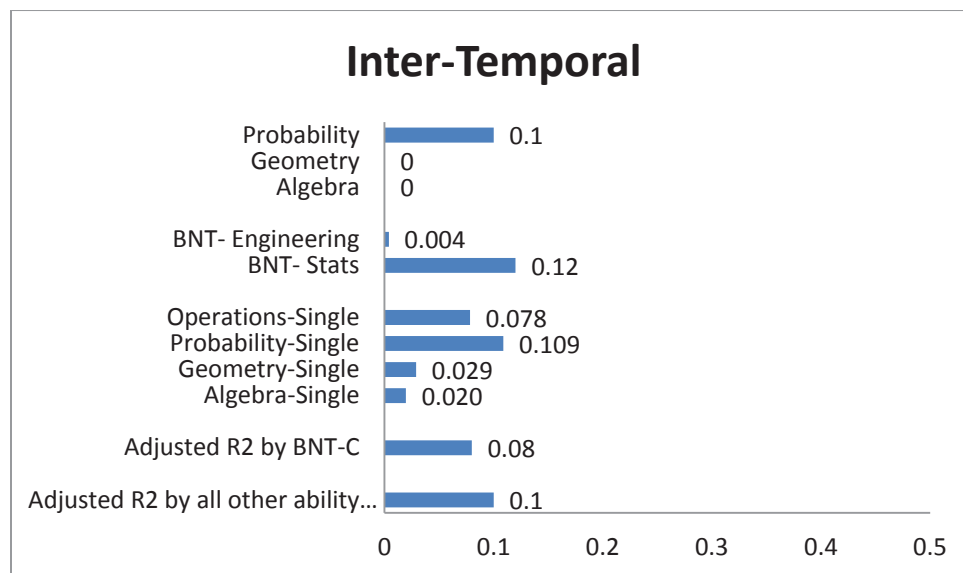


Figure L-13. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

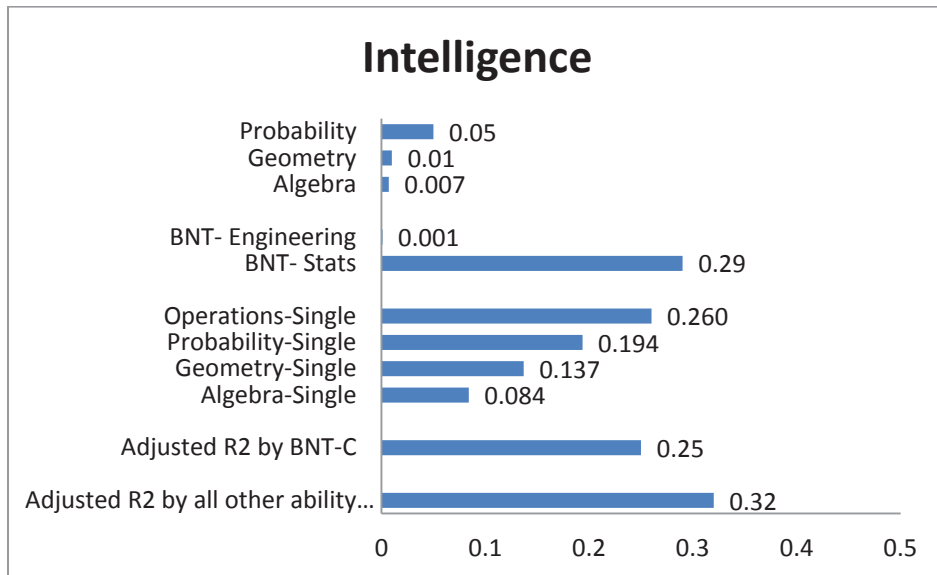


Figure L-14. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

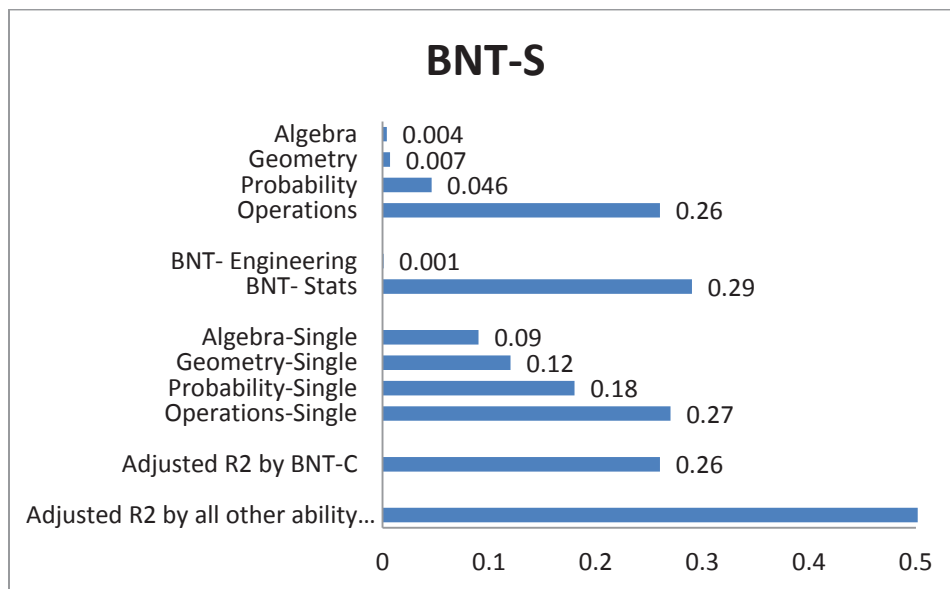


Figure L-15. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

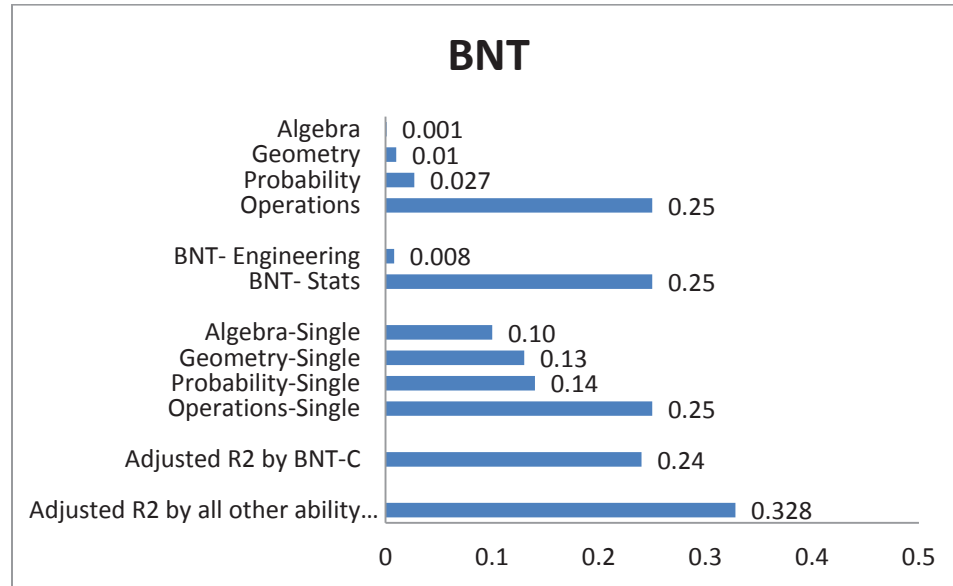


Figure L-16. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

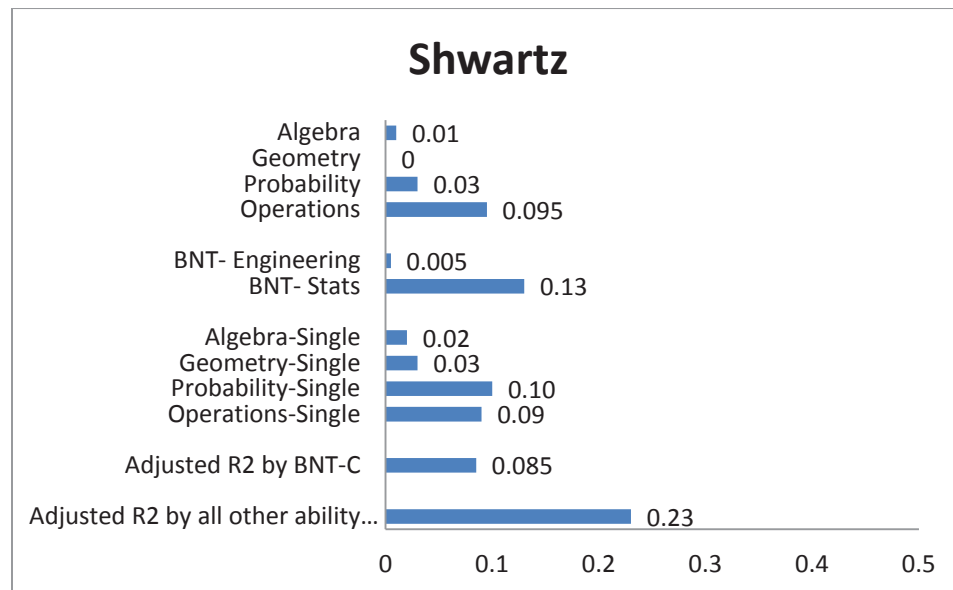


Figure L-17. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

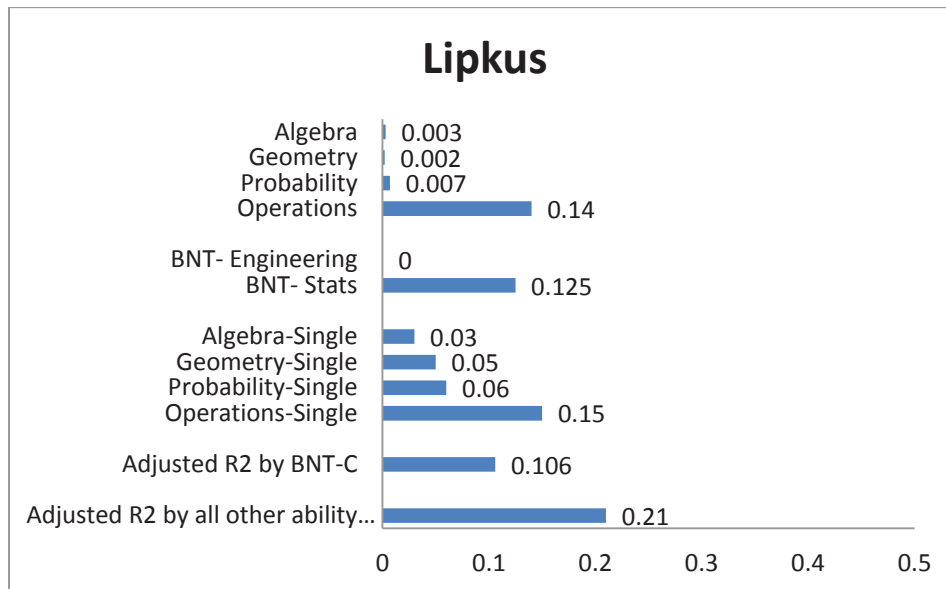


Figure L-18. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

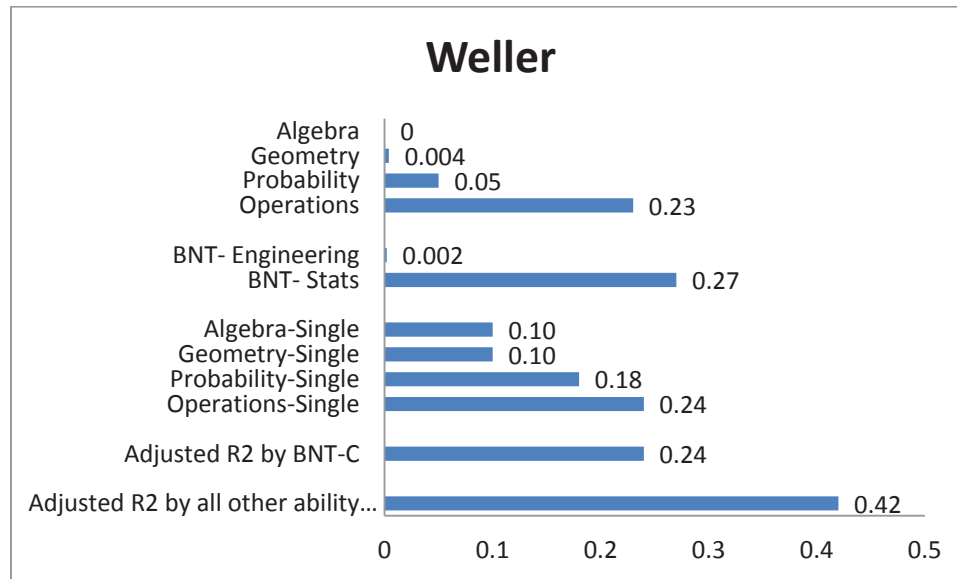


Figure L-19. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

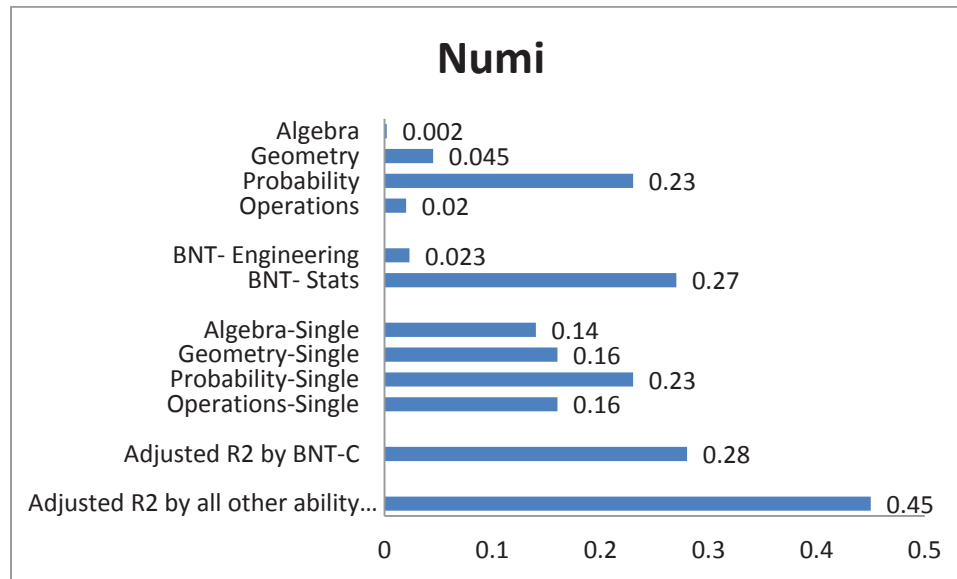


Figure L-20. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

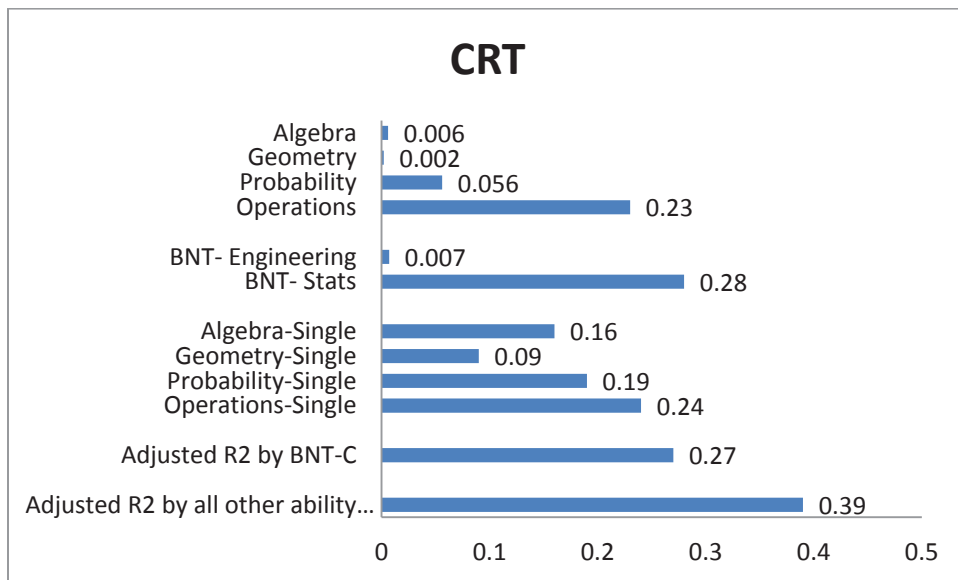


Figure L-21. Competitive analysis; results (length of bars showing adjusted R2) from 2 separate Hierarchical, 2 separate simultaneous regression analysis and one with BNT-C as independent predictor

Appendix M

Four components as single predictors

a) Decision Tasks

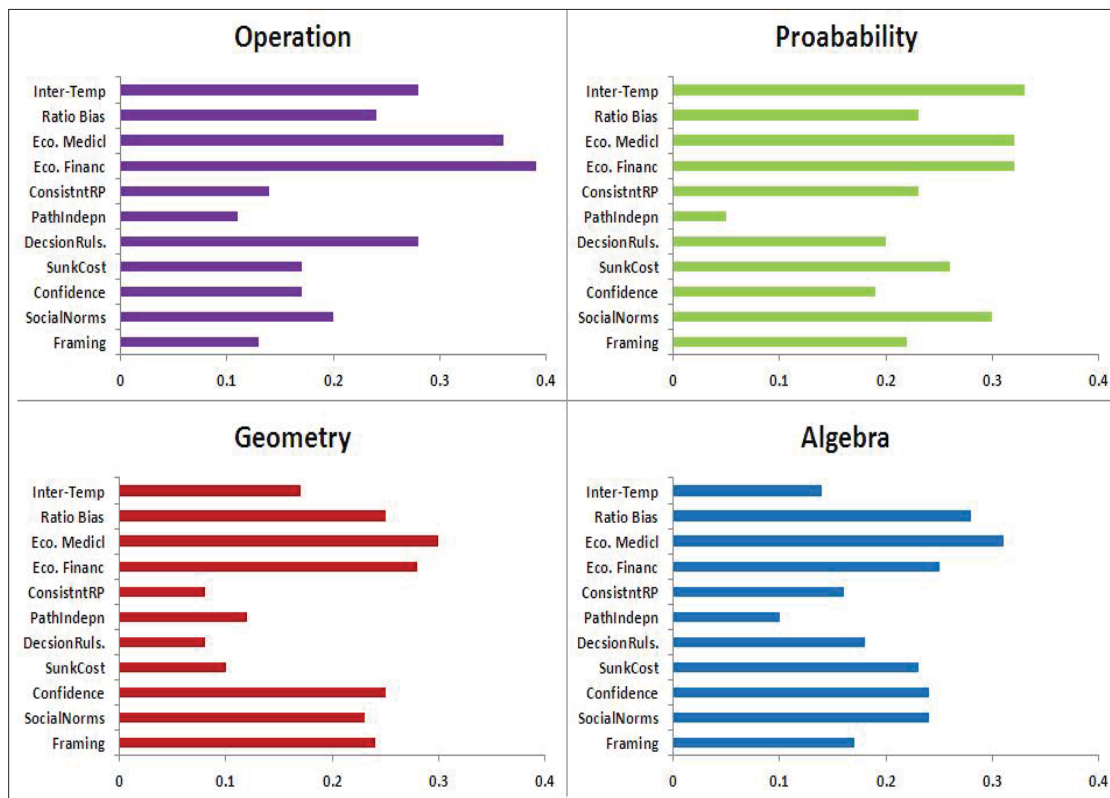


Figure M-1. Single predictor models; each component predicting all decision tasks as single predictors (length of bars showing Pearson correlations)

b) Ability Tasks

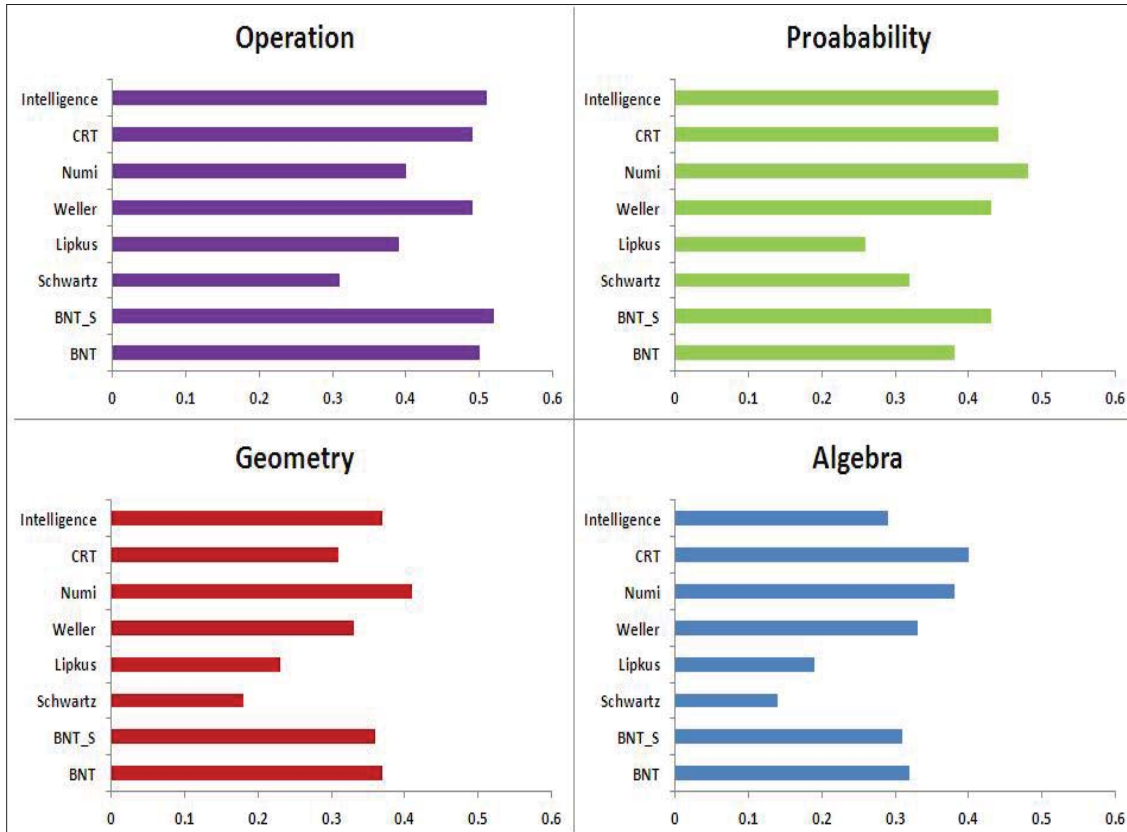


Figure M-2 Single predictor models; each component predicting all ability tasks as single predictors (length of bars showing Pearson correlations)

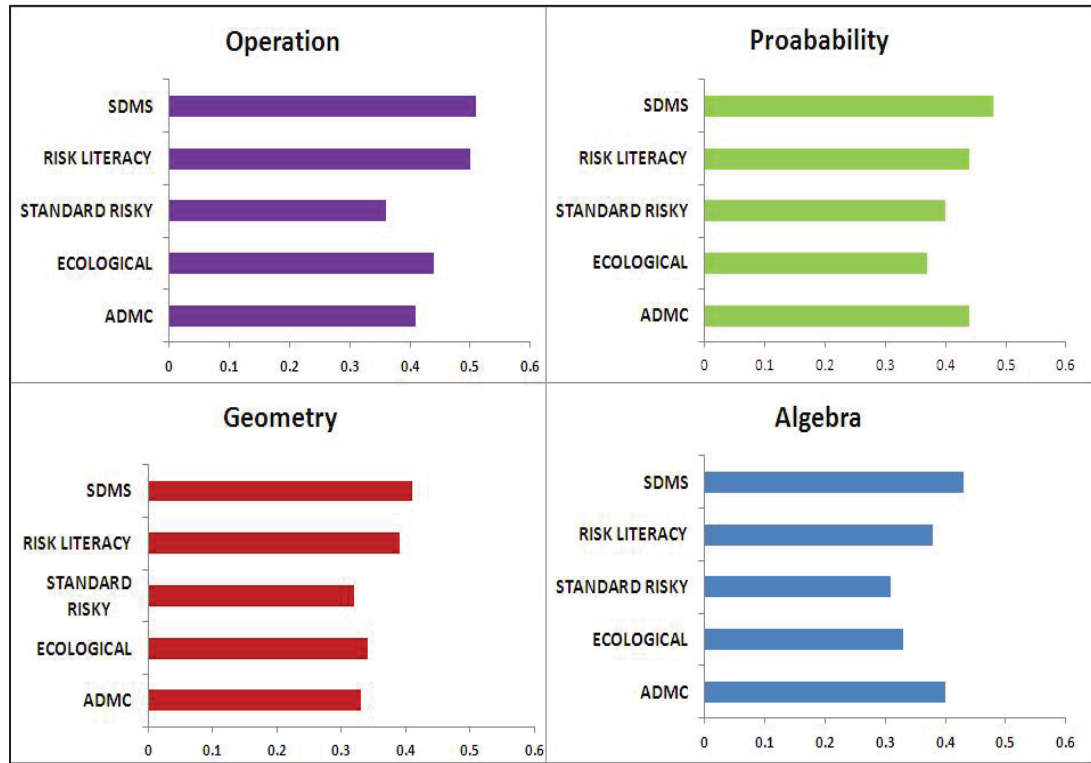
c) Overall Superior Decision Making Skill

Figure M-3 Single predictor models; each component predicting overall decision making as single predictors (length of bars showing Pearson correlations)