REAL-WORLD TASK PERFORMANCE: THE ROLE OF ABILITIES, NON-ABILITY TRAITS, AND PROXIMAL VARIABLES

A Thesis
Presented to
The Academic Faculty

by

Corey Evan Tatel

In Partial Fulfillment
of the Requirements for the Degree
Master's of Science in Psychology in the
School of Psychology

Georgia Institute of Technology
December, 2020

REAL-WORLD TASK PERFORMANCE: THE ROLE OF ABILITIES, NON-ABILITY TRAITS, AND PROXIMAL VARIABLES

Approved by:

Dr. Phillip Ackerman, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Ruth Kanfer
School of Psychology
Georgia Institute of Technology

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Date Approved: July 29, 2020

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my advisors, Dr. Phillip Ackerman and Dr. Ruth Kanfer, for their intellectual guidance on this project and for the extensive mentorship they have provided throughout my first two years of graduate school. Second, I'd like to thank Dr. Richard Catrambone for being a part of my committee and providing valuable insight.

I would also like to thank the graduate students who I have worked alongside of in the GT PARK Lab, Dr. Justin Sabree and Sibley Lyndgaard. Both have not only been fantastic colleagues, but great friends that I am lucky to share a graduate school experience with. Zack Tidler's insights have also been valuable as we share common research interests and he has allowed me to bounce ideas off of him throughout the process of conducting this study.

I want to acknowledge my team of research assistants that made data collection and analysis possible for this study. David Osayi, Maggie Cheng, Hanna Taylor, Daniel Zhou, Daniel Lewis, Sydney Rosenberg, and Tempel Dingman have all contributed by running participants, scoring video data, and conducting analyses. I am very grateful for each of their contributions and enjoyed working with this team.

Finally, I'd like to thank my parents, Jake and Leslie, and my brother, Zak, for their support as I pursue my educational goals. I look up to all three of my immediate family members and I have leaned on each of them for emotional support through my graduate school experience thus far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

K1	First keychain attempt.
K2	Second keychain attempt.
co/o	The proportion of overall time spent on the computer during the keychain task.
k/o	The proportion of overall time spent on the keychain during the keychain task.
i/o	The proportion of overall time spent looking at instructions during the chair task.
ch/o	The proportion of overall time spent working on the chair during the chair task.
r/o	The proportion of overall time spent researching during the traffic calming task.
p/o	The proportion of overall time spent preparing the PowerPoint deck during the traffic calming task.
co/k	The proportion of time spent on the computer vs. working on the keychain during the keychain task.
i/ch	The proportion of time spent looking at instructions vs. working on the chair during the chair task
r/p	The proportion of time spent researching vs. preparing the PowerPoint deck during the traffic calming task.
NS	Numer Series test.
DR	Diagramming Relations test.
SA	Spatial Analogies test.
GI	General Information test.
SSC	Spatial self-concept.
VSC	Verbal self-concept.
nAch	Need for achievement.

open	Openness to experience.
neur	Neuroticism.
extra	Extraversion.
TIE	Typical intellectual engagement.
DTL	Desire to learn.
compet	Competitiveness.
emot	Emotionality in achievement contexts.
Real	Realistic interests.
Inves	Investigative interests.
Con	Conventional interests.

SUMMARY

The objective of this study was to develop new, ecologically valid measures of adult intelligence that are representative of tasks that adults might encounter in their everyday lives. In order to accomplish this objective, 59 undergraduate students (35 male, 24 female) completed three hands-on, experimental tasks that were designed to be realistic and sample a variety of abilities. The three tasks were the construction of a paracord keychain, the assembly of an IKEA chair, and the development and presentation of a "traffic calming" solution. In order to maximize ecological validity of these tasks, participants were provided with access to the Internet so that they could utilize the full extent of resources that would have been at their disposal had they completed the task outside of a laboratory environment. While the results of the study were somewhat mixed in regard to specific hypotheses, they do highlight a gap between real-world intellectual performance and traditional ability assessment, as indicated by moderate and weak correlations between abilities and task performance. Results also suggest a moderate role of non-ability traits and proximal variables in determining task performance. Implications are discussed for the practical utility of intelligence testing and the importance of developing more realistic measures of human abilities.

CHAPTER 1. BACKGROUND AND LITERATURE REVIEW

Intelligence tests have evolved from measures that were originally designed to predict the future academic performance of children and identify children with intellectual disabilities (Binet & Simon, 1905; Terman, 1916). However, a disconnect has developed between what intelligence tests were designed for and how they have been used (Ackerman, 2017). While intelligence tests were designed to be used as *predictor* variables, it has become common for researchers and practitioners to use intelligence assessments as *criterion* variables. My review highlights the source of the historical disconnection between the purpose of intelligence assessment and practical utility that informs the need to develop more relevant measures of intellectual abilities within the context of adult populations. I proceed to review research that has investigated how people approach hands-on tasks and gather information from a variety of sources in order to solve problems. Finally, I introduce an experiment designed to identify gaps between traditional intelligence tests and intellectual performance in realistic settings and draw upon an investment model of intelligence (Ackerman, 1996) to hypothesize the influence of ability and non-ability traits on task performance.

1.1 The Criterion Problem in the Assessment of Adult Intelligence

Early intelligence tests were largely successful in predicting academic performances of children (Binet & Simon, 1905; Terman, 1916). However, as history has progressed researchers have shifted their focus to determining what predicts intelligence test performance rather than what intelligence tests predict. For example, a series of studies led investigators to conclude that enrollment in university-run nursery school generally raised intelligence quotient (IQ) scores, especially for individuals who had low IQ scores before enrollment (Wellman, 1940). While the findings were likely a statistical artifact, the result of regression-to-the-mean effects (McNemar, 1940), the design is characteristic of the criterion problem associated with intelligence testing because the researchers did not actually examine school performance. More recently, a study conducted by Brinch and Galloway (2012) suggested that compulsory schooling of Norwegian adolescents has led to increased intelligence quotient (IQ) scores among the population. Ackerman (2017) argued that the more relevant questions that studies such as these should investigate are whether university-run nursey schools or compulsory schooling of adolescents enhance future academic performance. If these interventions enhance IQ scores but do not enhance future academic performance, then the interventions are of little use given that IQ scores are designed to predict academic performance but are not themselves measures of academic performance.

Similarly, standardized tests such as the SAT or GRE are aptitude tests that are designed to predict future academic success of students who apply to undergraduate or graduate education programs, respectively. The academic system has been plagued by a similar criterion problem in which an emphasis has been placed on achieving the highest possible standardized score, as evidenced by the United States standardized test preparation industry that is predicted to reach value of $32.13 billion in 2021 (Technavio, 2017). This trend characterizes an issue similar to the of IQ testing – predictor variables are being used as criterion variables. Researchers as well as practitioners are interested in investigating what leads to higher scores rather than what higher scores predict.

1.2 The Need to Develop More Relevant Measures of Adult Intelligence

By focusing on what might predict higher scores on intelligence and ability tests, researchers and practitioners have lost sight of the most important question surrounding intelligence and ability testing; namely what behaviors or outcomes do intelligence measures predict in the real world? Ackerman (2017) called for future research to focus on developing sufficient criterion measures of adult intelligence that are representative of intellectually demanding tasks that adults are likely to encounter in their daily lives, whether in work or non-work domains. Initial empirical evidence of the need for such measures was illustrated by the findings of a study conducted over 60 years ago. Demming and Pressey (1957) grew dissatisfied with the contradiction that ability test scores decline with age, yet experience typically enhances performance in many domains. As a result, they designed tasks thought to be "indigenous" to adults and examined their relationships with several traditional tests of intellectual ability. They found that while younger adults outperformed older adults on the traditional measures of intelligence (e.g., the Army Beta test, Otis tests, Minnesota Paper Form Board), older adults outperformed younger adults on tasks that were more relevant to daily life (e.g., locating information in a phone book, understanding common legal terms, knowing which occupations to turn to for particular services) (Demming & Pressey, 1957).

While Demming and Pressey's work shed light on the need to evaluate the influence of intelligence on tasks "indigenous" to adults, the tasks that they investigated are no longer encountered by the typical adult. With advances in technology, a large majority of the adult population no longer needs to locate information in a phone book, understand complex legal terms, or memorize which occupations to turn to when services are needed. Such information is now easily accessible through the Internet. This study builds upon Demming and Pressey's (1957) work by creating tasks that are designed to be "indigenous" of modern-day adults and answers the call to assess the relationship between intelligence and how adults approach and solve cognitively-demanding tasks in their daily lives (Ackerman, 2017).

1.3 Real-World Task Performance

The initial goal of this research was to develop a sample of tasks that are representative of the types of intellectually demanding activities that adults are likely to confront in the context of modern life (i.e., outside contexts such as school or the laboratory). While past research on the relationship between intellectual abilities and hands-on, real-world tasks is sparse, there are two areas in psychological literature that are relevant to the context of the current study. In the following two sub-sections, I review the literature on research that has investigated assembly tasks and the ability of adults to assess and integrate information from various sources.

1.3.1 Research on Assembly Tasks

Assembly tasks are one example of an ecologically valid, intellectually demanding task. Many adults encounter the need to assemble objects from multiple component parts at some points their lives. Prominent examples include assembling a piece of furniture for a living room or constructing a swing set for a yard. The strategies that adults use to complete such tasks and the quality of the final product is likely a function of individual differences in knowledge, ability, and non-ability traits. Researchers have used assembly tasks in experimental contexts to gain a better understanding of what factors determine whether or not people can efficiently and accurately construct a target object. This research has typically focused on how different types of instructions influence task assembly performance (Daniel & Tversky, 2012; Fiorella, von Gog, & Hoogerheid, 2017; Morrell & Park, 1993; Novick & Morse, 2000; Wiking et al., 2016), how assembly task characteristics influence the difficulty of the task (Richardson, Jones, & Torrence, 2004; Richardson, Jones, Torrence, & Baugley, 2004), and how intellectual abilities influence task performance (Daniel & Tversky, 2012; Morell & Park, 1993).

The majority of empirical research utilizing assembly tasks has focused on how participants use and interpret different types of instructions. Findings have indicated that people find both text-based instructions and diagram-based instructions useful (Daniel & Tversky, 2012) and make fewer errors when they are provided with both types of instructions rather than one or the other (Morrell & Park, 1993; $r = 0.27$). Additionally, people are less likely to make errors when they are provided with step-by-step instructions rather than diagrams of completed objects (Wiking et al., 2016, $d = 0.61$). Recently, Fiorella and colleagues (2017) extended the study of instructions to assess how characteristics of instructional videos (e.g., the perspective from which they were filmed), rather than text or diagrams, impact individuals' abilities to complete assembly tasks.

Another stream of research has investigated how particular assembly task characteristics influence the task's perceived difficulty level. Richardson and colleagues (2004, 2006) used hierarchical task analysis to identify a series of task characteristics that determine the difficulty of an assembly task (e.g., number of fastening points, number of novel assemblies) and linked the majority of the task characteristics to how participants read instructions ($\beta_{\text{novel assemblies}} = 0.32$, $\beta_{\text{fastenings}} = 0.15$, $\beta_{\text{component groups}} = 0.12$, $\beta_{\text{fastening components}} = -0.087$) and how long participants spent thinking before attempting assembly tasks ($\beta_{fastening\ components}$ = *0.34,* $\beta_{novel\ assemblies}$ = *0.30,* $\beta_{selections}$ = *0.25,* $\beta_{symmetrical\ planes}$ = *-0.28).*

Three of the studies previously mentioned are particularly relevant to the current study because they included investigations of the influence of individual differences in cognitive ability on assembly task performance. Morell and Park (1993) found that age-related differences in assembly errors were attributed to differences in working memory ability ($r_{spatial\ working\ memory,\ errors}$ = -0.32, $r_{verbal\ working\ memory,\ errors}$ = -0.28) and text comprehension (average r = -0.22). Two other studies indicated the importance of spatial ability in assembly task performance. The Wiking et al. (2016) study was designed to investigate IKEA's claim that women are more effective in assembling their furniture than men because women are more likely to read the instructions. The findings refuted the claim by showing that men assembled an IKEA kitchen trolley faster (r = 0.34) and more accurately (r = 0.33) than women. The crucial component of the experiment was that gender differences in assembly time and accuracy were accounted for by scores on the Mental Rotation Task (MRT; Peters et al., 1995), a traditional assessment of spatial ability (Wiking et al., 2016; r = 0.48). Similarly, Daniel and Tversky (2012) found that individuals higher in spatial ability assembled a television stand faster than individuals lower in spatial ability (r = 0.56).

1.3.1.1 Gaps

While past research has provided evidence for the influence of instruction characteristics, task characteristics, and cognitive ability on assembly task time and quality of completed objects, there are two significant gaps in the literature that pertain to this study. First, many of the tasks that have been studied in experimental studies are somewhat artificial. Assembling models of electrical circuits (Fiorella et al., 2017), origami objects (Novick & Morse, 2000), or three-dimensional objects constructed from Lego blocks and pieces of foam board (Morell & Park, 1993) are not tasks that are representative of what adults are likely to encounter in their everyday lives.

Second, the studies that did use realistic tasks did not allow participants access to resources that would be available to them if they were completing the task outside of a laboratory environment. For example, studies have investigated how people assemble furniture using instructions that were provided to them (Daniel & Tversky, 2012; Wiking et al., 2016). However, adults are no longer limited to the instructions that furniture companies provide. They have access to additional resources on the Internet including demonstration videos, forums, and diagrams that may aid in assembly. This study aimed to address these gaps by investigating the influence of individual differences in abilities, personality, and self-concept on speed and accuracy of assembly tasks that adults are likely to encounter with access to resources that would be at their disposal outside of a laboratory environment.

1.3.2 Research on Assessing and Integrating Information

Given the vast array of resources that adults can use to accomplish intellectually demanding tasks, it is important to understand how intellectual abilities are related to the efficacy of navigating these resources. A philosophical theory called the "extended cognition hypothesis" proposes that cognition is not limited only to humans' minds, but can be extended to their environment and the tools that they use, such as technology (Rowlands & Mark, 1999). While this theory lacks empirical evidence and has been challenged (Rupert, 2004), it highlights the importance of investigating the relationship between intellectual abilities and the use of technological resources. In this sub-section, I review past research that has studied the ability to comprehend and integrate information. I begin with a review of investigations of how people integrate information from sources provided to them before I proceed to discuss investigations of Internet search abilities.

1.3.2.1 Comprehension/Integration of Provided Resources

Different types of intelligence have been linked to a person's ability to understand information that is presented to him or her. Intelligence has traditionally been divided into process-related intelligence (i.e., fluid intelligence) and content-related intelligence (i.e., crystalized intelligence) (Cattell, 1963). Both types of intelligence have been linked to how humans understand information that is presented to them. Meta-analytic findings suggest that two indicators of fluid intelligence, attention and verbal working memory, have the strongest correlations with reading comprehension levels (Carretti, Borella, Cornoldi, & De Beni, 2009; $r_{\text{verbal WM, comprehension}} = 0.41$, $r_{\text{WM updating measure, Comprehension}} = 0.37$). Domain knowledge, a specific component of crystalized intelligence, has been linked to both reading and viewing comprehension of variety of topics (Schroeders et al., 2013; $\beta = 0.78$).

In order to navigate an array of resources for the purposes of developing an approach or solution to a task, it is necessary to integrate information from multiple sources. A descriptive model called the Multiple Document Task-Based Relevance Assessment and Content Extraction Model (MD-TRACE Model) identified five core processes that individuals use during functional reading: the construction of a task model, the assessment of one's information needs, the selection/processing/integration of information, the construction of a task product, and the assessment of product quality (Rouet & Britt, 2011).

1.3.2.2 Internet Search Abilities

While there has been extensive research on how people evaluate and understand information from sources that are provided to them, there is significantly less research on how individuals evaluate and understand information as they search the Internet. A descriptive model has been proposed that describes the skills that people use to solve problems with information that they find on the Internet (Brand-Gruwel, Wopereis, & Walraven, 2009). Through three studies involving think-aloud methodologies, Brand-Gruwel and colleagues identified the following constituent skills as essential in the Internet information problem solving process: defining the information problem, searching for information, scanning information, processing information, and organizing/presenting information. The authors also proposed that regulatory (orientation, monitoring, steering, and evaluating) and conditional skills (reading skills, evaluating skills, and computer skills) can offset a lack of prior knowledge throughout the process.

Past research that is most analogous to the purposes of this study investigated the relationships between fluid intelligence, crystalized intelligence, and digital literacy (Moehring, Schroeders, Leichtmann, & Wilhelm, 2016). Prior to this study, assessments of the capability to solve problems using the Internet took place in artificial contexts. For example, the Educational Testing Service iSkills Assessment used simulation software to assess cognitive problem-solving skills and critical thinking skills while using the Internet (Katz, 2007). However, Moehring and colleagues conducted two experiments that used a more ecologically valid context. Participants in the studies were given a series of medical and health-related questions and given a time limit to search the Internet for the information required to answer the question. Items ranged from asking participants to identify symptoms of a disease to asking participants to identify the name of a surgical intervention depicted by an X-ray image. Results from the first experiment indicated that fluid intelligence, and to a lesser extent crystalized intelligence, played a significant role in determining task performance ($r_{\text{fluid}} = 0.78$, $r_{\text{crystalized}} = 0.46$) and that crystallized and fluid intelligence combined to account for 83% of the variance in performance. When domain-specific health knowledge was included in the second experiment, it was the strongest predictor of performance ($\beta = 0.81$).

1.3.2.3 Gaps

There has been significantly more research on how people understand and assess sources of provided information than on how effectively people search for relevant information on the Internet. Additionally, the majority of investigations of search abilities have taken place in an environment that is not ecologically valid. Only one study has assessed the relationship between individual differences in knowledge, ability, and performance on an information-seeking problem using the Internet (Moehring et al., 2016). While this investigation sheds light on the relationships between intellectual processes, prior knowledge, and digital literacy, the tasks assessed are not likely to be encountered by most adults. Medical professionals might be tasked with identifying symptoms of a disease or identifying a surgical procedure depicted on an X-ray image, but that is only a small portion of the adult population.

Additionally, other individual differences besides fluid and crystalized intelligence likely influence a person's performance on problem solving tasks using the Internet. Ackerman (1996) proposed a theory that delineated the critical role of non-ability traits in adult intellectual development. The theory contained four components: intelligence-as-process, personality, interests, and intelligence-as-knowledge (PPIK) (Ackerman, 1996). Intelligence-as-process and intelligence-as-knowledge are related to the traditional divisions of intelligence into fluid and crystalized categories (Cattell, 1963). However, the PPIK framework recognizes that non-ability traits such as personality and interests influence the subjects and activities that people invest cognitive effort in, which then determine domain-specific knowledge (i.e., intelligence-as-knowledge). Ackerman and Heggestad (1997) provided empirical evidence in support of the PPIK framework by demonstrating that non-ability traits are in fact correlated with intellectual abilities and by identifying trait complexes that include abilities, interests, and personality traits that are typically related. Investigating such non-ability traits might lead to a more holistic understanding of peoples' ability to solve problems using different resources.

The current study addresses these gaps by: 1) investigating the influence of individual differences in knowledge and ability on performance on tasks that adults are likely to encounter; and 2) by including non-ability traits in the study's design.

1.4 The Current Study

The current study was designed to develop a sample of tasks that are representative of the types of cognitively demanding tasks that adults are likely to confront. Specifically, the study used three experimental tasks: 1) construction of a paracord keychain; 2) assembly of an IKEA "Ingolf" chair; and 3) development/presentation of a solution to a traffic calming problem. The ultimate goal of developing these tasks is to predict performance on other real-world, cognitively demanding activities by maximizing the ecological validity of the tasks studied in a laboratory environment. Initially, the current study investigates whether performance on these tasks provides information that has properties unique in comparison to what traditional measures of knowledge and ability typically account for.

This study aimed to address two important gaps in current research. First, many of the reviewed studies involve tasks that adults are relatively unlikely to encounter (e.g., assembly an electrical circuit model or objects made of Legos and foam board). The tasks in the current study were chosen because they are representative of a variety of tasks that adults could encounter outside of a laboratory environment. Many adults will likely face the challenge of constructing or assembling an object from various components. The IKEA chair and paracord keychain were designed to represent such situations. It is also likely the adults will encounter situations in which they must solve complex problems that require them to gather information needed to develop a solution. The traffic calming problem was designed to be an example of such a problem.

Second, although some tasks outlined in the studies above were representative of tasks that are relevant to adults, they were not carried out in contexts that are representative of how adults would approach such tasks in the 21^{st} century. Several studies have investigated performance on relevant everyday intellectual tasks, but they typically limited individuals to constrained resources to support task solution. This study allowed access to a range of resources that participants would have access to in the real world including the Internet, demonstration videos, instructions, and diagrams. Additionally, participants completed the chair assembly task in pairs because adults often enlist help from friends or family when completing this task at home. By assessing participants' ability to complete these tasks in a context that is more representative of how they would approach the tasks outside of a laboratory environment, I attempted to identify gaps between what traditional intelligence tests measure and performance on real-world intellectual tasks.

1.5 Hypotheses

This section contains a series of hypotheses for relationships between task performance and traditional measures of intelligence, personality, interests, self-concept, motivational traits, prior experience, and task self-efficacy, as well as relationships between performance on the individual tasks themselves. The rationale for the strength and direction of each hypothesis is described.

1.5.1 Traditional Measures of Ability

Traditional measures fluid intelligence and crystalized intelligence are likely to be correlated with performance and speed of completion on each of the three tasks. Fluid and crystalized intelligence have been previously linked to assembly task performance (Morell & Park, 1993), the ability to process information from a multiple sources (Braten et al., 2011; Rouet et al., 1997), and the ability to solve problems using the Internet (Moehring et al., 2016). Accordingly, it was predicted that scores on traditional measures of fluid intelligence and crystalized intelligence would be correlated with real-world task performance and timing.

The utility of this approach to the assessment of adult intelligence was determined by the strength of the correlations between task performance and traditional measures. Moderate correlations between task performance and traditional measures indicate that the tasks provide additional information that is not accounted for by traditional measures, while large correlations indicate that they are redundant to traditional measurements and small correlations indicate that the tasks may not be measuring intellectual abilities. Therefore, I anticipated moderate correlations between traditional ability measures and task criteria.

However, I did not expect correlations between crystalized intelligence and task performance to be as large as correlations between fluid intelligence and task performance. Fluid intelligence typically has a more powerful influence on performance in a restricted environment such as a laboratory or a classroom, while crystalized intelligence has a more powerful influence on performance in a less restricted environment (e.g., in one's free time) (Beier & Ackerman, 2005). While this study was designed to maximize ecological validity and minimize the restriction of the environment, participants were still aware that they were being observed and timed during each task they attempted. It was expected that fluid intelligence would be more strongly correlated with task performance than crystalized intelligence due to the restricted nature of a laboratory environment. According to Cohen (1988), a correlation of between 0.10 and 0.29 is considered small, between 0.30 and 0.49 is considered moderate, and 0.50 or greater is considered large. Based on these criteria, I hypothesized a correlation of approximately 0.30 between crystalized intelligence and performance and correlations ranging from 0.35 to 0.49 between fluid intelligence and task performance.

Additionally, I expected that traditional measures of fluid and crystalized intelligence would be related to the amount of time spent on each task. I predicted that individuals or pairs of individuals with higher levels of abilities and knowledge would complete the paracord keychain construction task and the chair assembly task faster than individuals or pairs of individuals with lower levels of abilities and knowledge. These predictions are justified based on the various theories and investigations that suggest speed is an important component of intellectual performance (Beck, 1933; Thorndike, 1924). However, the relationships between ability, knowledge, and time spent on the traffic calming task are likely more complex. While I did not expect to see overall differences in completion time, I did expect to see differences in how people spend their time during the task. For example, individuals with higher levels of ability or knowledge might spend more time evaluating resources and developing an ideal solution while individuals with lower levels of ability or knowledge might spend more time preparing their PowerPoint presentations. All hypotheses associated with time allocation, including ability and non-ability traits, are exploratory in nature given that there is not sufficient theory to confidently predict the direction of such correlations. All specific hypothesis associated with abilities traits are depicted **Table 1.**

Table 1 – Hypothesized relationships between abilities and task outcomes.

	Keychain Quality	Keychain Completion Speed	Chair Completion Speed	Traffic Calming Presentation
Fluid intelligence	*Hypothesis 1a* r = 0.35 to 0.39	*Hypothesis 1b* r = 0.35 to 0.39	*Hypothesis 1c* r = 0.35 to 0.39	*Hypothesis 1d* r = 0.35 to 0.39
Crystallized intelligence	*Hypothesis 1e* r = 0.30	*Hypothesis 1f* r = 0.30	*Hypothesis 1g* r = 0.30	*Hypothesis 1h* r = 0.30

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.5.2 Non-Ability Traits

In this section, I outline a series of hypotheses predicting relationships between non-ability traits and performance on each of the three tasks. I include relationships between traits delineated in the PPIK framework (interests and personality), two other non-ability traits (self-concept and motivational traits), and task performance and timing. While I predicted moderate correlations ranging from 0.35-0.49 between traditional measures of fluid intelligence and task performance, I did not expect correlations between non-ability traits and task performance and timing to be as large. The PPIK framework specifies that non-ability traits influence adult intelligence because they determine what people *choose* to invest their cognitive resources in and consequently, the knowledge that they acquire. Participants in this study were not given a choice as to which task they preferred to attempt; they were instructed to complete all three tasks. Similar to the hypothesized associations between crystalized intelligence and task performance, I posited the correlations between non-ability traits and task performance to be approximately 0.30; a medium-sized association.

1.5.2.1 <u>Interests.</u>

The general consensus framework for assessing vocational interests is Holland's model (1963). Holland identified six interest themes: Realistic (individuals who typically prefer motor and physical activities), Investigative (individuals who typically prefer intellectually engaging activities), Artistic (individuals who "prefer indirect relations with others"), Social (individuals who typically prefer activities that involve helping or interacting with others), Enterprising (individuals who prefer activities that allow them to utilize persuasive or sales abilities), and Conventional (individuals who typically prefer highly structured activities) (Holland 1963).

Ackerman & Heggestad (1997) delineated relationships between interests and measures of intellectual ability. Particularly relevant to the current study, verbal ability is associated with Investigative and Artistic interests with effect sizes ranging from $r = 0.20$ to 0.37 (Ackerman, Kanfer, & Goff, 1995; Kanfer, Ackerman, & Heggestad, 1996; Randahl, 1991; Rolfus & Ackerman, 1996), spatial ability is associated with Realistic interests with effect sizes ranging from $r = 0.24$ to 0.34 (Ackerman et al., 1995; Kanfer, et al., 1996; Randahl, 1991; Rolfus & Ackerman, 1996), and perceptual speed is associated with conventional interests with effect sizes ranging from $r = 0.15$ to 0.16 (Kanfer, et al., 1996; Randahl, 1991; Rolfus & Ackerman, 1996). In the context of this study, the keychain construction task and chair assembly task were expected to require spatial ability and perceptual speed while the traffic calming task was expected to require verbal ability. Based on empirically established relationships between abilities and interests and the expected relationships between abilities and task performance, I hypothesized that Realistic and Conventional interests would be associated with performance and timing on the keychain construction and chair assembly tasks while Investigative and Artistic interests would be associated with performance and timing on the traffic calming task. Hypothesized relationships between vocational interest themes and task outcomes are depicted in **Table 2.**

Table 2 – Hypothesized relationships between vocational interests and task outcomes.

	Keychain Quality	Keychain Completion Speed	Chair Completion Speed	Traffic Calming Presentation
Realistic interests	*Hypothesis 2a* r = 0.30	*Hypothesis 2b* r = 0.30	*Hypothesis 2c* r = 0.30	
Conventional interests	*Hypothesis 2d* r = 0.30	*Hypothesis 2e* r = 0.30	*Hypothesis 2f* r = 0.30	
Investigative interests				*Hypothesis 2g* *r = 0.30*
Artistic interests				*Hypothesis 2h* *r = 0.30*

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.5.2.2 Personality

Personality traits also fall under the non-ability "umbrella" and are likely to influence outcomes related to the tasks in the current study. In their meta-analysis, Ackerman and Heggestad (1997) found 52% of the personality-ability correlations investigated to be significantly different from zero. The correlations found between two trait complexes and intellectual abilities are particularly noteworthy for the purposes of this study. On one hand, openness to experience and typical intellectual engagement (TIE), which is one's "desire to engage and understand their world, their interest in a wide variety of things, and their preference for a complete understanding of a complex topic or problem" (Goff & Ackerman, 1992), are positively associated with intellectual abilities. On the other hand, traits broadly related to neuroticism and psychoticism are negatively associated with intellectual abilities.

Additionally, Ackerman and Heggestad (1997) outlined associations between interests and personality. Of particular interest to the current study, conscientiousness is positively associated with Conventional interests (Gottfredson, Jones, & Holland, 1993, $r = 0.25$) and openness/TIE is positively associated with Investigative and Artistic interests (Kanfer et al., 1996, $r_{openness/TIE,\ Investigative} = 0.42$, $r_{openness/TIE,\ Artistic} = 0.35$). Based on the established relationships between personality, interests, and intelligence, I hypothesized that performance and timing on the keychain construction and chair assembly tasks would be associated with conscientiousness/need for achievement, performance and timing on the traffic calming task would be associated with openness/TIE, and performance and timing on all three tasks would be associated with neuroticism.

Considering the context in which the chair assembly task is completed, extraversion was posited to play a role in task outcomes. Extraversion has been linked to performance in team settings (Morgeson, Reider, & Campion, 2005, $r = 0.21$), likely because teams high in extraversion communicate more effectively in order to achieve the desired goal. Accordingly, I predicted that extraversion would be associated with performance and timing in the chair assembly task because participants completed the task in pairs. Hypothesized relationships between personality traits and task outcomes are depicted in **Table 3**.

Table 3 – Hypothesized relationships between personality traits and task outcomes.

	Keychain Quality	Keychain Speed	Chair Completion Speed	Traffic Calming Presentation
Conscientiousness	*Hypothesis 3a* $r = 0.30$	*Hypothesis 3b* $r = 0.30$	*Hypothesis 3c* $r = 0.30$	
nAch	*Hypothesis 3d* $r = 0.30$	*Hypothesis 3e* $r = 0.30$	*Hypothesis 3f* $r = 0.30$	
Openness				*Hypothesis 4a* $r = 0.30$
TIE				*Hypothesis 4b* $r = 0.30$
Neuroticism	*Hypothesis 5a* $r = -0.30$	*Hypothesis 5b* $r = -0.30$	*Hypothesis 5c* $r = -0.30$	*Hypothesis 5d* $r = -0.30$
Extraversion			*Hypothesis 6* $r = 0.30$	

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.5.2.3 Self-Concept

Another non-ability trait that should be related to task performance is self-concept. Self-concept is a person's beliefs about their general abilities in specific domains (Shavelson, Hubner, & Stanton, 1976). It is related to, yet theoretically distinct from other types of self-estimates such as self-esteem and self-efficacy (Pajares & Miller, 1994; Valentine, DuBois, & Cooper, 2004). Self-esteem is the most general of the three and consists of a person's perception of his or her abilities across different domains (Rosenberg, 1965), while self-efficacy is the most specific of the three and consists of a person's perception of his or her ability to accomplish a specific task or achieve a specific outcome (Bandura, 1977). In terms of scale, self-concept falls between self-esteem and self-efficacy (Ellingsen, 2013). It is useful to analyze self-estimates at the level of self-concept given that different self-concept domains, such as verbal and mathematical, are often polarized (Ackerman & Wolman, 2007).

Past research has investigated the relationships between self-concept, achievement, abilities, and personality. Kornilova, Kornilov, and Chumakova (2009) have established a large association between academic self-concept and academic achievement ($r = 0.60$). In terms of associations with ability, results have been somewhat mixed. Effect sizes for the relationship between self-estimates of ability and objective ability range from correlations of 0.08 (Verbal Test of Spatial Ability) to 0.45 (Vocabulary) (Ackerman & Wolman, 2007). However, Ackerman and Wolman's (2007) study indicated that self-estimates of ability are somewhat reliable in particular domains. Additionally, self-concept has been linked to a variety of personality traits. For example, openness is correlated with verbal self-concept ($r = 0.49$) and conscientiousness is correlated with both math self-concept ($r = 0.26$) and problem-solving self-concept (r = 0.28) (Marsh, Trautwenin, Ludtke, & Koller, 2006). Additionally, TIE is associated with both verbal (Ackerman, et al., 1995, r = 0.35) and academic self-concept (Kanfer, et al., 1996, r = 0.18). Ackerman and Wolman (2007) also found higher correlations of trait complexes with measures of verbal self-concept than with other self-concept measures.

Based on established links between self-concept, ability, and personality, as well as abilities associated with each task, I hypothesized that performance and timing on the keychain construction task and chair assembly task would be associated with spatial self-concept while performance on the traffic calming task would be associated with verbal self-concept. Specific hypotheses associated with self-concept are depicted in **Table 4.**

Table 4 – Hypothesized relationships between self-concept and task outcomes.

	Keychain Quality	Keychain Completion Speed	Chair Completion Speed	Traffic Calming Presentation
Spatial self-concept	*Hypothesis 7a* r = 0.30	*Hypothesis 7b* r = 0.30	*Hypothesis 7c* r = 0.30	
Verbal Self-concept				*Hypothesis 7d* r = 0.30

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

.

1.5.2.4 Motivational Traits

The last non-ability traits that were considered in the context of this study were motivational traits; given that what drives individuals to direct and sustain effort is likely to influence their performance on the various tasks. Kanfer and Heggestad (1997) integrated research on personality and motivation to distinguish between motivational traits and motivational skills. Motivational traits, they argued, are stable and likely to be a function both personality and motivation. Specifically, Kanfer and colleagues integrated personality taxonomy measures, achievement motivation measures, and goal orientation measures to create the Motivational Trait Questionnaire (MTQ) (Heggestad & Kanfer, 2000; Kanfer & Ackerman, 2000; Kanfer & Heggestad, 1997). The MTQ consists of 3 scales each containing two sub-scales: approach-oriented motivation (desire to learn, mastery), competitive excellence (other-referenced goals, competitiveness), and aversion-related motivational traits (worry, emotionality). I expected each of these scales to be associated with outcomes on all three tasks. Approach-oriented motivation and competitive excellence were hypothesized to positively influence task outcomes, while aversion-related motivation was hypothesized to negatively influence task outcomes. Specific hypotheses associated with each sub-scale of the MTQ are presented in **Table 5.**

Table 5 – Hypothesized relationships between MTQ sub-scales and task outcomes.

	Keychain Quality	Keychain Completion Speed	Chair Completion Speed	Traffic Calming Presentation
Desire to learn	*Hypothesis 8a* *r* = 0.30	*Hypothesis 8b* *r* = 0.30	*Hypothesis 8c* *r* = 0.30	*Hypothesis 8d* *r* = 0.30
Mastery	*Hypothesis 8e* *r* = 0.30	*Hypothesis 8f* *r* = 0.30	*Hypothesis 8g* *r* = 0.30	*Hypothesis 8h* *r* = 0.30
Other-oriented goals	*Hypothesis 9a* *r* = 0.30	*Hypothesis 9b* *r* = 0.30	*Hypothesis 9c* *r* = 0.30	*Hypothesis 9d* *r* = 0.30
Competitiveness	*Hypothesis 9e* *r* = 0.30	*Hypothesis 9f* *r* = 0.30	*Hypothesis 9g* *r* = 0.30	*Hypothesis 9h* *r* = 0.30
Worry in achievement contexts	*Hypothesis 10a* *r* = -0.30	*Hypothesis 10b* *r* = -0.30	*Hypothesis 10c* *r* = -0.30	*Hypothesis 10d* *r* = -0.30
Emotionality in achievement contexts	*Hypothesis 10e* *r* = -0.30	*Hypothesis 10f* *r* = -0.30	*Hypothesis 10g* *r* = -0.30	*Hypothesis 10h* *r* = -0.30

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.5.3 Proximal Predictors

There is a theoretical distinction in the skill acquisition literature between distal predictors and proximal predictors (Ackerman et al., 1995). On one hand, distal predictors are general in nature and often refer to traits that drive skill acquisition. Both ability and non-ability traits are considered distal predictors. On the other hand, proximal predictors are more context-specific and are typically directly associated with a particular task. Distal factors influence behavior through proximal factors (Kanfer & Ackerman, 1989) and in some cases, proximal predictors provide incremental validity in prediction of task performance over distal traits (Ackerman et al., 1995). Two proximal predictors were examined in this study: prior experience and task self-efficacy. The specific hypotheses associated with each proximal predictor is depicted in **Table 6.**

1.5.3.1 Prior experience

In addition to ability and non-ability traits, prior experience was posited to play a role in task-related outcomes. Individuals with domain-specific knowledge, which is developed primarily through experience (Ackerman, 2000), exhibit higher levels of digital literacy (Moehring et al., 2016). Declarative knowledge and procedural skills are essential components of how adults complete intellectual tasks in the real world (Ackerman, 2000). In the context of this study, it is likely that experience with tasks related to those which have been chosen lead to more successful outcomes. For example, someone who has experience assembling furniture is likely more skilled at the chair assembly task and a person with knowledge of traffic calming strategies is likely better equipped to develop a high-quality solution for the traffic calming task. Given that domain knowledge is a component of crystalized intelligence and crystalized intelligence typically has smaller correlations with performance in laboratory environments than fluid intelligence does (Beier & Ackerman, 2005), I predicted a moderate correlation of approximately 0.30 between prior experience and performance on all three tasks.

1.5.3.2 Task Self-Efficacy

Self-efficacy refers to a person’s perception of his or her capability in accomplishing a specific achievement goal (Kanfer, 2012). People develop self-efficacy beliefs from mastery experiences, vicarious experiences, social persuasion, and physiological arousal (Bandura, 1977). Self-efficacy has been linked to performance on a variety of outcomes, including but not limited to academic performance (Schunk, 1989), computer programming task performance (Wiedenbeck, 2005), and air traffic controller tasks (Kanfer & Ackerman, 1989). Self-efficacy is typically assessed at the level of a specific task or outcome (Zimmerman, 2000) and is therefore likely to predict performance on each task in the current study. Given the that task self-efficacy estimates are directly related to the specific task context, I anticipated that task self-efficacy would be strongly correlated with task timing and task performance (r = 0.50 - 0.60).

Table 6 – Hypothesized relationships between proximal predictors and task outcomes.

	Keychain Quality	Keychain Completion Speed	Chair Completion Speed	Traffic Calming Presentation
Prior knowledge/experience	*Hypothesis 11a* *r* = 0.30	*Hypothesis 11b* *r* = 0.30	*Hypothesis 11c* *r* = 0.30	*Hypothesis 11d* *r* = 0.30
Task self-efficacy	*Hypothesis 12a* *r* = 0.50-0.60	*Hypothesis 12b* *r* = 0.50-0.60	*Hypothesis 12c* *r* = 0.50-0.60	*Hypothesis 12d* *r* = 0.50-0.60

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.5.4 Correlations between task performance

There is a common debate among differential psychologists as to whether broad or narrow measurements of individual differences in both ability and non-ability domains are superior. This debate is related to the perspective of Brunswik symmetry, which can be traced back to Brunswik's lens model (Brunswik, 1952) and has been advanced significantly by Wittman (Wittman, 1991; Wittman, 1995; Wittman & Sub, 1999). Wittman grounds the foundation of Brunswik symmetry in the distinction between experimental and correlational psychology, which was delineated by Cronbach (1957). According to Cronbach, experimental psychologists are concerned with explaining behavior by controlling situational variables while correlational psychologists are concerned with using individual differences to predict criterion variables of interest. Brunswik symmetry is of particular interest to correlational psychologists.

Psychological theories and measures are often represented by hierarchical models. Common examples include Sternberg's Triarchic theory of intelligence (Sternberg, 1988) and the Five Factor Model of personality (Costa & McCrae, 1988). The central idea of Brunswik symmetry is that in order to maximize the predictive power of dispositional differences, the predictor and criterion variables should be measured at the same level of aggregation (Wittman & Sub, 1999). That is, broad traits predict broad criterion while narrow traits predict narrow criterion. For example. a broad personality trait such as conscientiousness may not be an effective predictor of any single behavioral episode such as determining whether or not an employee will succeed in making a single sale. It would be a more accurate predictor of a broader criterion that aggregates behavior across many episodes, such as job performance over the course of a year.

In order to predict the relationships between performance on each of the three tasks in this study, it is crucial to consider the hierarchical structure of human abilities and the concept of Brunswik symmetry. An extensive review of abilities has revealed seven categories of abilities (each with component abilities at lower levels) that are nested under a general intelligence factor. Those seven abilities are fluid intelligence, crystallized intelligence, visual perception, perceptual speed, learning and memory, knowledge and achievement, and ideational fluency (Carroll, 1993). In the context of this study, the keychain and chair tasks likely require spatial abilities (i.e., visual perception and perceptual speed) while the traffic calming task likely requires verbal abilities (i.e., ideational fluency)

Given the hierarchical nature of abilities based on this model, correlations between performance on these three tasks that likely require different abilities should provide insight into whether broad or narrow predictors are beneficial for determining real-world intelligence. If correlations between the three tasks are relatively large, that would suggest the need for broad predictors. However, smaller correlations would suggest specialization that requires more narrow predictors. Based on the general factor in Carroll's theory, as well as other theories of intelligence (e.g., Spearman, 1927), I expected correlations of traffic calming task performance with keychain and chair task performance to be between 0.20 and 0.34. However, I also expected the magnitude of correlations between task performance to display specialization. I predicted that outcomes on the keychain construction and chair assembly task would be moderately correlated, but at a magnitude larger than correlations between outcomes of either task with traffic calming task outcomes

(r = 0.35-0.39). Specific hypotheses associated with correlations between task outcomes are presented in **Table 7.**

Table 7 – Hypothesized relationships between task outcomes

	Keychain Quality	Keychain Completion Speed	Chair completion Speed	Traffic Calming Presentation
Keychain Quality				
Keychain Completion Speed				
Chair Completion Speed	*Hypothesis 13a* r = 0.35 to 0.49	*Hypothesis 13b* r = 0.35 to 0.49		
Traffic Calming Presentation	*Hypothesis 13c* r = 0.20 to 0.34	*Hypothesis 13d* r = 0.20 to 0.34	*Hypothesis 13e* r = 0.20 to 0.34	

Note. All time variables are reverse scored to represent speed. Higher correlations indicate better performance while lower correlations indicate worse performance.

1.6 Summary

In sum, this study aimed to provide initial evidence of the disconnect between intelligence testing and real-world intelligence by investigating relationships between abilities, attitudes, and task performance. In order to accomplish this objective, the current study utilized traditional ability tests, non-ability trait measures, and developed measures of proximal variables. These measures were used a baseline by which to compare performance on the three “real world” tasks. The tasks utilized were developed with the intention of maximizing ecological validity, so that they would represent performance outside of a laboratory to the greatest extent possible. It was expected that comparisons of performance on the keychain assembly, chair assembly, and traffic calming tasks would reveal moderate correlations with traditional intelligence assessments and non-ability trait measures. The anticipated moderate correlations would indicate that there is significant room for growth in the way that we measure adult intelligence in a modern context.

CHAPTER 2. METHOD

2.1 Participants

While it is less than ideal to use an undergraduate sample for an adult intelligence study, it is important to note that this a proof-of-concept study. The primary purpose of the study is to determine the validity of using ecologically valid task performance to assess intelligence in a sample that is above average in ability. The use of this sample is justified given that if the results indicate the utility of this approach, it is likely to translate to a less highly selected adult population. Georgia Institute of Technology undergraduate students were recruited as participants through the Georgia Tech participant pool using SONA. Students were compensated with 4.5 hours of course research credit.

In order to be eligible to participate in the study, students were required to be at least 18 years of age and fluently speak and read English. Power analysis using *G*Power*, an open source software package, indicated that the sample size needed to detect the smallest correlation hypothesized was sixty-seven. Therefore, I aimed to recruit at least 70 participants. Unfortunately, data collection was halted prematurely due to the global pandemic associated with COVID-19. Sixty-one participants completed the experiment before data collection concluded. Two participants did not follow instructions on multiple portions of the study and were therefore excluded from analysis, leaving 59 participants (35 men and 24 women) with complete data that were included in analysis.

2.2 Materials and Measures

2.2.1 Traditional Measure of Intelligence

2.2.1.1 Fluid Intelligence

Fluid intelligence was assessed using three traditional measures: Spatial Analogies (Ackerman & Kanfer, 1993), Diagramming Relations from the Educational Testing Service Kit (Ekstrom, French, Harman, & Derman, 1976), and Number Series (Thurstone, 1962). In the Spatial Analogies Test, participants were given nine minutes to complete a thirty-item multiple choice test that involved using analogical reasoning to answer questions containing spatial content (Ackerman, 2000). In the Diagramming Relations test, participants were shown items with three objects and instructed to use logical reasoning to choose from a set of overlapping circles that best characterized the relationship between the objects (Ackerman, 2000). Finally, in the Number Series test, participants were provided items consisting of a series of numbers that follow a particular rule and instructed to use inductive reasoning to determine the next number in the series (Ackerman, 2000).

2.2.1.2 Crystallized Intelligence

Crystalized intelligence was assessed using three traditional measures: Multidimensional Ability Battery (MAB) Comprehension (Jackson, 1998), Wechsler Adult Intelligence Scales-Revised (WAIS-R) Information Test (Wechsler, 1981), and the Extended Range Vocabulary from the ETS Kit (Ekstrom et al., 1976). The MAB Comprehension Test assessed cultural knowledge, the WAIS-R Information Test assessed general knowledge, and the ETS Extended Range Vocabulary Test presented words to participants and required them to choose the words that most closely matched each from a list of choices (Ackerman, 2000).

2.2.2 Non-Ability Traits

2.2.2.1 Interests

The Unisex Edition of the American College Testing Interest Inventory (UNIACT) was used to assess vocational interest themes (Lamb & Prediger, 1981). This 90-item measure provided individual scores for each of Holland's (1963) six interest themes: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional. Responses to the measure were given on six-point scale ranging from 1 ("Strongly Dislike") to 6 ("Strongly Like").

2.2.2.2 Personality Traits

Four of the big five personality traits (extraversion, conscientiousness, openness, and neuroticism), and need for achievement (nAch), were measured using items the from the International Personality Item Pool (IPIP) collection (Goldberg, 2008). Each of these scales contains 10 items. Typical intellectual engagement was measured using a 12-item, short-form scale. Responses to all personality scale items were reported on a scale ranging from 1 ("Very untrue of me") to 6 ("Very true of me").

2.2.2.3 Self-Concept

Verbal and spatial self-concept were measured using a scale in which participants rated the extent to which they agreed with 12 statements reflecting their skills and abilities in each of the two domains. Six items included statements representative of verbal skills and abilities and six items included statements representative of spatial skills and abilities. An example of an item on the verbal scale is: "I can recognize correct English usage (that is, grammar and punctuation)." An example of an item on the spatial scale is: "I can imagine how objects look from a different perspective." Responses were given on a scale ranging from 1("Strongly disagree") to 6 ("Strongly agree").

2.2.2.4 Motivational Traits

Motivational traits were measured using the short-form version of the previously mentioned MTQ (Heggestad & Kanfer, 1997; Kanfer & Ackerman, 2000). The measure contains 48 items, three scales (approach-oriented motivation, competitive excellence, and aversion-related motivation), and six sub-scales (desire to learn, mastery, other-oriented goals, competitiveness, worry in achievement contexts, and emotionality in achievement contexts). Approach-oriented motivation includes desire to learn and mastery, competitive excellence includes other-oriented goals and competitiveness, and aversion-related motivation includes worry in achievement contexts and emotionality in achievement contexts.

2.2.3 Proximal Variables

2.2.3.1 Prior Experience

A prior experience scale was constructed for the purpose of assessing the extent to which participants had exposure to either tasks that were similar to those performed in the experiment or tasks that require skills similar to those performed in the experiment. The scale consisted of 6 items per task, for a total of 18 items. For each of the 18 items, participants were asked to report how often they had engaged in the presented task on a five-point scale ranging from 1 ("Never) to 5 ("Many times"). See **Table 8** for the specific items included in the scale.

Table 8 – Prior experience measures.

In your past experiences, how often have you done each of the following?	
Keychain Task	1. Completed arts and crafts projects in your spare time. 2. Created friendship bracelets, keychains, or related items. 3. Learned a new artistic technique (sketching technique, weaving technique, etc.). 4. Used instructions or diagrams to complete an arts/crafts project. 5. Searched the internet for help completing arts/craft projects. 6. Taught Somebody else to complete and art and crafts project.
Chair Task	7. Used basic tools (screwdriver, wrench, etc.). 8. Assembled basic items (furniture, toys, etc.). 9. Performed basic maintenance on such items (tightening a loose chair, fixing a wobbly table, repairing a broken toy, etc.). 10. Used instruction manuals to put together items. 11. Searched the internet for help putting together items. 12. Worked with other people to assemble items.
Traffic Calming Task	13. Researched a civics problem on the web. 14. Designed a detailed PowerPoint presentation. 15. Delivered a public presentation. 16. Prepared for a debate. 17. Persuaded a group to go along with your solution to a problem. 18. Participated in discussions on how to solve civics problems.

2.2.3.2 Task Self-Efficacy

Similar to prior experience, participants completed self-efficacy measures designed for each specific task. The self-efficacy measures were created using a traditional method in which respondents indicate their level of confidence in completing each task by responding to five items of increasing difficulty on a scale ranging from 0 (“No confidence”) to 8 (“Certain that I can do it”) (Bandura, 1986). For the keychain and chair tasks, difficulty was operationalized as timing. Participants were presented with an image of the completed keychain and chair and asked how confident they were that they could assemble/construct each object in a decreasing amount of time. The keychain task measure asked participants their confidence that they could complete the task in 8 intervals ranging from 40 to 8 min and the chair task measure asked participants their confidence that they could complete the task in 5 intervals ranging from 30 to 10 min. Each of these two measures consisted of 5 items. For the traffic calming task, participants were informed that they would be required to present a brief set of recommendations for a specific suburban roadway/civics project and that they would be allowed access to the Internet. They then indicated their confidence that they could create presentations of increasing quality: ranging from “fairly detailed and documented” to “highly detailed and expertly documented”. The scale for this measure consisted of 4 items.

2.3 Procedure

The procedure for this study took place in two parts. In the first part of the experiment, participants were informed of their rights are research participants, completed the proper consent documentation, and completed the 6 intelligence tests (3 fluid intelligence tests and 3 crystallized intelligence tests) as well as the non-ability trait measures (vocational interests, personality traits, self-concept, and motivational traits) and proximal variable measures (prior experience and task self-efficacy). At the conclusion of the first session, participants were compensated with 2 hours of course research credit. In the second portion of the experiment, participants returned to the lab to complete each of the three real-world tasks. Participants were allowed a five-minute break between the second and third task of the session. The procedure for each of the three tasks is described in the following sections.

2.3.1 Paracord Keychain Task

To begin this task, research assistants provided participants with a two-colored strings of parachute chord that were fused together, a key ring, and a carabiner clip.

Participants were shown an image of a paracord keychain constructed from the materials (see **Figure 1**), asked to construct their own, and given access to Internet to search for help resources (e.g. demonstration videos, written instructions, step-by-step diagrams). Participants were informed that the goal of the task was to construct the keychain as quickly and accurately as possible. As participants completed the task, their actions were record using video cameras and computer screen recording software. Participants were given a maximum of 45 minutes to complete the task because the longest completion time during pilot testing was approximately 45 min. If participants completed the task in less than 45 min, they attempted to make a second keychain in the time remaining. The second keychain was used in exploratory analyses.

Figure 1 – Paracord keychain image presented to participants.

Note. This image was retrieved from: https://www.youtube.com/watch?v=YBwwqB153Ys

2.3.2 Traffic Calming Task

To begin this task, research assistants informed participants that they would be presented with a real-world problem and they were to use the Internet to search for a solution to the problem, develop a PowerPoint presentation communicating their ideal solution, and present their solution to the research assistants at the conclusion of the task. Participants were also informed that they would be scored on the quality of their solution as well as their presentation. Participants were then presented with the following prompt:

> *"Imagine that you are a homeowner in a neighborhood in which the streets do not have sidewalks. Over the course of the past year, you have grown concerned over a safety issue in your community. Drivers are frequently driving well above the speed limit on the streets in your neighborhood, endangering the safety of children playing and adults walking.*
>
> *You have raised issue over the concern to your homeowner's association and have been chosen to represent the homeowners of your committee at a town planning/safety commission meeting. Your job is to develop an efficient plan to* ***calm traffic*** *so as to improve the safety of your street and neighborhood. You will then deliver your plan in a five-minute presentation.*
>
> *Please use this session to develop and prepare your presentation. You may use the computer provided to search the Internet for information that will help you develop your ideal plan. You will present your PowerPoint to researchers at the end of the session."*

Participants were allowed a maximum 60 min to research potential solutions and prepare their presentation because the longest completion time during pilot testing was approximately 60 min. Similar to the keychain construction task, participants' actions were recorded using video cameras and computer screen recording software. At the conclusion of preparation time, participants were given a maximum of five minutes to deliver their presentation to the research assistant, who video recorded the presentation.

2.3.3 Chair Assembly Task

As much as was possible, participants were paired up to complete the chair assembly task. The sessions were arranged so that the chair task was the last task completed by a participant scheduled during an earlier timeslot and the first task completed by a participant scheduled in later timeslot. However, some participants completed the task on their own because no other participant signed up for the slot before or after them or because other participants failed to attend their scheduled session. A total of 32 participants (16 dyads) completed the task in pairs and a total of 27 participants completed the task individually.

To begin the task, research assistants provided participants with the component parts of an "Ingolf" IKEA chair and the instructions provided by IKEA. See **Figure 2** for images of the chair before and after assembly. Research assistants also provided participants with access to the Internet and informed them that they were free to search for additional help resources (e.g. demonstration videos, alternative instructions, completed diagrams). Participants were also informed that the goal of the task was to assemble the chair as quickly and accurately as possible. The participants' actions were recorded using video cameras. Although participants were given the option to use the Internet, none opted to take advantage of this resource. All participants relied on the instruction manual.

Participants were given a maximum of 30 min to complete the task because the longest assembly time during pilot testing was approximately 30 min.

Figure 2 – The IKEA Ingolf chair before and after assembly.

Note. Photographs taken during pilot testing.

At the conclusion of the three tasks, participants were debriefed on the purpose of the experiment and reward an additional 2.5 hours of research credit for the second portion of the study.

2.4 Scoring Video Data

At the conclusion of the study, task-related variables (e.g., performance, timing, time allocation) were determined by viewing video and screen recordings. Rubrics for scoring each of variable of interest were developed and tested for clarity using 10 randomly selected participants. Three research assistants independently scored each variable for these 10 participants. The research assistants then met with me to discuss disagreements and adjust the rubric for clarity. After using the revised rubric to score each of the 10 participants a second time, interrater reliabilities were examined. Based on the recommendation of Koo & Li (2016), Intraclass Correlation 3 (ICC3) was used to calculate interrater reliability. Koo and Li indicated that ICC values of less than 0.5 indicate poor reliability, between 0.5 and 0.75 indicate moderate reliability, between 0.75 and 0.9 indicate good reliability, and greater than 0.90 indicate excellent reliability. While such cutoffs could be considered arbitrary considering interrater reliability is dependent on the extent to which variables differ in terms of subjectivity, all variables in this study exceeded 0.75 (i.e., “good” or “excellent” reliability) except for one -- the quality of traffic calming solution presentations (which had “moderate” reliability). This is somewhat unsurprising given that scoring the quality of a solution to a problem without a definitive correct answer involves more subjectivity than the other ratings. All interrater reliability estimates are presented in **Table 9.** After testing and revising the rubrics for clarity, each participant was scored by one of the three research assistants. To account for the lower interrater reliability associated with the quality of traffic calming solution presentations, two research assistants scored each presentation and a final score was calculated as the average of the two scores. The variables and rubrics associated with each task are described in further detail in the following sections.

Table 9 – Interrater reliability.

Variable	ICC3	Reliability
Keychain quality	0.91	Excellent
Percentage of keychain complete	0.94	Excellent
Keychain time allocation: Computer	0.84	Good
Keychain time allocation: Keychain	1.00	Excellent
Chair time allocation: Instructions	0.84	Good
Chair time allocation: Chair	0.93	Excellent
Dominance (chair task)	0.89	Good
Traffic calming solution presentation score	0.69	Moderate
Traffic calming time allocation: Research	0.95	Excellent
Traffic calming time allocation: PowerPoint	0.87	Good

2.4.1 Keychain Task

The rubric used to guide research assistants through scoring the keychain task and entering data for the task included 7 variables: completion time for both keychain, quality of both keychains, an overall completion score for the task, time spent using the computer, and time spent working on the keychain.

2.4.1.1 Completion Time

Completion time for both keychains was tracked using a stopwatch and recorded.

2.4.1.2 Keychain Quality

Research assistants used a provided rubric to score the quality of both keychains. The maximum quality score for each keychain was 120 points. Research assistants were given a list of the following six errors identified during pilot testing and instructed to deduct 20 points for each error: 1) keychain was constructed using the incorrect weave, 2) keychain was noticeably short (less than 3.5 inches), 3) keychain was curved rather than straight, 4) keychain weave was asymmetrical, 5) paracord was not attached to the carabiner clip correctly, and 6) a significant amount of excess paracord hung from either the carabiner clip or the key ring.

2.4.1.3 Overall Completion score

An overall completion score was given to each participant ranging from 0 to 200 based on the percentage he or she completed of each keychain. For example, if a participant completed 100% of the both keychains, he or she received a score of 200. If a participant completed 100% of one keychain and 50% of the second keychain, he or she received a score of 150. If a participant only completed 75% of their first keychain, he or she received a score of 75.

2.4.1.4 Keychain Task Time Allocation

Two variables were recorded to determine how participants allocated their time during the keychain task. For each minute of video, researched assistants marked whether participants used the computer, worked on the keychain, or both. This led to two different time allocation variables that were independent of one another: time spent on the computer and time spent working on the keychain.

2.4.2 Chair task

The rubric used to guide research assistants through scoring the chair task and entering data included 4 variables: completion time, time spent looking at instructions, time spent working on the chair, and dominance.

2.4.2.1 Completion Time

Completion time for chair assembly was tracked using a stopwatch and recorded.

2.4.2.2 Chair Task Time Allocation

Similar to the keychain task, two variables were recorded to determine how participants allocated their time during the chair task. For each minute of video, research assistants marked whether participants were looking at the instruction manual, assembled the chair, or both. This led to two different time allocation that were independent of one another:

2.4.2.3 Dominance

An additional variable was scored in order to take into account the collaborative nature of the task for participants who assembled the chair in dyads. This variable was included because there was a desire to investigate whether abilities, non-ability traits, and proximal variables influence behavior in a collaborative setting and whether this behavior subsequently influences performance. For each dyad, research assistants gave participants a score ranging from 1-100, with 1 representing “completely submissive” and 100 representing “completely dominant”. The research assistants were given a list of dominant and submissive behaviors that were common during pilot testing to refer to when making their ratings. The list of dominant behaviors included giving instructions, taking a tool from a teammate, maintaining control of the instruction manual, and speaking out loud frequently. The list of submissive behaviors included following teammates’ instructions without offering instructions, giving a tool to a teammate when he or she could have used it on their own, offering the instruction manual to a teammate when he or she could have read it themselves, and not speaking out loud frequently. Each participants’ rating was independent of his or her partner, meaning that there were cases where there were two dominant teammates, two submissive teammates, and one dominant and one submissive teammate.

2.4.3 Traffic Calming Task

The rubric used to guide research assistants through scoring the traffic calming task and entering data for the task included 4 variables: traffic calming solution presentation quality, overall completion time, time spent researching, and time spent preparing the PowerPoint deck.

2.4.3.1 <u>Traffic Calming Solution Presentation Quality</u>

Research assistants were provided with a rubric to score the quality of traffic calming solution presentations. The rubric contained 5 components and each participant received a score ranging from 0 to 15 on each component. Therefore, presentation scores could range from 0 to 75. The five components were selected based on characteristics of high-quality presentations during pilot testing. The five components of the rubric were: the extent to which presentations included a variety of potential options, the extent to which presentations weighed the benefits and downsides of each option, the quality and strength of the argument made by the participants, the evidence used to defend the argument, and the extent to which presentations addressed financial considerations.

2.4.3.2 Traffic Calming Time Allocation

Similar to the other two tasks, two variables were recorded to determine how participants allocated their time during the traffic calming task. For each minute of screen recording video, research assistants marked whether participants were researching traffic calming solutions, preparing their PowerPoint deck, or both. This led to two different time allocation variables that were independent of one another.

CHAPTER 3. RESULTS

The results for this study are presented in four sections. I begin by providing information on high quality and low quality performance on the three real-world tasks. I also provide descriptive statistics for the three tasks. I then present the descriptive statistics and factor solutions for ability and non-ability trait measures. Next, I present the correlational analyses that were used to test hypotheses specified before data collection. Finally, I conclude with a series of exploratory analyses. The exploratory analyses section contains hierarchical regression analysis, factor extension analysis, correlational analyses that include task variables not specified in hypotheses (e.g., performance on the second keychain, time allocation), and gender differences.

Throughout the results section, a series of abbreviations were developed to improve clarity and efficiency. While each abbreviation is described below, it is also provided in the **List of Symbols and Abbreviations** at the beginning of the document. K1 indicates the first keychain attempt, while K2 indicates the second keychain attempt. Traffic calming presentation score was abbreviated as pres. The subscript of co/o indicates the proportion of overall time spent on the computer during the keychain task, k/o indicates the proportion of overall time spent on the keychain during the keychain task, i/o indicates the proportion of overall time spent looking at instructions during the chair task, ch/o indicates the proportion of overall time spent on the chair during the chair task, r/o indicates the proportion of overall time spent researching during the traffic calming task, and p/o indicates the proportion of overall time spent preparing the PowerPoint during the traffic calming task. Also, the subscript of co/k indicates the proportion of time spent on the computer vs. the keychain during the keychain task, i/ch indicates the proportion of time spent on instructions vs. the chair during the chair task, and r/p indicates the proportion of time spent researching vs. preparing the PowerPoint during the traffic calming task. In terms of ability tests, NS indicates the Number Series test, DR indicates the Diagramming Relations test, SA indicates the Spatial Analogies test, and GI indicates the General Information test. Finally, additional subscript abbreviations were developed for non-ability traits. SSC indicates spatial self-concept, VSC indicates verbal self-concept, consc indicates conscientiousness, nAch indicates need for achievement, open indicates openness, neur indicates neuroticism, extra indicates extraversion, TIE indicates typical intellectual engagement, DTL indicates desire to learn, compet indicates competitiveness, emot indicates emotionality, Real indicates Realistic, Inves indicates Investigative, and Con indicates Conventional.

3.1 Task Outcomes

3.1.1 Task Performance

In order to gain a better interpretation of the descriptive statistics associated with the three tasks incorporated in the study, it is necessary to understand not only what constitutes positive performance on each task, but also what constitutes negative performance. The target products for the keychain and chair tasks are represented in **Figures 1 and 2,** respectively. Additional images of high-quality keychains constructed by participants are displayed in **Figure 3. Figures 1-3** can be referenced as representations of ideal performance.

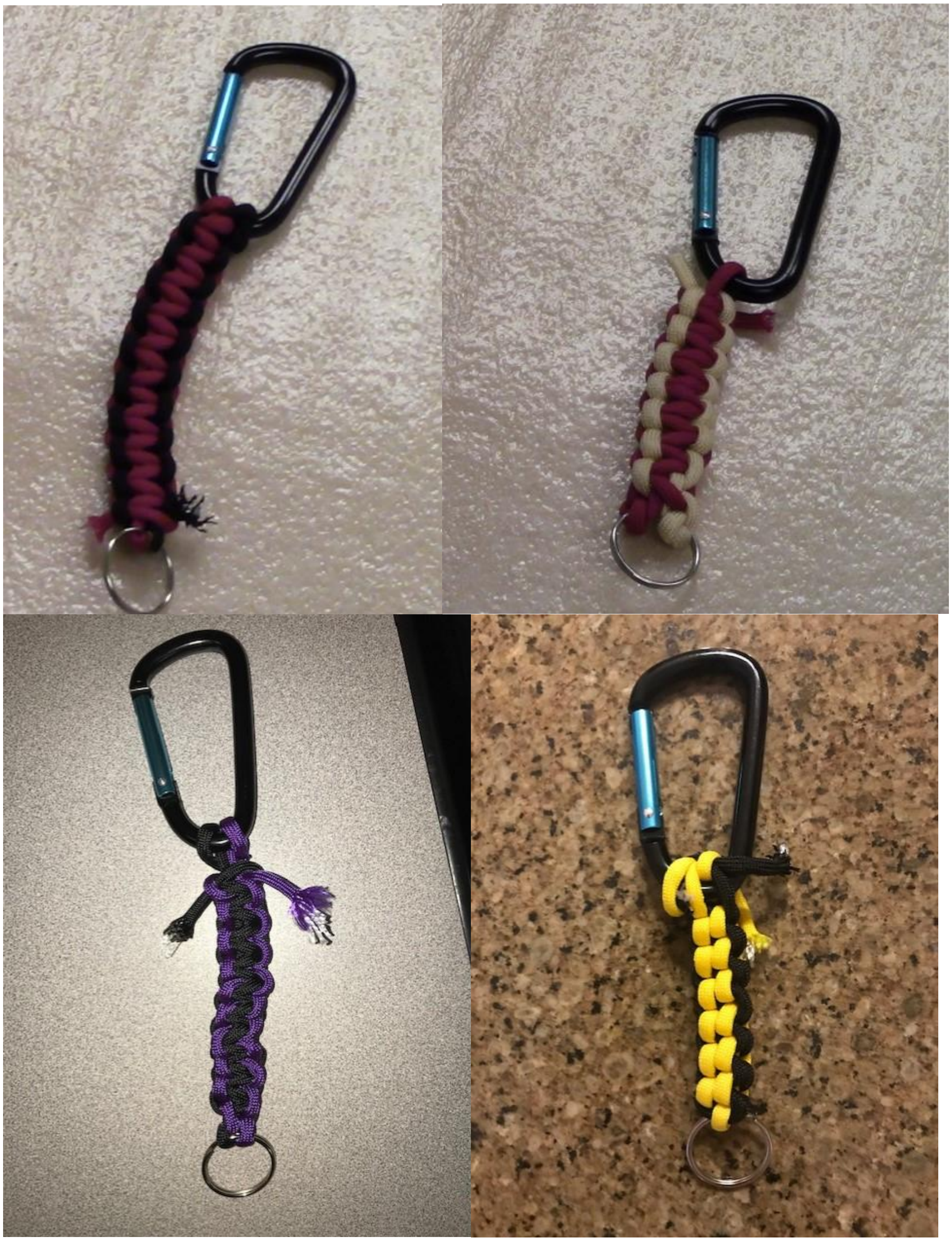
There were six mistakes that were commonly made during the keychain task and were included in the rubric that scored keychain performance. Each of these six mistakes is represented in the images provided in **Figure 4.** The earliest mistake that was made during keychain construction came at the onset of the task. Participants were instructed to construct the keychain depicted in **Figure 1**, which was made using a “cobra” weave. However, some participants did not use the cobra weave and used a different type of weave instead. For example, the image on the bottom right of **Figure 4** displays a keychain constructed using a “snake knot”. The next set of mistakes occurred once participants utilized the proper weave. Participants were penalized for errors during the weaving process, such as creating an asymmetrical weave or a curved weave. At the end of the process, two additional errors were commonly made. Participants often failed to correctly connect the keychain to the key and/or ring or left significant excess paracord hanging off of the end of the keychain. In the image provided to participants, both strings of paracord were connected to the clip and the ring. Many participants only connected one of the strings. Finally, the lowest performers often constructed noticeably short keychains (i.e., less than 3.5 inches in length).

The final chair constructed by participants always looked identical. However, the most common error during assembly occurred when participants put a piece of the chair on backwards and tightened the screws. Inevitably, participants would notice the mistake later in the process and were forced to backtrack to correct it, severely impacting their speed of completion. Traffic calming task performance was a bit more abstract. The lower quality presentations were characterized by a variety of weaknesses that I describe below. The more effective presentations typically included each of these components. First, many participants did not develop or defend a specific solution. They simply presented a laundry list of traffic calming solutions that appeared in their online searches. Next, many participants failed to consider the financial implications of their solutions. Many of the lower quality presentations also did not include evidence (e.g., research studies, statistics) to support their argument. Finally, many participants only presented a single solution and did not consider alternatives or weigh the pros and cons of potential options. The slide deck of a high-quality presentation is provided in **Appendix D** and the slide deck of a low quality presentation is provided in **Appendix E.** The presentation provided in **Appendix D** received a high score because the considered multiple options, weighed the potential benefits and consequences of each outcome, made a definitive argument for a specific solution, defended the proposed solution with evidence, and considered the financial implications. The presentation provided in **Appendix E** received a high score because the participants simply listed a few options rather than arguing for a particular solution, hardly addressed the benefits or consequences of each solution, and did not provide evidence in terms of research or statistics to enhance the presentation.

Figure 3 – High quality keychains.

Note. Both of these keychains received the maximum score of 120 for keychain quality.

Figure 4 – Common keychain mistakes.



Note. The errors during keychain construction that counted against participants were an incorrect weave (bottom right), a curved weave (top left), an asymmetric weave (bottom left), a noticeably short weave (top right), incorrect attachment of paracord to clip and/or ring (top right), excessive paracord hanging off of the keychain (bottom left).

3.1.2 Descriptive Statistics

Means, standard deviations, and ranges for all variables associated with task outcomes were calculated and presented in **Table 10**. There are a few observations that can be noted from these statistics. The keychain quality statistics reveal that the average keychain included 1-2 errors indicated by the rubric, given that 20 points for each error were deducted from a maximum point total of 120. Additionally, participants generally constructed their second keychain 9 minutes faster than their first ($M_{K1} = 19.86$, $SD_{K1} = 10.20$, $M_{K2} = 9.64$, $SD_{K2} = 4.69$). This difference in speed of completion was statistically significant ($t = 6.78$ (82.49), $p < 0.01$) and yielded a large effect size ($d = 1.29$).

In terms of time allocation, three proportions were calculated for each task: a proportion of overall time spent on each task component (i.e., two proportions) and a proportion of time spent on one component vs. the other component (i.e., one proportion). It is worth nothing that using this method, the two proportions involving overall time do not necessarily add up to 1.00 because a participant might look at the computer *and* work on the keychain in the same minute. If a participant did this consistently, he or she might have ended up with two proportions that have a sum somewhat substantially higher than 1.00.

Table 10 – Task outcome descriptive statistics.

Variable	N	Mean	SD	Possible Range	Range
Keychain Task					
K1 Quality	59	85.98	34.67	0.00-120.00	1.00-120.00
K1 Time	59	19.86 min	10.20 min	0.00 min - 45.00 min	5.85-45.00 min
K2 Quality	55	91.29	32.28	0.00-120.00	0.00-120.00
K2 Time	55	9.64 min	4.69 min	1.00 min - 44.00 min	3.15 min - 22.80 min
Overall Keychain Completion	59	180.51	48.12	0.00-200.00	5.00-200.00
K1 Computer time/Overall Time	59	0.48	0.22	0.00-1.00	0.00-1.00
K1 Keychain Time/Overall Time	59	0.90	0.09	0.00-1.00	0.55-1.00
K2 Computer time/Overall time	55	0.10	0.14	0.00-1.00	0.00-0.55
K2 Keychain Time/Overall Time	55	0.85	0.22	0.00-1.00	0.32-1.00
K1 Computer time/Keychain Time	59	0.50	0.26		0.00-1.28
K2 Computer time/Keychain Time	55	0.12	0.14		0.00-0.55
Chair Task					
Chair Completion Time	59	11.79 min	4.72 min	1.00 min - 30.00 min	6.68 min - 15.58 min
Dominance	32	53.06	16.52	0.00-100	5.00-80.00
Instruction time/Overall Time	59	0.52	0.17	0.00-1.00	0.15-0.86
Chair Time/Overall Time	59	0.93	0.09	0.00-1.00	0.67-1.00
Instruction Time/Chair Time	59	0.56	0.18		0.22-0.88

Traffic Calming Task					
Presentation Score	59	49.63	13.03	0.00-75.00	8.50-67.00
Overall Preparation Time	59	45.78 min	14.19 min	1.00-60.00 min	9.67 min - 60.00 min
Research Time/Overall Time	59	0.48	0.19	0.00-1.00	0.00-0.85
Ppt Time/Overall Time	59	0.70	0.16	0.00-1.00	0.31-0.96
Research Time/Ppt Time	59	0.70	0.31		0.000-1.53

Note. K1 = Keychain 1, Key 2 = Keychain 2, SD = Standard deviation.

During the keychain task, participants rarely went an entire minute without working on the keychain, as indicated by high averages ($M_{K1} = 0.85$, $M_{K2} = 0.90$). There was very little variance in proportion of overall time spent working on the first keychain ($SD = 0.09$), which reflects that most participants explored the task in a hands-on manner. Finally, participants spent a lower proportion of their time their time using the Internet on their second attempt, indicating less reliance as they became more familiar with the task. This is reflected by lower means for both proportion of overall time spent on the computer ($M_{K1} = 0.48$, $SD_{K1} = 0.22$, $M_{K2} = 0.10$, $SD_{K2} = 0.14$) and proportion of time spent on computer vs. time spent on keychain on participants' second attempt compared to their first ($M_{K1} = 0.50$, $SD_{K1} = 0.26$, $M_{K2} = 0.12$, $SD_{K2} = 0.14$). Both of these differences in time allocation between keychain attempts were statistically significant and characterized by large effect sizes ($t_{co/o}(96.89) = 10.06$, $p_{co/o} < 0.01$, $d = 2.06$; $t_{co/k}(91.64) = 9.86$, $p_{co/k} < 0.01$, $d = 1.82$).

In regard to the chair assembly task, participants displayed an average completion time of just under 12 minutes ($M_{time} = 11.79$, $SD_{time} = 4.72$). Similar to the keychain task, participants rarely went an entire minute without working on the chair ($M_{ch/o}$ 0.93, SD $_{ch/o}$ = 0.09). They also spent around half of their time looking at the instruction manual, as evidence by both the proportion of overall time spent looking at the instruction manual ($M_{i/o} = 0.52$, $SD_{i/o} = 0.17$) and the proportion of time spent on the instructions vs. the chair ($M_{i/ch} = 0.56$, $SD_{i/ch} = 0.18$). Finally, the average participant displayed behaviors that were moderately dominant ($M_{dom} = 53.06$, $SD_{dom} = 16.52$), with a few extremely dominant or submissive participants that led to a leptokurtic distribution characterized by heavier tails (kurtosis = 1.42).

On the traffic task, participants achieved an average presentation score of just under 50, meaning the average presentation lost approximately 25 points (M_{pres} = 49.63, SD_{pres} = 13.03). Participants spent an average total of approximately 46 minutes total on the task ($M_{prep\ time}$ = 45.78, $SD_{prep\ time}$ = 14.19). In terms of time allocation, the average participant spent time researching during at least a portion of half of the minutes they spent preparing ($M_{r/o}$ = 0.48, $SD_{r/o}$ = 0.19) and at least a portion of around 70% of the minutes they spent preparing working on their PowerPoint deck ($M_{p/o}$ = 0.70, $SD_{p/o}$ = 0.16).

3.2 Abilities, Non-Ability Traits, and Proximal Variables

3.2.1 Trait Composites and Complexes

Trait complexes are constructed by considering common variance among individual differences of different domains (e.g., intelligence, personality, interests). They are useful in scenarios in which researchers desire a parsimonious explanation for a large number of traits that predict knowledge or skill acquisition (Ackerman, Kanfer, & Beier, 2013). Trait complexes were calculated in two steps. First, exploratory factor analysis was used to ensure that factors representing traits that typically share common variance were extracted. Second, trait complex composites were calculated using sums of unit-weighted z-scores of the individual measures.

In the case of ability tests, there were two abnormalities in the data. First, the Number Series Test, a test of fluid intelligence, had unusually low correlations with the other two tests of fluid intelligence ($r_{NS, SA}$ = 0.11, $r_{NS, DR}$ = 0.17) and an unusually high correlation with a test of crystallized intelligence ($r_{NS, GI}$ = 0.50). The correlations among the ability tests are displayed in **Table 11.** As a result of this abnormality, the Number Series test was excluded from analysis. Additionally, an exploratory factor analysis resulted in a Heywood case – meaning that communalities were estimated at values greater than 1.00. In order to account for this, a principal components analysis was conducted and rotated using a varimax rotation. The results indicated two distinct components: one representing fluid intelligence and one representing crystallized intelligence. The rotated component matrix is provided in **Table 12.** Tests of crystallized intelligence have higher loadings on Component 1 and tests of fluid intelligence have higher loadings on Component 2. Based on these results, a fluid intelligence composite measure was calculated using the Spatial Analogies and Diagramming Relations tests and a crystallized intelligence composite measure was calculated using the MAB Comprehension, General Information, and Extended Range Vocabulary tests. The fluid and crystallized intelligence composites were moderately correlated (r = 0.47).

In order to create non-ability trait composites, all personality traits, vocational interest themes, and motivational traits were included in a principal axis factor analysis and the solution was rotated using a direct artificial personal probability function rotation (DAPPFR) (Tucker and Finkbeiner, 1981). Parallel analysis indicated a seven-factor solution. However, this relatively liberal estimate is likely a result of the study's modest sample size. Because a seven-factor solution would significantly over-factor the number of variables included (i.e., 18), a more appropriate five-factor solution was used to create the trait complexes. See **Table 13** for the rotated five-factor solution and **Table 14** for the factor correlation matrix associated with the solution. Four non-ability trait complexes were derived based on the factor solution: 1) Conscientiousness/nAch/Desire to learn/Mastery, 2) Other-oriented goals/Competitiveness/Enterprising interests/Conventional interests/Extraversion, 3) Neuroticism/Worry in achievement contexts/Emotionality in achievement contexts, and 4) Openness/TIE/Investigative interests/Artistic interests.

Table 11– Ability test correlation matrix.

	Comprehension	Information	Vocabulary	Number Series	Spatial Analogy	Diagram Relations
Comprehension						
Information	0.41					
Vocabulary	0.31	0.69				
Number Series	0.20	0.50	0.26			
Spatial Analogies	0.19	0.35	0.25	0.11		
Diagram Relations	0.46	0.41	0.39	0.17	0.50	

Note. $N = 59$. Correlates larger than $r = 0.35$ are significant at the $p < 0.05$ level.

Table 12 – Ability test rotated component matrix.

	Component 1	Component 2
MAB Comprehension	*0.55*	*0.33*
General Information	*0.86*	0.22
Extended Range Vocabulary	*0.88*	0.10
Spatial Analogies	0.09	*0.88*
Diagramming Relations	*0.37*	*0.77*

Note. MAB = Multidimensional Aptitude Battery. Loadings greater than 0.30 are presented in *italics* to indicate salience.

Table 13 – Non-ability trait rotated factor solution.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Openness	***0.71***	-0.11	0.08	0.06	0.04
Conscientiousness	*0.33*	-0.03	-0.08	***0.54***	0.06
Extraversion	0.16	***0.44***	-0.03	-0.04	*0.60*
Neuroticism	0.07	-0.10	***0.48***	-0.09	-0.22
nAch	0.01	0.06	-0.07	***0.67***	-0.03
TIE	***0.61***	0.04	-0.07	*0.32*	*-0.31*
Desire to learn	*0.44*	-0.08	0.01	***0.65***	*-0.34*
Mastery	0.28	0.05	0.06	***0.77***	-0.24
Other-oriented goals	-0.11	***0.60***	*0.39*	0.07	-0.10
Competitiveness	-0.22	***0.47***	0.09	0.22	-0.02
Worry in achievement contexts	0.06	0.04	***0.76***	-0.03	-0.01
Emotionality in achievement contexts	-0.05	0.14	***0.88***	0.02	0.04
Investigative interests	***0.60***	0.03	0.19	-0.01	-0.10
Realistic interests	*0.43*	0.03	0.03	-0.03	*-0.53*
Artistic interests	***0.62***	0.08	-0.07	*-0.30*	-0.00
Social interests	*0.55*	0.49	0.21	0.23	0.02
Enterprising interests	0.10	***0.75***	0.03	-0.02	0.07
Conventional interests	-0.01	***0.49***	-0.13	-0.01	-0.27

Note. Traits included in the trait complexes associated with Factors 1-4 are presented in **bold**. Loadings greater than 0.30 are presented in *italics* to indicate salience.

Table 14 – Non-ability trait factor correlation matrix.

	Factor 1	**Factor 2**	**Factor 4**	**Factor 4**	**Factor 5**
Factor 1					
Factor 2	-0.01				
Factor 3	-0.21	-0.23			
Factor 4	0.17	0.24	-0.40		
Factor 5	0.36	-0.12	-0.23	0.51	

Note. $N = 59$. Correlates larger than $r = 0.35$ are significant at the $p < 0.05$ level.

3.2.2 Descriptive Statistics and Reliability Estimates

Descriptive statistics and internal consistency reliability estimates for all ability tests, non-ability traits, and proximal measures were calculated and presented in **Tables 15 and 16.** The tables include descriptive statistics for each individual measure, as well as for trait complexes. All statistics were calculated using the sample of 59 participants except for task self-efficacy. Two participants has scores of zero self-efficacy for all three tasks. These participants were removed from all analyses including self-efficacy because the self-efficacy scales were designed for participants to indicate some minimal level of confidence.

Table 15 – Descriptive statistics and reliability estimates for ability tests.

Test	Items	Mean	sd	Possible range	Range	α	S-B coef.
Fluid Intelligence							
Spatial Analogies	30	21.93	4.96	0-30	8-30	0.86	
Diagramming Relations	30	23.10	6.38	0-30	5-30	0.91	0.85
Number Series	20	12.37	2.50	0-20	7-20	0.71	
Crystallized Intelligence							
Comprehension	28	21.53	2.94	0-28	12-26	0.62	
General Information	24	9.53	2.58	0-24	4-15	0.64	
Vocabulary	48	21.24	7.31	0-48	3-36	0.87	0.89
Composites							
Fluid Intelligence	2	0.00	1.00		-2.94 - 1.42	0.65	
Crystallized Intelligence	3	0.00	1.00		-2.46 - 1.94	0.60	

Note. sd = standard deviation. α = Chronbach's alpha internal consistency estimate. S-B coef = Spearman-Brown coefficient derived from the Spearman-Brown prophecy formula designed to estimate total test reliability from split-half reliability.

Measure	Items	Mean	sd	Possible Range	Range	α
Self-Concept						
Verbal Self Concept	6	30.36	4.86	6-36	13-36	0.88
Spatial Self-Concept	6	30.36	3.75	6-36	20-36	
Personality						
Typical Intellectual Engagement	12	51.32	9.08	12-72	29-67	0.85
Extraversion	10	39.34	10.04	10-60	17-58	0.91
Openness to experience	10	45.81	8.17	10-60	30-60	0.82
Need for achievement	10	48.97	7.34	10-60	27-60	0.88
Conscientiousness	10	43.19	7.85	10-60	19-57	0.88
Neuroticism	10	30.29	9.39	10-60	13-49	0.88
Motivational Traits						
Desire to Learn	8	30.12	5.19	8-48	26-48	0.83
Mastery	8	38.93	5.22	8-48	23-47	0.80
Other-Oriented Goals	7	29.29	7.18	7-42	8-41	0.92
Competitiveness	6	21.51	6.55	6-36	7-33	0.89
Worry in Achievement Contexts	10	40.80	10.06	10-60	15-60	0.90
Emotionality in Achievement Contexts	9	29.41	9.39	9-54	11-50	0.86
Interests						
Realistic	15	52.71	13.07	15-90	22-86	0.88
Investigative	15	57.64	15.77	15-90	15-84	0.93
Artistic	15	49.75	16.76	15-90	15-78	0.92
Social	15	64.66	11.10	15-90	36-86	0.85
Enterprising	15	52.63	15.29	15-90	23-82	0.92
Conventional	15	49.22	16.05	15-90	20-85	0.93

Composites					
Conscientiousness, nAch, Desire to learn, Mastery	4	0.00	1.00	-3.00 – 1.69	0.88
Neuroticism, Worry in achievement contexts, Emotionality in achievement contexts	3	0.00	1.00	-2.16 – 2.09	0.87
Openness, TIE, Artistic, Investigative	4	0.00	1.00	-2.29 – 1.58	0.67
Other-oriented goals, Competitiveness, Enterprising, Conventional, Extraversion	5	0.00	1.00	-2.28 – 1.92	0.66

Table 16– Descriptive statistics and reliability estimates for non-ability trait and proximal variable measures.

Note. sd = standard deviation. α = Chronbach's alpha internal consistency estimate.

3.3 Correlational Analysis: Task Performance

In order to investigate the hypotheses specified prior to the study, a series of correlations were computed and compared against the magnitude of correlations hypothesized prior to data collection. For each task, abilities, non-ability traits, and proximal variables were correlated with task performance variables and the observed correlation was tested against the hypothesized correlation using Fisher's *r*-to-Z transformation coupled with a *z*-test. In this scenario, results are consistent with hypotheses when the observed correlation is *not* statistically different from the hypothesized correlation. These results are presented in **Appendices A, B, and C. Appendix A** provides correlations associated with the keychain construction task, **Appendix B** provides correlations associated with the chair assembly task, and **Appendix C** provides correlations associated with the traffic calming task.

However, there is an important caveat associated with these analyses. Given that the analytical approach implemented to test hypotheses implies that results are consistent with a given hypothesis when the hypothesis test is not significant, it is important to consider the likelihood that an alternative correlation would have been detected had it existed in the population. Many researchers often conduct post-hoc analyses to determine if studies held sufficient power to detect an effect significantly different from zero. However, this is not necessarily appropriate because a correlation significantly different from zero is not necessarily meaningful (Ackerman & Hambrick, 2020). In this scenario, it is more meaningful to determine whether the study held sufficient power to detect a correlation significantly different from the *hypothesized* value had it existed in the population. In order to investigate this, post-hoc power analyses were conducted using a correlation bivariate normal model test. Cohen's criteria of small, moderate, and large correlations are differentiated by 0.20: 0.10, 0.30, and 0.50, respectively (Cohen, 1988). Therefore, I conducted power analyses to determine the likelihood that I would have detected a correlation that was 0.20 less than the correlation hypothesized given the sample size utilized in the study. Detecting such a small difference in correlations often requires a relatively large sample size. Collecting data from this many participants in this study was not feasible for two reasons: 1) the study required a significant amount of time due to the hands-on nature of the tasks, and 2) data collection was halted prematurely due to the pandemic. See **Appendix F** for a table containing the power that I achieved to detect meaningful differences from each hypothesized correlation and the sample size that I would have needed to detect such a correlation.

I did not achieve a power of 0.80 to detect a correlation meaningfully different from any of the values hypothesized. Therefore, the conclusions that can be drawn from the hypothesis tests are limited. As a result, I have placed the hypothesis test analyses in appendices and devoted this section to discussing whether or not the correlations between variables associated with each hypothesis is *meaningful* rather than *significant.* There is an important distinction between correlations that are significant and correlations that are meaningful (Ackerman & Hambrick, 2020). A small correlation might be significantly different from zero or not statistically significant form the hypothesized value, but the magnitude of such a correlation may not be meaningful in the context of the study. Therefore, this space it is better utilized discussing meaningfulness rather than significance. In the context of this study, I consider correlations between abilities and task outcomes that are greater than or equal to +/- 0.30 and correlations between non-ability traits and task outcomes that greater than or equal to +/- 0.20 to be meaningful. I set the threshold for meaningfulness of correlations of task outcomes with non-ability traits lower than those of abilities because of the fact that the study was completed in a laboratory environment that was likely to elicit a maximal performance situation rather than a typical performance situation (Beier & Ackerman. 2005).

3.3.1 Keychain Task

In order to investigate relationships between person attributes and keychain task performance, abilities, non-ability traits, and proximal variables were correlated with the quality of the first keychain constructed and speed of the first keychain constructed. Performance on the second keychain and time allocation were not addressed in hypotheses prior to data collection and were subsequentially included in exploratory analysis. All correlations associated with keychain task performance included the full sample of 59 participants except for those associated with task-self efficacy. Two participants indicated zero self-efficacy for the task, which goes against the way the self-efficacy measurements were designed. They were designed so that each participant could indicate a minimal level of efficacy. Subsequentially, these two participants were removed from analyses associated with keychain self-efficacy.

Of the many correlations investigated, 10 correlations between individual difference measures and task performance can be interpreted as meaningful according to the criteria provided above (i.e., at least +/- 0.30 for abilities and at least +/- 0.20 for non-ability traits). In terms of abilities, participants with higher fluid intelligence constructed higher quality keychains than participants with lower fluid intelligence (r = 0.36). While correlations between fluid intelligence and completion speed (r = 0.18) as well as crystallized intelligence and keychain quality (r = 0.24) were in the expected directions, neither crossed the threshold necessary to be interpreted as a meaningful correlation. Interestingly, while participants with higher crystallized intelligence tended to construct high quality keychains, they also took longer to construct their keychains than participants with lower crystallized intelligence (r = -0.25). This is somewhat surprising, but the magnitude of the correlation is not strong enough to be considered meaningful.

Two personality traits that were hypothesized to show relationships with performance on the keychain task (conscientiousness and neuroticism) showed meaningful correlations with completion speed but not with keychain quality. Individuals who were higher on conscientiousness and lower on neuroticism were more likely to complete their first keychain attempt faster than individuals who were lower on conscientiousness and higher on neuroticism (r_{consc} = 0.24; r_{neur} = -0.21). However, the correlations between these same traits and quality of keychain constructed were not meaningful (r_{consc} = 0.01; r_{neur} = 0.08), which suggests that perhaps these personality traits play a more crucial role in speed than accuracy in this specific task. One other personality trait that was not included in hypotheses prior to data collection displayed a meaningful correlation with keychain quality. Participants higher on openness to experience generally constructed higher quality keychains than participants who were lower on openness to experience (r_{open} = 0.22).

There are three other non-ability traits that displayed meaningful correlations with keychain task performance. Spatial self-concept was positively correlated with both keychain quality and completion speed at magnitudes that can be interpreted as meaningful ($r_{SSC, qual}$ = 0.26, $r_{SSC, speed}$ = 0.29). Therefore, self-concept likely played a key role in determining both speed and accuracy components of performance for this task.

Additionally, two vocational interest themes showed meaningful correlations with keychain quality. As expected, individuals who scored higher on the Realistic interest theme constructed higher quality keychains than individuals who scored lower on the Realistic interest theme ($r = 0.37$). No predictions were made in regard to the relationship between keychain task performance and Artistic interest theme. However, the correlation between Artistic interests and keychain quality ($r = 0.20$) crossed the threshold necessary to be interpreted as meaningful. In hindsight, this is not surprising given that the keychain task likely required artistic skills that individuals who score higher on the Artistic interest theme are more likely to possess compared top the rest of the sample. All MTQ sub-scales as well as the remaining personality traits and vocational interest themes measured did not display meaningful correlations with keychain task performance.

In terms of proximal variables, prior experience and task self-efficacy displayed very different results associated with keychain task performance. Prior experience, as expected, was positively and meaningfully correlated with both keychain quality and completion speed ($r_{\text{PE, qual}} = 0.28$, $r_{\text{PE, speed}} = 0.38$). However, the results associated with task self-efficacy were surprising. Although the correlations were relatively small and the magnitudes did not cross the threshold needed to be interpreted as meaningful, individuals with higher task self-efficacy actually constructed lower-quality keychains than individuals with lower task self-efficacy ($r = -0.08$), but completed the task faster ($r = 0.12$).

3.3.2 Chair Task

The analyses associated with the chair assembly task are somewhat complex given that 32 participants completed the task in pairs and 27 participants completed the task individually. The dependent variable associated with performance on this task is speed of completion. In order to determine whether there was a fundamental difference in chair task performance between the dyads and individuals, an independent samples *t*-test was conducted to assess whether there was a significant difference in chair assembly completion speed between the two groups. On average, dyads assembled the chair around 4 minutes faster than individuals ($M_{\text{dyads}} = 9.89$, $M_{\text{individuals}} = 14.03$). The *t*-test revealed that this difference was in fact significant ($t = -3.51$, $p < 0.01$). Because of this difference, data for individuals and dyads were not combined. Ultimately, completing the chair in dyads is a critical component of the study's design. This was crucial because allowing participants to complete the task with another person enhanced the ecological validity of the experiment. Additionally, hypotheses were developed under the assumption that participants would complete the task in pairs. Accordingly, only the data used from the dyads were used to assess the study's hypotheses and included in this section. See **Appendix B** for a table that describes the observed correlations and hypothesis tests. One participant indicated that they had zero self-efficacy for the task and they were removed from analyses associated with chair task self-efficacy, just as the two participants who reported zero self-efficacy for the keychain task. A component of additional analyses was conducted using the data from participants who completed the task individually in which data from both dyads and individuals were integrated, and those results were included in the exploratory analyses section.

Abilities did not display meaningful relationships with chair task performance. The correlations of chair completion speed with fluid and crystallized intelligence were both in the expected directions ($r_{\text{fluid}} = 0.18$, $r_{\text{crystallized}} = 0.10$), but both were smaller than expected and neither magnitude was large enough to cross the threshold used to define meaningful correlations in the context of this study. However, there were eight non-ability traits that were meaningfully correlated with chair task performance. In terms of personality, the meaningful findings were consistent with the relationships hypothesized prior to data collection. Individuals who were higher on conscientiousness and nAch and lower of neuroticism were likely to assemble that chair faster than individuals who were lower on conscientiousness and nAch and higher on neuroticism ($r_{\text{consc}} = 0.40$, $r_{\text{nAch}} = 0.34$, $r_{\text{neur}} = -0.26$). Results also suggest that motivational traits play an important role in performance on the chair task. Participants who scored higher on the approach-oriented motivation sub-scales (i.e., desire to learn, mastery) and lower on the aversion-relation motivation sub-scales (i.e., worry in achievement contexts, emotionality in achievement contexts) assembled the chair faster than participants who scored lower on the approach-oriented motivation sub-scales and higher on the aversion-related motivation sub-scales ($r_{\text{DTL}} = 0.48$, $r_{\text{mast}} = 0.22$, $r_{\text{worry}} = -0.46$, $r_{\text{emot}} = -0.43$).

It is worth noting that while many of the correlations associated with non-ability traits and chair task performance are of magnitudes large enough to be considered meaningful, several are noticeably higher than hypotheses predicted (e.g., conscientiousness, desire to learn, worry in achievement contexts). All vocational interest themes as well as the personality traits and MTQ sub-scales not mentioned displayed non-meaningful correlations with completion speed on the chair task. Finally, the results associated with the relationships of prior experience and task-self efficacy with chair task performance were similar to those associated with the keychain task. Specifically, prior experience ($r_{\text{PE}} = 0.34$) displayed a meaningful correlation with completion time, while completion time ($r_{\text{PE}} = 0.16$) did not.

3.3.3 Traffic Calming Task

In order to investigate the relationships between with traffic calming task performance, abilities, non-ability traits, and proximal variables were correlated with participants' traffic calming solution presentation scores. Task time allocation was not addressed in specific hypotheses and was therefore included in the exploratory analyses section. Similar to the keychain task, two participants indicated zero self-efficacy to complete this task and were therefore excluded from analysis containing task self-efficacy measures.

Only two individual differences measures showed meaningful correlations with traffic calming task performance. Consistent with the results from the keychain task, individuals higher in fluid intelligence constructed higher-quality traffic calming solution presentations than individuals lower in fluid intelligence (r = 0.42) and the magnitude of the correlation was high enough for the relationship to be interpreted as meaningful. The correlation between crystallized intelligence task performance was in the direction anticipated (r = 0.21), however the magnitude was smaller than 0.30, indicating that the association is not meaningful in the context of this study. Self-concept also played a meaningful role in task performance. While higher levels of spatial self-concept were associated with higher levels of performance on the keychain and chair tasks, higher levels of verbal self-concept were associated with higher levels of performance on the traffic calming task (r = 0.29). This is in line with expectations because the traffic calming task was designed to sample verbal abilities with the other two tasks were designed to sample spatial and perceptual speed abilities. The remaining non-ability traits did not display meaningful correlations with traffic calming performance. However, it is worth noting that several personality traits (e.g., openness to experience, TIE) and vocational interest themes (e.g., Investigative, Artistic) were correlated with performance at magnitudes just shy of the 0.20 threshold. These traits might warrant further investigation in future studies. Finally, both proximal variables measured in this study did not display meaningful correlations with task performance ($r_{PE} = 0.01$, $r_{SE} = 0.01$).

3.3.4 Correlations between task outcomes.

In order to investigate the potential of broad vs. narrow predictors of real-world intellectual performance, each task outcome was correlated with other task outcomes. See **Table 17** for the complete correlation matrix associated with these analyses. Correlations between keychain task and traffic calming task performance used the entire sample of 59 participants, while correlations involving the chair task included only the 32 participants who completed the task in dyads. Additionally, all keychain variables were derived from participants' first attempt at the task.

In parallel with expectations, higher performance was correlated on each of the three tasks. This means that all things being equal, participants who performed well on one task generally performed well on the other two tasks as well, but not by an overwhelming amount. Also unsurprisingly, performance on the two dependent variables associated with keychain task (i.e., quality and speed) displayed a higher raw correlation than variables from different tasks. However, between-task correlations were generally smaller than expected. Correlations between keychain quality and chair completion time, keychain quality and traffic calming solution presentation quality, keychain completion time and chair completion time, and chair completion time and traffic calming all fell below the threshold of +/- 0.30 that typically characterizes moderate correlations. The only hypothesis that fell within the hypothesized range was the correlation between keychain completion time and traffic calming solution presentation quality. This correlation is also the only correlation between task outcomes that crosses the 0.30 threshold that characterizes meaningful correlations in the context of this study. Together, these results tend to favour the "narrow" side of the debate between broad vs. narrow predictors because the small correlation between task outcomes indicates specialization.

Table 17 – Correlations between task outcomes.

	Keychain quality	Keychain Speed	Chair Speed	Traffic Solution
Keychain quality				
Keychain speed	0.46			
Chair speed	0.16	0.14		
Traffic Solution	0.18	0.33	0.13	

3.4 Exploratory Analyses

Although the analyses presented in this section were not conducted to test relationships or effects hypothesized prior to data collection, they provide supplementary information that aids in interpretation of the study's results. The exploratory analyses are divided into seven sub-sections. The first two sub-sections build upon the analyses presented in the previous sections; they involve the relationships between person attributes and "real-world" task performance. All results presented in the previous section are correlational. The results in these two sub-sections extend these by investigating the extent to which person attributes predict task performance and the relationship between task performance and the ability and non-ability trait factor structures presented in **Section 3.2**.

The next three sub-sections in this section involve task measures that were collected but were not included in hypotheses for one of two reasons: 1) theory did not provide sufficient justification to make confident predictions, or 2) the variables were not indicative of performance, yet provide additional insight into how individuals with different attributes approach real-world tasks. The next sub-section involves integrating the unused data from the chair task (i.e., the participants who completed the task individually). Finally, gender differences in both criterion and predictor variables are explored in the last sub-section. All exploratory analyses were conducted using ability composites and non-ability trait complexes for the purpose of parsimony and to avoid issues of multicollinearity in regression analyses.

3.4.1 Hierarchical Regression

For each of the three tasks, a series of 4 linear regression models was computed in order to determine which person attributes predict performance and the extent to which additional attributes provide incremental validity in prediction of performance. For each task outcome, variables were entered in order from distal to proximal. Ability composites were entered first, followed by non-ability trait complexes, followed by self-concept measures, followed by proximal variable measures (i.e., prior experience and task self-efficacy). A model comparison approach was taken to assess incremental validity at each step of the process. **Table 1**8 provides the hierarchical regression results for all 4 task performance variables.

Hierarchical regression analyses revealed somewhat modest results. Abilities accounted for a significant proportion of variance in keychain quality (R^2 = 0.13, F = 3.94, p = 0.03), keychain speed (R^2 = 0.16, F = 4.96, p = 0.01), and traffic calming presentation score (R^2 = 0.20, F = 6.88, p < 0.01). The only performance variable in which abilities did not account for a significant proportion of variance was chair completion time (R^2 = 0.04, F = 0.53, p = 0.60). The four non-ability trait complexes provided incremental validity over and above abilities only in the prediction of chair completion time (ΔR^2 = 0.31, Adj ΔR^2 = 0.22, F = 2.88, p = 0.04), which is interesting given that the chair task was completed in pairs while they keychain and traffic tasks were completed individually. The change in R^2 associated with entering self-concept scores and proximal measures were not significant for any of the four variables, indicating that these measures did not provide incremental validity in the prediction of performance over and above ability and non-ability traits.

Table 18 – Hierarchical regression predicting task performance

	K1 Quality				K1 Speed				Chair Speed				Traffic Calming Presentation			
	R^2	ΔR^2	Adj. R^2	Δ Adj. R^2	R^2	ΔR^2	Adj. R^2	Δ Adj. R^2	R^2	ΔR^2	Adj. R^2	Δ Adj. R^2	R^2	ΔR^2	Adj. R^2	Δ Adj. R^2
Step 1	**0.13**		**0.10**		**0.16**		**0.12**		0.04		-0.03		**0.20**		**0.17**	
Step 2	0.15	0.02	0.05	-0.05	0.20	0.04	0.11	-0.01	0.35	**0.31**	0.19	**0.22**	**0.25**	0.05	**0.16**	-0.01
Step 3	0.20	0.05	0.07	0.02	0.23	0.03	0.10	-0.01	0.42	0.10	0.21	0.02	**0.29**	0.04	**0.18**	0.02
Step 4	0.28	0.08	0.12	0.05	0.27	0.04	0.11	0.01	0.42	0.00	0.14	-0.07	0.29	0.00	0.14	-0.04

Note. R^2 = Coefficient of determination which represents the proportion of variance in the dependent variable accounted for by the independent variables. Adj. R^2 = Coefficient of determination corrected for the number of predictors included in the model. Statistics that are significant at the $p = 0.05$ level are presented in **bold**. Step 1 = ability trait complexes, Step 2 = non-ability trait complexes, Step 3 = self-concept measures, Step 4 = proximal variables measures.

3.4.2 Dwyer's Extension Analysis

While hierarchical regression analyses provide insights into the prediction of task performance based on person attributes, one might inquire about a related yet distinct question: To what extent are task performance variables related to the principal components and factor solutions derived from the attributes in question. In order to investigate this question, I used a factor extension method commonly known as "Dwyer's extension" analysis (Dwyer, 1937). This method essentially extends a factor solution derived from a set of original variables to a second set (an "extension set") of variables so that one can determine how the second set of variables loads on the factors derived from the original solution. This approach is preferable to including all variables in question in a single factor analysis because oftentimes, the original and extension variables contain linear dependencies. Perhaps most importantly, the dependent variables can influence the predictor factor structure. Dwyer's extension analysis is a method that allows for the calculation of relevant factor loadings without influencing the original "predictor" factor structure (Gorsuch, 1997).

Horn (1973) provided a straightforward equation to calculate an extension analysis solution: $P_e = R_e P_c\,(P_c{}' P_c)^{-1}\,\mathrm{R}_1^{-1}$, where P_e is the pattern matrix associated with extension variables, R_c is a matrix of correlations between the original variables and the extension variables, P_c is the pattern matrix associated with the original solution, and R_1 is a matrix of correlations between the factors from the original solution. This formula was used in order to determine the loadings of the task performance variables associated with the keychain and traffic calming tasks on the principal component solution associated with the ability tests and the factor solution associated with the non-ability traits that were presented in **Section 3.2.1.** Performance on the chair task was not included in this analysis because only 32 participants completed the task as designed (i.e., in pairs). The results of the extension analysis are presented in **Tables 19 and 20.** Each table presents the loadings of the extension loading below the original solution.

There are several important things to note based on these results. First, none of the extension loadings are as large as those associated with the predictor solution (especially in the case of non-ability traits), which fits the narrative that there are components of real-world task performance that are not accounted for by traditional measures. Second, all three task variables loaded more substantially on fluid intelligence than they did on crystallized intelligence. Third, keychain quality and traffic calming presentation quality meaningful loadings on Factor 1, which is associated with openness to experience, typical intellectual engagement, Investigative interests, and Artistic interests. This provides initial evidence that these traits play a role in real-world task performance across two different domains. Additionally, keychain quality displayed a negative loading on Factor 3, which is mostly associated with neuroticism, worry in achievement contexts, and emotionality in achievement contexts. This loading is unsurprising given that that the keychain task involved more time pressure than the traffic calming task did. The positive loading of traffic calming presentation on this factor can likely be explained by the positive loadings of Investigative and Social interests on the same factor. Finally, both variables associated with the keychain task displayed negative loadings on Factor 5, which suggests that individuals high on extraversion who lack approach-oriented motivation and realistic interest performed poorly on the task.

Table 19 – Dwyer's Extension: Abilities.

	PCA solution of abilities	
	Component 1	Component 2
MAB Comprehension	*0.55*	*0.33*
General Information	*0.86*	0.22
Extended Range Vocabulary	*0.88*	0.10
Spatial Analogies	0.09	*0.88*
Diagramming Relations	*0.37*	*0.77*

	Extension of Task Variables	
	Component 1	Component 2
Keychain Quality	0.15	*0.33*
Keychain Speed	-0.11	0.17
Traffic Calming Presentation	0.14	*0.36*

Note. Loadings greater than 0.30 are presented in *italics* to indicate salience.

Table 20 – Dwyer's Extension: Non-ability Traits.

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Openness	***0.71***	-0.11	0.08	0.06	0.04
Conscientiousness	*0.33*	-0.03	-0.08	***0.54***	0.06
Extraversion	0.16	***0.44***	-0.03	-0.04	*0.60*
Neuroticism	0.07	-0.10	***0.48***	-0.09	-0.22
nAch	0.01	0.06	-0.07	***0.67***	-0.03
TIE	***0.61***	0.04	-0.07	*0.32*	*-0.31*
Desire to learn	*0.44*	-0.08	0.01	***0.65***	*-0.34*
Mastery	0.28	0.05	0.06	***0.77***	-0.24
Other-oriented goals	-0.11	***0.60***	*0.39*	0.07	-0.10
Competitiveness	-0.22	***0.47***	0.09	0.22	-0.02
Worry in achievement contexts	0.06	0.04	***0.76***	-0.03	-0.01
Emotionality in achievement contexts	-0.05	0.14	***0.88***	0.02	0.04
Investigative interests	***0.60***	0.03	0.19	-0.01	-0.10
Realistic interests	*0.43*	0.03	0.03	-0.03	*-0.53*
Artistic interests	***0.62***	0.08	-0.07	*-0.30*	-0.00
Social interests	*0.55*	*0.49*	0.21	0.23	0.02
Enterprising interests	0.10	***0.75***	0.03	-0.02	0.07
Conventional interests	-0.01	***0.49***	-0.13	-0.01	-0.27
			Extension of Task Variables		
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Keychain Quality	*0.32*	-0.07	-0.13	-0.01	-0.21
Keychain Speed	0.02	0.03	-0.07	-0.13	0.41
Traffic Calming	*0.29*	0.00	0.21	0.10	0.09

Note. Traits included in the trait complex associated with Factors 1-4 are presented in **bold**.

Loadings greater than 0.30 are presented in *italics* to indicate salience.

3.4.3 *Second Keychain Attempt*

As previously mentioned, participants who completed their first keychain before the time limit of 45 minutes were given the opportunity to construct a second keychain in the time remaining. This was done as a way to obtain extra data in the time that participants were in the lab. While no hypotheses were made regarding the second keychain, analyses led to some observations consistent with what was found regarding the first keychain attempt. Ability composites and non-ability trait complexes, along with self-concept and proximal measures, were correlated with keychain quality on the second attempt, as well an overall score for keychain completion. These correlations are depicted in **Table 21.**

Four of the 59 participants did not complete a keychain and therefore did not attempt a second. Therefore, correlations associated with keychain completion included the full sample, while the correlations associated with quality of the second attempt only included the 55 participants who attempted the second keychain. Additionally, the participants who indicated zero self-efficacy was excluded from the self-efficacy correlations just as they were with the analyses associated with the first keychain attempt, leaving 57 participants who were included in the correlation between self-efficacy and overall completion and 53 participants who were included in the correlations between self-efficacy and keychain quality associated with the second attempt.

In line with findings associated with performance on the first keychain, spatial self-concept displayed moderate correlations with both quality of keychain 2 (r = 0.43) and overall keychain completion (r = 0.32). In addition to spatial self-concept, prior experience and task self-efficacy displayed meaningful correlations with task outcomes. Specifically, individuals with higher levels of prior experience were more likely to complete both keychains ($r = 0.30$) and individuals with higher task self-efficacy completed higher quality keychains on their second attempt ($r = 0.28$). Interestingly, task self-efficacy played a more substantial role in keychain 2 quality than it did in keychain 1 quality ($r = -0.08$). Finally, openness to experience, typical intellectual engagement, Investigative interests, and Artistic interests as well as other-oriented goals, competitiveness, Enterprising interests, Conventional interests, and extraversion displayed small correlations with keychain completion but near-zero relationships with keychain 2 quality.

3.4.4 Dominance

While the primary hypotheses associated with the chair task involved the influence of person attributes on performance, one could also posit that behavior during the task is linked to both performance to person attributes. Correlations were computed in order to explore the relationships between behavior during task engagement (i.e., participant dominance as rated by research assistants) and each variable of interest. Dominance is of particular interest in the context of this task because participants completed it in pairs. Based on this, only the 32 participants who completed the task with a partner were included in this portion of the analysis. These correlations are depicted in **Table 21.** Just as in the primary analysis, an additional participant was removed from the correlation between task self-efficacy and dominance because he or she indicated zero self-efficacy for the task.

Results indicated that dominance did not play a substantial role in task performance given that the correlation between dominance and completion speed was relatively small ($r = 0.09$). However, stronger relationships between person attributes and dominance were apparent. Specifically, dominance was moderately and positively correlated with fluid intelligence ($r = 0.31$), conscientiousness/nAch/desire to learn/mastery ($r = 0.39$), openness/TIE/Investigative/Artistic (*r* = 0.31), spatial self-concept (*r* = 0.42), and prior experience (*r* = 0.36). Crystallized intelligence, however, displayed the opposite effect. Although people with higher crystallized intelligence tended to complete the task slightly faster than individuals with lower crystallized intelligence (*r* = 0.10), they were also less likely to display dominant behavior during task engagement (*r* = -0.37). Finally, one might expect that the trait complex associated with other-oriented goals, competitiveness, Enterprising interests, Conventional interests, and extraversion would be positively correlated with dominance. However, findings indicated a lack of support for this expectation given that the correlation between this trait complex and dominance was relatively trivial (*r* = -0.01).

3.4.5 Integrate chair task groups

Given the study's modest sample size, there is a desire to utilize as much data as possible. However, it is not theoretically justifiable to simply analyse the performance of individuals and dyads as if they are the same for two reasons. First, the pairing of participants was an essential component of the study that was designed to enhance ecological validity. Second, the dyads completed the task at a faster rate by an average of approximately four minutes. As a part of exploratory analyses, I attempted to integrate the data from the individuals and dyads by accounting for the differences in completion time. Specifically, I subtracted four minutes from the completion times of individuals who completed the task on their own and computed correlations between person attributes and completion speed for the entire sample of 59 participants. The purpose of this analysis was to explore whether I may have seen similar trends had the individuals completed the task in dyads. These correlations are depicted in **Table 21.**

The results of this analysis revealed a somewhat similar pattern to what was revealed in by correlations between person characteristics of individuals who completed the task in dyads and chair performance in the main analyses. However, several correlations were attenuated slightly after integration. Specifically, analyses revealed that fluid intelligence ($r = 0.13$), openness/TIE/Investigative/Artistic ($r = 0.19$), spatial self-concept ($r = 0.27$), verbal self-concept ($r = 0.19$), prior task experience ($r = 0.19$), and task-self efficacy ($r = 0.28$) displayed small, positive non-trivial correlations with integrated completion speed. Conversely, Neuroticism/Worry/Emotionality ($r = -0.13$) and Other-oriented goals/Competitiveness/Enterprising/Conventional/Extraversion ($r = -0.27$). were negatively correlated with completion speed. It is particularly interesting that the trait complex including competitive excellence, Enterprising interests, and extraversion was negatively correlated with completion speed given that over half of participants completed the task in dyads.

Table 21 – Exploratory Correlational Analyses.

	Fluid intelligence	Crystallized intelligence	Conscientiousness, nAch, Desire to learn, Mastery	Openness, TIE, Investigative, Artistic	Neuroticism, Worry, Emotionality	Other, Competitiveness, Enterprising, Conventional, Extraversion	Spatial Self-Concept	Verbal Self-Concept	Prior experience	Task Self-efficacy	Chair speed
Keychain 2 Quality	0.23	0.14	0.02	-0.00	-0.08	0.02	0.43	-0.13	0.15	0.28	
Overall Keychain Completion	0.24	-0.04	-0.00	0.14	-0.12	-0.18	0.32	0.06	0.30	-0.01	
Dominance	0.31	-0.37	0.39	0.31	-0.10	-0.01	0.42	0.10	0.36	0.16	0.09
Integrated Speed	0.13	0.04	0.02	0.19	-0.13	-0.27	0.24	0.19	0.19	0.28	

3.4.6 Time Allocation

Completion time has been previously analyzed for both the keychain and chair tasks. However, overall completion time is not the only time-based variable of interest. In this section, I analyzed the differences in how participants allocated the time they engaged with each task. The precedence for considering question of time allocation during task engagement can be traced to Sternberg's (1977) classic study that found that high ability individuals tend to spend more time in preparation stages and less time in task engagement. For the purpose of this study's context, I used three different variables for each task in order to capture time allocation: with overall time spent on different task components, the proportion of overall time spent on difference task components, and the proportion of time spent on one task components vs. another with both task performance variables and participants' scores on individual difference measures. The correlations between time allocation variables and performance variables are presented in **Table 22** and the correlations between time allocation variables and individual differences are presented in **Table 23.**

Interestingly, the relationships between time allocation and completion time differed between the two tasks that required spatial ability and perceptual speed. People who spent more time on the computer generally took longer to construct their keychains ($r_{co/o}$ = -0.21, $r_{co/k}$ = -0.23), while people who spent more time looking at instructions generally assembled the chair faster ($r_{i/o}$ = 0.32, $r_{i/ch}$ = 0.38). This difference might be attributed to the difference in task ambiguity. Perhaps individuals who spent more time on the computer took extensive time to search for different resources (therefore slowing them down), while individuals who spent more time looking at instructions developed a plan that allowed them to assemble the chair quickly, in a straightforward manner. In regard to the traffic calming task, participants who spent more time researching produced slightly higher quality traffic calming solutions ($r_{\text{r/o, qual}} = 0.14$).This is in line with Sternberg's (1977) findings as more preparation time was correlated with better performance.

The correlations computed between individual difference measure scores and task performance variables revealed somewhat different patterns than what one might expect based on Sternberg's (1979) findings. In both the keychain task and the chair task, higher ability individuals generally spent higher proportions of their time on the task itself ($r_{\text{fluid, c/o}} = 0.22$; $r_{\text{fluid, i/o}} = 0.21$), yet ability had very little influence on the amount of time spent in preparation stages ($r_{\text{fluid, k/o}} = 0.02$; $r_{\text{fluid, ch /o}} = 0.01$) or time ratio of time spent on each component($r_{\text{fluid, co/k}} = 0.01$; $r_{\text{fluid, i/ch}} = -0.01$). The same pattern held true for more proximal variables, such as spatial self-concept, prior task experience, and task self-efficacy. This implies that in the case of these hands-on tasks, high performers were able spend time working on the task (i.e., either constructing the keychain or assembling the chair) while spending a similar amount of time on preparation. In other words, they were likely switching back and forth between preparation and task engagement more frequently. In a sense, the high ability individuals were able to multi-task. Conversely, the personality trait complex most strongly associated with traffic calming task performance, Openness/TIE/Investigative/Artistic ($r = 0.26$), and prior experience ($r = 0.25$) were positively correlated with time spent on research This is more indicative of support for Sternberg's findings. These differences may point to a fundamental difference in how high performers allocate their time on tasks that are more hands on vs. mentally demanding.

Table 22 – Correlations between time allocation and performance variables.

	co/o	k/o	co/k	K1 quality	K1 speed	i/o	ch/o	i/ch	Chair speed	r/o	p/o	r/p	pres
co/o													
k/o	0.04												
co/k	0.97	-0.13											
K1 quality	0.00	0.02	0.02										
K1 speed	-0.21	0.12	-0.23	0.46									
i/o	0.22	0.06	0.22	-0.04	0.10								
ch/0	-0.07	0.14	-0.07	0.11	0.12	0.21							
i/ch	0.25	0.04	0.24	-0.05	0.06	0.95	-0.08						
Chair speed	0.07	0.05	-0.05	0.09	0.28	0.32	-0.20	0.38					
r/o	-0.21	0.02	-0.20	0.11	0.01	-0.28	-0.02	-0.28	-0.03				
p/o	0.18	0.03	0.19	0.11	-0.19	0.05	0.09	0.07	0.09	0.16			
r/p	-0.23	0.01	-0.23	0.09	0.09	-0.30	-0.13	-0.30	-0.07	0.81	-0.39		
pres	-0.16	0.33	-0.20	0.18	0.33	0.09	0.05	0.11	0.17	0.14	-0.04	0.13	

Table 23 – Correlations between time allocation and individual differences

	co/o	k/o	co/k	i/o	ch/o	i/ch	r/o	p/o	r/p
Fluid intelligence	0.02	0.22	0.01	0.01	0.21	-0.01	-0.03	0.08	-0.08
Crystallized intelligence	0.03	-0.01	0.05	-0.17	-0.07	-0.14	0.11	0.23	-0.02
Conscientiousness, need for achievement, desire to learn, mastery	0.15	0.08	0.12	0.11	0.16	0.06	-0.02	0.06	-0.01
Openness, TIE, Investigative, Artistic	-0.01	-0.01	-0.01	-0.01	-0.15	0.04	0.26	-0.04	0.24
Neuroticism, Worry, Emotionality	-0.26	-0.26	-0.24	-0.12	-0.02	-0.11	0.12	-0.01	0.07
Other-oriented goals, competitiveness, Enterprising, Conventional, Extraversion	0.06	0.06	0.08	-0.10	0.17	-0.19	0.16	-0.07	0.16
Spatial self-concept	0.08	0.21	0.04	0.07	0.20	0.01	0.05	0.18	-0.01

Verbal self-concept	-0.11	0.04	-0.12	-0.09	-0.27	-0.01	0.06	-0.06	0.10
Prior experience	-0.05	0.18	-0.07	0.09	0.11	0.05	0.25	0.13	-0.07
Task Self-Efficacy	0.05	0.19	0.04	-0.00	0.13	-0.04	-0.04	0.07	-0.01

3.4.7 Gender differences

In Chapter 1, I reviewed a particularly relevant study that investigated gender differences in the ability to assemble IKEA furniture. The study found that men were more efficient than women in assembling an IKEA kitchen trolley and that the differences in assembly speed could be accounted for by gender differences in spatial ability (Wiking et al., 2016). This study's findings suggest that it would be relevant to investigate gender differences on each of the four major criterion variables associated with the tasks utilized in this study. In contrast to Wiking et al.'s (2016) findings, I did not find a significant difference between men and women in terms of assembly speed on the chair task ($t = -1.73$, $p = 0.09$). There was also not a significant gender effect in terms of keychain quality ($t = -0.26$, $p = 0.80$). However, there were significant gender effects for keychain speed ($t = -2.62$, $p = 0.01$) and traffic calming solution presentation ($t = -2.76$, $p = 0.01$). Women finished their first keychain attempt faster than men ($M_{men} = 22.40$, $SD_{men} = 11.34$, $M_{women} = 16.16$, $SD_{women} = 6.96$) and delivered higher quality traffic calming solutions than men ($M_{men} = 46.30$, $SD_{men} = 14.90$, $M_{women} = 54.48$, $SD_{women} = 7.66$).

Given the two significant gender effects on criterion variables associated with the study, I also investigated whether or not there were significant gender effects on predictor variables that might account for the effects described. The only significant gender effect found related to the abilities, non-ability traits, and proximal variables measured in this study was on prior experience related to the keychain task ($t = -4.83$, $p < 0.01$). In this instance, women had higher levels of prior experience associated with the keychain task than men which might explain their ability to complete the task at a more rapid pace. However, it is worth noting that both significant effects might simply be a result of restriction of range associated with the small sample size (i.e., 24 women). Women had noticeably lower standard deviations than men on both significant criterion variables, indicating that this may indeed be the case. As a result, these results should be interpreted with caution.

CHAPTER 4. DISCUSSION

4.1 Summary and Implications

4.1.1 The Gap Between Traditional Assessment and Real-World Performance

My primary objective for conducting this study was to highlight a gap between real-world intellectual performance and the way that psychologists traditionally measure abilities. The approach involved updating and expanding Demming and Pressey's work (1957) by creating three ecologically valid experimental tasks designed to represent tasks that adults might encounter in day-to-day activities. There was a large number of correlations hypothesized prior to data collection, but the overall pattern of results was consistent with the claim that there is a substantial difference between what traditional assessments measure and task performance in a hands-on, real-world context. **Appendices A, B, and C** display 65 correlations between traditional individual differences measures and task performance. Of these 65 correlations, none would be classified as a "large" correlation under Cohen's (1988) criteria. Each correlation was either moderate or small.

The fact that all correlations were moderate or small suggests that there may be more that goes into adult intellect than the traditional assessments capture. This gap is likely related to key situational differences between the ways that traditional assessments are conducted and the way that these tasks were implemented. On one hand, the traditional assessments largely involved questions presented in multiple-choice format in which participants were required to quickly recognize a single correct answer. The environment was very structured and participants were restricted in terms of resources available to them. On the other hand, the tasks completed during the experiment represented less structured and less restrictive environments. Participants were given more hands-on, open-ended tasks in which there was no one best answer or no one best approach. They were also given access to Internet resources that are readily available to them in everyday environments. The situational characteristics of the experimental tasks were intended to more closely resemble situational characteristics of real-world problem-solving contexts, including both work and non-work domains. The small to moderate relationships that this study found between traditional assessments and task performance support Ackerman's (2017) call to develop more relevant assessments of adult capabilities that are critical for performance in the 21st century.

4.1.2 Abilities, Non-Ability Traits, and Task performance

Fluid intelligence and crystallized intelligence were both related to task performance relatively consistently. Additionally, abilities generally predicted aspects of task performance that involved the *quality* of the task solution (i.e., quality of keychain, quality of traffic calming solution) more consistently than aspects of task performance involving *speed* of completion (i.e., keychain and chair completion time). This suggests that abilities might play a more substantial role in real-world tasks in which speed is a not a primary factor in determining performance, as it sometimes is in traditional assessments of fluid and crystallized intelligence. In addition to these correlational findings, results from factor extension analysis highlight a gap between traditional ability assessments and real-world problem-solving capabilities. Task performance variables displayed trivial loadings on a crystallized intelligence components and loadings on a fluid intelligence component that were noticeably lower than those of the Spatial Analogies or Diagramming Relations tests. Together, these results indicate that adult abilities are more nuanced than the traditional way that fluid and crystallized intelligence are measured.

One way that the gap might be closed between ability tests and real-world intellectual performance is by considering non-ability traits that share common variance with abilities

(Ackerman & Heggestad, 1997). Non-ability traits may play an even more critical role in realistic contexts than they do in traditional testing environments because realistic contexts represent "typical" performance while testing environments represent "maximal" performance (Ackerman, 1994). Generally speaking, two of the non-ability trait complexes were positively related to "real-world" task performance and two of the complexes were negatively related to performance. Individuals high on the trait complexes that included conscientiousness, nAch, and approach-oriented motivation, as well as openness to experience, typical intellectual engagement, Investigative interests, and Artistic interests generally displayed above average scores on all tasks. Individuals high on the trait complexes associated with neuroticism and avoidance-related motivation as well as competitive excellence, Enterprising interests, Conventional interests, and extraversion generally displayed below average scores on all tasks. These results indicate that non-ability traits may play a non-trivial role in real-world task performance and should be considered in ecologically valid assessments of real-world problem solving.

Proximal variables were related to task performance as well. Self-concept was moderately correlated with each task performance variable in the study. Spatial self-concept was correlated with performance on both of the hands-on tasks that required spatial ability and perceptual speed (i.e., the keychain and chair tasks) and verbal self-concept was correlated with performance on the more verbally oriented traffic calming task. The role of prior experience and task self-efficacy was a bit inconsistent across tasks. Prior experience seemed to play more substantial role in the keychain and chair tasks than it did in the traffic calming task. These results indicate the possibility that prior experience is more important for tasks that require hands-on engagement and spatial ability/perceptual speed.

Correlations of task-self efficacy with chair task performance were closer to expectations than correlations of task-self efficacy with keychain or traffic calming performance. A potential explanation for this difference might be that individuals had more experience with skills related to the chair task (e.g., using tools, following instructions) than they did for the more unusual traffic or keychain task components. Perhaps participants were able to make more accurate efficacy judgements for this task as a result. While the influence of proximal variables varied in terms of consistency across task domains, it remains plausible that these variables play a role in real-world problem solving. The results of this study indicate that distal non-ability traits and more proximal individual differences might play a role in bridging the gap between how we typically assess intelligence and how adults approach problems in realistic contexts.

4.1.3 Broad vs. Narrow Measurement of Real-World Intelligence

Another critical question pertaining to this study involves the extension of the classic debate between a broad vs. narrow measurements of intelligence (i.e., Brunswik symmetry) to a more ecologically valid, realistic context. This debate was investigated by analyzing correlations between outcomes associated with different tasks. Results indicated that only one correlation between task outcomes (keychain completion time and traffic calming solution score) crossed Cohen's threshold of 0.30 that typically characterizes a moderate correlation. Each of the other correlations between task outcomes are considered small correlations. These small correlations suggest support for the perspective that narrow factors of intelligence accounted for performance. However, one would expect that if narrow factors did play a role, the outcomes on the keychain and chair tasks would be more strongly correlated than outcomes on the traffic calming task because the abilities required by the keychain and chair tasks are somewhat similar. This, however, was not the case. All correlations between keychain and chair task variables were classified as small. Therefore, the evidence gathered to investigate the debate between broad and narrow measurement of real-world intelligence provided somewhat mixed support.

4.2 Limitations

There are five limitations that should be taken into account when interpreting the results of this study. The most obvious limitation to consider is the study's modest sample size. Unfortunately, the data collection phase of this study was truncated because the global pandemic associated with COVID-19, preventing the experimental protocol from being carried out in person. As a result, the study's sample was smaller than expected. This limitation is particularly strong for the results involving the chair assembly task. Only 32 participants were able to complete the task in pairs before data collection was halted. Therefore, any conclusions drawn from the chair assembly task should be interpreted with caution.

Next, the results of the hypothesis tests associated with this study must be interpreted with an important qualification related to power. As Ackerman and Hambrick (2020) pointed out, it is often difficult to detect small differences in correlations because it requires a large sample size to achieved the desired power to do so. For example, the authors point out that it would require a sample size of $N = 1000$ to detect a difference of 0.1 when the smallest correlation is $r = 0.3$. While this might explain the relatively low power estimates revealed by post-hoc analyses, it does not justify concluding without doubt that there are not meaningfully different correlations between predictor and criterion variables in the population.

This study is also limited by the fact that one of the tasks did not turn out to be as difficult as expected. Prior to the study, the chair task was designed to be representative of a challenging furniture assembly task. When the modern adult attempts such a task, he or she often turn to the Internet for help, whether in the form of YouTube videos or diagrams. However, the chair task turned out not to be much of a challenge. Participants were aware that they could access the Internet during task engagement, but none of them took advantage of this option. All participants were able to assemble the chair using the instruction manual provided. The conclusions that can be drawn from the results associated this task are limited because the task did not challenge participants in the way it was intended to.

Another key limitation of this study involves characteristics of the sample population. While the scope of this study extends beyond college undergraduates, an undergraduate sample was used. This was a proof of concept study – meaning that the goal was to highlight an important gap in measurement vs. real-world performance and this sample was sufficient for doing so. However, it is possible that adults of different age groups approach tasks such as these in fundamentally different ways. Also, the sample of Georgia Tech undergraduates is much higher in mathematical, spatial, and verbal abilities than the population large. It is possible that the average 20-25 year old might not be able to construct the paracord keychain in the time allotted. For these reasons, the takeaways of this study should not be generalized to an adult population given that the sample did not accurately represent diverse adult populations.

A final limitation of this study involves two peculiarities encountered during data analysis. First, the Number Series test (a well-validated fluid intelligence test) displayed unusually high correlations with tests of crystallized intelligence and unusually low correlations with tests of fluid intelligence. As a result, I was unable to use this test in my analyses. Second, a "Heywood case" prevented the extraction of a principal axis factor solution for the ability data. This was likely a consequence of the study's modest sample size, but it required the use of a principal components solution rather than a principal axis solution. Each of these circumstances prevented the full use of data analysis procedures that would have been desired and this should be taken into consideration when interpreting this study's results.

4.3 Recommendations for Future Research

While this study reveals mixed support for many hypotheses, it was successful in accomplishing its primary goal: to highlight a disconnect between how abilities are typically measured and the way that adults use abilities in the real-world. Now that this disconnect has been identified, there are a number of ways that this research agenda could be advanced. In this section, I offer three recommendations for future research to consider. First, there is a significant need to develop more relevant assessments of adult knowledge and skills. This study has shown what traditional measures of knowledge and skills *fail* to account for. The next step in this research should involve improving upon these measures and creating new ones. This study provides a potentially promising direction: using tasks that are more representative of the 21st century problem-solving contexts. Future research should build off of this idea and continue developing realistic ways to assess problem-solving. This should involve allowing participants access to modern technology that typically enhances problem-solving (i.e., the Internet). The next step in this process should involve identifying profiles of individuals who can utilize information search strategies and articulate solutions to open-ended problems that do not have definitive solutions.

Second, future studies should integrate measures of abilities, distal non-ability traits, and more proximal variables. At this point, there is extensive empirical evidence of shared variance among ability and non-ability traits (Ackerman 2000; Ackerman, Charro-Premzic, & Furnham, 2011; Ackerman & Heggestad, 1997). I suggest that each trait level (i.e., distal to proximal) plays a role in real-world performance. If there is a desire to improve predictions of intellectual performance in the real-world, it will be necessary to consider a broad variety of traits that range from distal to proximal. Researchers should integrate traits from these different domains in order to capitalize on common variance and create a more holistic picture of real-world performance.

Third, future studies should build upon this work and extend the research approach to broader contexts. This broader context should include both different samples and different task requirements. In terms of samples, future research should examine whether the same gap that exists between traditional assessment and real-world performance exists in samples other than college undergraduates. For example, researchers might look at older populations or specific employment groups. In regard to task characteristics, these tasks were selected as examples of tasks that adults might encounter in the real-world. There are other tasks that adults might encounter in everyday life that require different abilities than the tasks that were used in this study. For example, none of these tasks required numerical ability. Future studies might design tasks that require numerical ability, but are more practically relevant than traditional measures that test numerical ability (i.e., the Number Series test.). Additionally, future studies might investigate tasks that are encountered in everyday job tasks in order to more accurately reflect the problem-solving environment of particular jobs or employment groups.

APPENDIX A. KEYCHAIN TASK CORRELATIONAL HYPOTHESES AND ANALYSIS

Variable 1	Variable 2	Hypothesized Correlation	Observed correlation	Hypothesis Test (Compared against a fixed value)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
Fluid intelligence	Keychain 1 Quality	0.35 to 0.49	0.36	Against 0.35: z = 0.06 p = 0.95 Against 0.49: z = 0.84 p = 0.40	No	Yes	Yes	Yes
Fluid intelligence	Keychain 1 Speed	0.35 to -0.49	0.18	Against -0.35: z = 0.97 p = 0.33 Against -0.49: z = 1.87 p = 0.06	No, but upper bound is close	Yes	No	Mixed
Crystallized intelligence	Keychain 1 Quality	0.30	0.24	z = 0.34 p = 0.73	No	Yes	No	Mixed
Crystallized intelligence	Keychain 1 Speed	0.30	-0.25	z = 2.99 p < 0.001	Yes	No	No	No
Conscientiousness	Keychain 1 Quality	0.30	0.01	z = 1.58 p = 0.11	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Conscientiousness	Keychain 1 Speed	0.30	0.24	z = 0.34 p = 0.73	No	Yes	Yes	Yes
nAch	Keychain 1 Quality	0.30	0.06	z = 1.32 p = 0.19	No	Yes	No	No, can't reject hypothesis but correlation is near-zero

Variable 1	Variable 2	Hypothesized Correlation	Observed correlation	Hypothesis Test (Compared against a fixed value)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
nAch	Keychain 1 Speed	0.30	0.18	z = 0.67 p = 0.50	No	Yes	No	Mixed
Desire to learn	Keychain 1 Quality	0.30	0.12	z = 1.00 p = 0.32	No	Yes	No	Mixed
Desire to learn	Keychain 1 Speed	0.30	-0.07	z = 2.01 p = 0.04	Yes	No	No	No
Mastery	Keychain 1 Speed	0.30	0.05	z = 1.37 p = 0.17	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Mastery	Keychain 1 Speed	0.30	0.02	z = 0.53 p = 0.13	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Spatial Self-Concept	Keychain 1 Quality	0.30	0.26	z = 0.23 p = 0.82	No	Yes	Yes	Yes
Spatial Self-Concept	Keychain 1 Speed	0.30	0.29	z = 0.06 p = 0.95	No	Yes	Yes	Yes
Realistic interests	Keychain 1 Quality	0.30	0.37	z = 0.42 p = 0.68	No	Yes	Yes	Yes
Realistic interests	Keychain 1 Speed	0.30	0.16	z = 0.78 p = 0.43	No	Yes	No	Mixed
Prior Experience	Keychain 1 Quality	0.30	0.28	z = 0.12 p = 0.91	No	Yes	Yes	Yes
Prior Experience	Keychain 1 Speed	0.30	0.38	z = 0.48 p = 0.63	No	Yes	Yes	Yes
Task Self-Efficacy	Keychain 1 Quality	0.50 to 0.60	-0.05	Against 0.50: z = 3.11 p < 0.01 Against 0.60: z = 3.86 p < 0.01	Yes	No	No	No

Variable 1	Variable 2	Hypothesized Correlation	Observed correlation	Hypothesis Test (Compared against a fixed value)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
Task Self-Efficacy	Keychain 1 Speed	0.50 to 0.60	0.12	Against -0.50: $z = 2.23$ $p = 0.03$ Against -0.60: $z = 2.98$ $p < 0.01$	Yes	No	No	No
Neuroticism	Keychain 1 Quality	0.30	0.08	$z = 1.10$ $p = 0.27$	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Neuroticism	Keychain 1 Speed	-0.30	-0.21	$z = 0.46$ $p = 0.64$	No	Yes	Yes	Yes
Worry in achievement contexts	Keychain 1 Quality	-0.30	-0.10	$z = 1.11$ $p = 0.27$	No	Yes	No	Mixed
Worry in achievement contexts	Keychain 1 Speed	-0.30	-0.15	$z = 0.84$ $p = 0.40$	No	Yes	No	Mixed
Emotionality in achievement contexts	Keychain 1 Quality	-0.30	-0.16	$z = 0.78$ $p = 0.43$	No	Yes	No	Mixed
Emotionality in achievement contexts	Keychain 1 Speed	-0.30	-0.07	$z = 1.27$ $p = 0.21$	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Conventional interests	Keychain 1 Quality	0.30	-0.14	$z = 2.38$ $p = 0.02$	Yes	No	No	No
Conventional interests	Keychain 1 Speed	0.30	-0.19	$z = 2.66$ $p = 0.01$	Yes	No	No	No
Other-oriented goals	Keychain 1 Quality	0.30	0.00	$z = 1.64$ $p = 0.10$	No	No	No	No, can't reject hypothesis but

Variable 1	Variable 2	Hypothesized Correlation	Observed correlation	Hypothesis Test (Compared against a fixed value)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
								correlation is in wrong direction and near-zero
Other-oriented goals	Keychain 1 Speed	0.30	-0.11	z = 2.22 p = 0.03	Yes	No	No	No
Competitiveness	Keychain 1 Quality	0.30	0.10	z = 1.11 p = 0.27	No	Yes	No	Mixed
Competitiveness	Keychain 1 Speed	0.30	0.04	z = 1.43 p = 0.15	No	Yes	No	No, can’t reject hypothesis but correlation is near-zero

APPENDIX B. CHAIR TASK CORRELATIONAL HYPOTHESES AND ANALYSIS (DYADS ONLY)

Variable 1	Variable 2	Hypothesized Correlation	Observed correlation for dyads	Hypothesis Test (Compared against a fixed value)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
Fluid intelligence	Time	0.35 to 0.49	0.18	Against 0.35: z =0.70 p = 0.48 Against 0.49: z = 1.35 p = 0.18	No	Yes	No	Mixed
Crystallized intelligence	Time	0.30	0.10	z = 0.80 p = 0.43	No	Yes	No	Mixed
Conscientiousness	Time	0.30	0.40	z = 0.43 p = 0.66	No	Yes	Yes	Yes
nAch	Time	0.30	0.34	z = 0.17 p =0.87	No	Yes	Yes	Yes
Desire to learn	Time	0.30	0.48	z = 0.81 p = 0.42	No	Yes	Yes	Yes
Mastery	Time	0.30	0.22	z = 0.33 p = 0.74	No	Yes	Yes	Yes
Spatial Self-Concept	Time	0.30	0.46	z = 0.72 p = 0.47	No	Yes	Yes	Yes
Realistic interests	Time	0.30	0.13	z = 0.68 p =0.50	No	Yes	No	Mixed
Extraversion	Time	0.30	0.04	z = 1.03 p = 0.30	No	Yes	No	No, can't reject hypothesis and correlation is in correct direction, but strength is near-zero

Conventional interests	Time	0.30	-0.05	z = 1.37 p = 0.17	No	No	No	No, can't reject hypothesis but correlation is in wrong direction and near-zero
Other-oriented goals	Time	0.30	-0.19	z = 1.91 p = 0.06	No	No	No	No, can't reject hypothesis but correlation is in wrong direction.
Competitiveness	Time	0.30	-0.09	z = 1.52 p = 0.13	No	No	No	No, can't reject hypothesis but correlation is in wrong direction and near-zero
Neuroticism	Time	-0.30	-0.26	z = 0.17 p = 0.87	No	Yes	Yes	Yes
Worry in achievement contexts	Time	-0.30	-0.46	z = 0.72 p = 0.47	No	Yes	Yes	Yes
Emotionality in achievement contexts	Time	-0.30	-0.43	z = 0.57 p = 0.57	No	Yes	Yes	Yes
Prior experience	Time	0.30	0.34	z = 0.17 p = 0.87	No	Yes	Yes	Yes
Task self-efficacy	Time	0.50 to 0.60	0.16	Against 0.50: z = 1.45 p = 0.15 Against 0.60: z = 1.99 p = 0.05	No	Yes	No	Mixed

APPENDIX C. TRAFFIC CALMING TASK CORRELATIONAL HYPOTHESES AND ANALYSIS

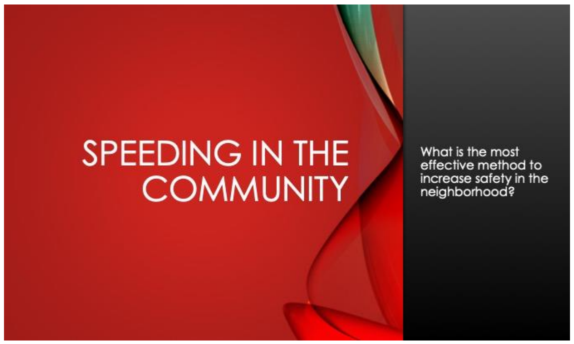
Variable	Variable 2	Hypothesized Correlation	Observed Correlation	Hypothesis Test (Compared against a fixed valued)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
Fluid intelligence	Presentation Score	0.35 to 0.49	0.42	Against 0.35: z = 0.44 p = 0.66 Against 0.49: z = 0.47 p = 0.64	No	Yes	Yes	Yes
Crystallized intelligence	Presentation Score	0.30	0.21	z = 0.51 p = 0.61	No	Yes	No	Mixed
Openness to experience	Presentation Score	0.30	0.18	z = 0.67 p = 0.50	No	Yes	No	Mixed
Typical intellectual engagement	Presentation Score	0.30	0.08	z = 1.21 p = 0.22	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Investigative interests	Presentation Score	0.30	0.18	z = 0.67 p = 0.50	No	Yes	No	Mixed
Artistic interests	Presentation Score	0.30	0.19	z = 0.62 p = 0.54	No	Yes	No	Mixed
Verbal self-concept	Presentation Score	0.30	0.29	z = 0.06	No	Yes	Yes	Yes

Variable	Variable 2	Hypothesized Correlation	Observed Correlation	Hypothesis Test (Compared against a fixed valued)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
				p = 0.95				
Desire to learn	Presentation Score	0.30	0.08	z = 1.21 p = 0.22	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Mastery	Presentation Score	0.30	0.13	z = 0.95 p = 0.34	No	Yes	No	Mixed
Other-oriented goals	Presentation Score	0.30	-0.07	z = 2.01 p = 0.04	Yes	No	No	No
Competitiveness	Presentation Score	0.30	0.01	z = 1.58 p = 0.11	No	Yes	No	No, can't reject hypothesis but correlation is near-zero
Neuroticism	Presentation Score	-0.30	0.01	z = 1.69 p = 0.09	No	No	No	No, can't reject hypothesis but correlation is in wrong direction and trivial
Worry in achievement contexts	Presentation Score	-0.30	0.05	z = 1.90 p = 0.06	No	No	No	No, can't reject hypothesis but correlation is in wrong direction and near-zero
Emotionality in achievement contexts	Presentation Score	-0.30	0.10	z = 2.17 p = 0.03	Yes	No	No	No
Prior experience	Presentation Score	0.30	0.01	z = 1.58 p = 0.11	No	Yes	No	No, can't reject hypothesis but correlation is in

Variable	Variable 2	Hypothesized Correlation	Observed Correlation	Hypothesis Test (Compared against a fixed valued)	Significantly different from hypothesis?	Correlation in correct direction?	Correlation meaningful?	Hypothesis supported?
Task self-efficacy	Presentation Score	0.50 to 0.60	0.01	Against 0.50: $z = 2.80$ $p < 0.01$ Against 0.60: $z = 3.55$ $p < 0.01$	Yes	No	No	wrong direction and near-zero No

APPENDIX D. HIGH-QUALITY TRAFFIC CALMING PRESENTATION

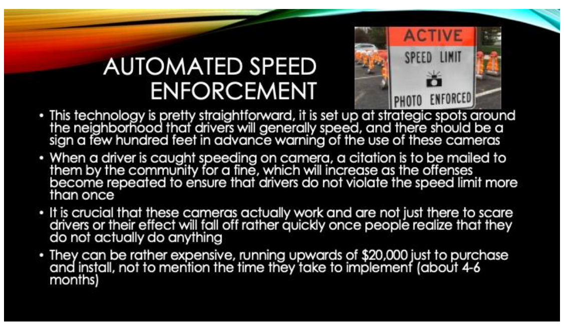
Slide 1



Slide 2

Slide 3

Slide 4



Slide 5

APPENDIX E. LOW-QUALITY TRAFFIC CALMING PRESENTATION

Slide 1

Neighborhood plan

Make a better/safer community

Slide 2

Add Sidewalks

- Add sidewalks to provide safety for pedestrians, especially the little kids
- Reduces crashes, mobility, and gives the neighborhood a healthier community
- Dimensions of sidewalk- about 48 inches so that two people can walk side by side
- Average cost- $4.72 per square foot

Slide 3

Signs

- Drop the speed limit by 5 mph
- Add other cautious signs in neighborhood
- Slow sign ($22)
- Caution children playing ($27)
- This would warn future drivers about kids near by and to keep their open

Slide 4

Possible neighborhood watch

- Anyone who wants to volunteer (no cost)
- Will provide safety to neighborhood and warn fast drivers to slow down

APPENDIX F. POST-HOC POWER ANALYSES

	n	Hypothesized correlation	Meaningfully different correlation	Power achieved	Sample size needed for a power of 0.80
Keychain & Traffic	57	0.60	0.40	0.64	88
Chair	31	0.60	0.40	0.42	88
Keychain & Traffic	57	0.50	0.30	0.55	110
Chair	31	0.50	0.30	0.36	110
Keychain & Traffic	59	0.49	0.29	0.56	112
Chair	32	0.49	0.29	0.36	112
Keychain & Traffic	59	0.35	0.15	0.49	137
Chair	32	0.35	0.15	0.32	137
Keychain & Traffic	59	0.34	0.14	0.48	138
Chair	32	0.34	0.14	0.32	138
Keychain & Traffic	59	0.30	0.10	0.47	144

Chair	32	0.30	0.10	0.31	144
Keychain & Traffic	59	0.20	0.00	0.45	153
Chair	32	0.20	0.00	0.30	153

REFERENCES

Ackerman, P. L. (1994). Intelligence, attention, and learning: Maximal and typical performance. *Current Topics in Human Intelligence*, *4*, 1-27.

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence*, *22*(2), 227-257.

Ackerman, P. L. (2000). Domain-specific knowledge as the" dark matter" of adult intelligence: Gf/Gc, personality and interest correlates. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *55*(2), P69-P84.

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist*, *38*(2), 85-93.

Ackerman, P. L. (2017). Adult intelligence: The construct and the criterion problem. *Perspectives on Psychological Science*, *12*(6), 987-998.

Ackerman, P. L., & Beier, M. E. (2012). The problem is in the definition: g and intelligence in I–O psychology. *Industrial and Organizational Psychology*, *5*(2), 149-153.

Ackerman, P. L., Chamorro-Premuzic, T., & Furnham, A. (2011). Trait complexes and academic achievement: Old and new ways of examining personality in educational contexts. *British Journal of Educational Psychology*, *81*(1), 27-40.

Ackerman, P. L., & Hambrick, D. Z. (2020). A primer on assessing intelligence in laboratory studies. *Intelligence*, *80*, 101440.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: evidence for overlapping traits. *Psychological Bulletin*, *121*(2), 219-245.

Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. J*ournal of Applied Psychology*, *78*(3), 413-432.

Ackerman, P. L., Kanfer, R., & Beier, M. E. (2013). Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, STEM persistence, and gender differences. *Journal of Educational Psychology*, *105*(3), 911-927.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, *1*(4), 270-304.

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, *13*(2), 57-78.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology*, *4*(3), 359-373.

Beck, L. F. (1933). The role of speed in intelligence. *Psychological Bulletin*, *30*(2), 169-178.

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

Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique, 11*, 191–244.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, *53*(4), 1207-1217.

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, *44*(1), 6-28.

Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, *21*(2), 180-192.

Bridgett, D. J., & Walker, M. E. (2006). Intellectual functioning in adults with ADHD: a meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychological Assessment*, *18*(1), 1-14.

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, *109*(2), 425-430.

Brunswik, E. (1952). The conceptual framework of psychology.(Int. Encycl. unified Sci., v. 1, no. 10.) Chicago: University of Chicago Press.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246-251.

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

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

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

Costa Jr, P. T., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, *55*(2), 258-265.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671-684.

Daniel, M. P., & Tversky, B. (2012). How to put things together. *Cognitive Processing*, *13*(4), 303-319.

Demming, J. A., & Pressey, S. L. (1957). Tests “indigenous” to the adult and older years. *Journal of Counseling Psychology*, *4*, 144–148.

Dwyer, P. S. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika, 2*(3), 173-178.

Ekstrom, R. B., French, J. W., Harman, H., & Derman, D. (1976). Kit of Factor-Referenced Cognitive Tests (revised edition). *Educational Testing Service*, Princeton, NJ.

Ellingsen, V. J. (2013). *Academic self-concept under typical and maximal environmental press* (Doctoral dissertation, Georgia Institute of Technology).

Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It’s all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology, 109*(5), 653-665.

Gorsuch, R. L. (1997). New procedure for extension analysis in exploratory factor analysis. *Educational and Psychological Measurement*, *57*(5), 725-740.

Goff, M., & Ackerman, P. L. (1992). Personality-intelligence relations: Assessment of typical intellectual engagement. *Journal of Educational Psychology, 8*4(4), 537-552.

Goldberg, L. R. (2008). International Personality Item Pool. http://ipip.ori.org.

Gottfredson, G. D., Jones, E. M., & Holland, J. L. (1993). Personality and vocational interests: The relation of Holland's six interest dimensions to five robust dimensions of personality. *Journal of Counseling Psychology, 40(4),* 518-524.

Heggestad, E. D., & Kanfer, R. (2000). Individual differences in trait motivation: Development of the Motivational Trait Questionnaire. *International Journal of Educational Research, 33*(7-8), 751-776.

Holland, J. L. (1963). Explorations of a theory of vocational choice and achievement: II. A four-year prediction study. *Psychological Reports*, *12*(2), 547-594.

Horn, J. L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research*, *8*(4), 477-489.

Jackson, D. N. (1998). Multidimensional aptitude battery-II. sigma assessment systems. *Port Huron, Michigan.*

Katz, I. R. (2007). Testing information literacy in digital environments: ETS's iSkills assessment. *Information Technology and Libraries*, *26*(3), 3-12.

Kanfer, R. (2012). Work Motivation: Theory, Practice, and Future Directions. In S.W.J. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology* (pp 455-495). New York, New York: Oxford University Press, Inc.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*(4), 657.

Kanfer, R., & Ackerman, P. (2000). Individual differences in work motivation: Further explorations of a trait framework. *Applied Psychology, 49*(3), 470-482.

Kanfer, R., Ackerman, P. L., & Heggestad, E. D. (1996). Motivational skills & self-regulation for learning: A trait perspective. *Learning and Individual Differences, 8*(3), 185-209.

Kanfer, R., & Heggestad, E. D. (1997). Motivational traits and skills: A person-centered approach to work motivation. Research in *Organizational Behavior 19*, 1-56.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155-163.

Kornilova, T. V., Kornilov, S. A., & Chumakova, M. A. (2009). Subjective evaluations of intelligence and academic self-concept predict academic achievement: Evidence from a selective student population. *Learning and Individual Differences, 19(4)*, 596-608.

Lamb, R. R., & Prediger, D. J. (1981). *Technical report for the Unisex Edition of the ACT Interest Inventory (UNIACT).* American College Testing Program.

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2006). Integration of multidimensional self-concept and core personality constructs: Construct validation and relations to well-being and achievement. *Journal of Personality*, *74*(2), 403-456.

McNemar, Q. (1940). A critical examination of the University of Iowa studies of environmental influences upon the IQ. *Psychological Bulletin*, *37*(2), 63-92.

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education*, *41*(3), 271-290.

Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale—Revised. *Psychological Assessment*, *5*(1), 34-40.

Moehring, A., Schroeders, U., Leichtmann, B., & Wilhelm, O. (2016). Ecological momentary assessment of digital literacy: Influence of fluid and crystallized intelligence, domain-specific knowledge, and computer usage. *Intelligence*, *59*, 170-180.

Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. P*ersonnel Psychology, 58*(3), 583-611.

Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging*, *8*(3), 389-399.

Novick, L. R., & Morse, D. L. (2000). Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition*, *28*(7), 1242-1256.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*(3), 228-242.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193-203.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39–58.

Randahl, G. J. (1991). A typological analysis of the relations between measured vocational interests and abilities. *Journal of Vocational Behavior*, *38*(3), 333-350.

Richardson, M., Jones, G., & Torrance, M. (2004). Identifying the task variables that influence perceived object assembly complexity. *Ergonomics*, *47*(9), 945-964.

Richardson, M., Jones, G., Torrance, M., & Baguley, T. (2006). Identifying the task variables that predict object assembly difficulty. *Human Factors*, *48*(3), 511-525.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology, 88*(1), 174-188.

Rosenberg, M. (1965). Rosenberg self-esteem scale (SES). *Society and the adolescent self-image.*

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. *Text relevance and learning from text*, 19-52.

Rouet, J. F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, *15*(1), 85-106.

Rowlands, M., & Mark, R. (1999). *The body in mind: Understanding cognitive processes*. Cambridge University Press.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy*, *101*(8), 389-428.

Salmerón, L., Cañas, J. J., Kintsch, W., & Fajardo, I. (2005). Reading strategies and hypertext comprehension. *Discourse Processes*, *40*(3), 171-191.

Schroeders, U., Bucholtz, N., Formazin, M., & Wilhelm, O. (2013). Modality specificity of comprehension abilities in the sciences. European Journal of Psychological Assessment.

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, *2*(1), 8-9.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162-173.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407-441.

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

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363-374.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2*(2-3), 191-210.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Lawrence Erlbaum.

Sternberg, R. J. (1980). Factor theories of intelligence are all right almost. *Educational Researcher*, *9*(8), 6-18.

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

Surber, J. R., & Schroeder, M. (2007). Effect of prior domain knowledge and headings on processing of informative text. *Contemporary Educational Psychology*, *32*(3), 485-498.

Technavio (2017). *Test preparation market in the US 2017-2021.*

Terman, L. M. (1916). *The measurement of intelligence. Boston, MA: Houghton Mifflin.*

Thorndike, E. L. (1924). Measurement of Intelligence. *Psychological Review, 31*(3), 219-252.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs.*

Thurstone, T. G. (1962). PMA. *Chicago: science research associates.*

Tucker, L. R., & Finkbeiner, C. T. (1981). Transformation of factors by artificial personal probability functions. *ETS Research Report Series*, (2), i-19.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist, 39*(2), 111-133.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler Adult Intelligence Scale. *Contemporary Educational Psychology.*

Wellman, B. L. (1940). Iowa studies on the effects of schooling. In G. M. Whipple (Ed.), *The thirty-ninth yearbook of the National Society for the Study of Education: Intelligence: Its nature and nurture. Part II. Original studies and experiments* (pp. 377–399). Bloomington, IL: Public School Publishing.

Wiedenbeck, S. (2005, October). Factors affecting the success of non-majors in learning to program. In *Proceedings of the first international workshop on Computing Education Research* (pp. 13-24).

Wiking, S., Brattfjell, M. L., Iversen, E. E., Malinowska, K., Mikkelsen, R. L., Røed, L. P., & Westgren, J. E. (2016). Sex Differences in Furniture Assembly Performance: An Experimental Study. *Applied Cognitive Psychology*, *30*(2), 226-233.

Wittmann, W. W. (1991). Meta-analysis and brunswik symmetry. *Social prevention and the social sciences: Theoretical controversies, research problems, and evaluation strategies*, 381-393.

Wittmann, W. W. (1995). The significance of Brunswik-Symmetry for psychological research and assessment. *European Journal of Psychological Assessment*, *11*(1), 59-60.

Wittmann, W.W. and Süß, H.-M. (1999) 'Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik Symmetry', in P.L.Ackerman, P.C.Kyllonen and R.D.Roberts (eds) Learning and Individual Differences: Process, Trait, and Content Determinants, pp. 77–108. Washington, DC: American Psychological Association.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*(1), 82-91.