

**From the holdings of the Department of Special Collections , Princeton University Library,
Princeton, NJ**

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyright law. We do not charge any permission or usage fees for the publication of images of material in our collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff. For more information regarding copyright, credit, and citation please review the information on the Copyright, Credit and Citation Guidelines page of our website.

<https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines>. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. <https://library.princeton.edu/special-collections/ask-us>

Doctoral Dissertations, 2011-2026

Psychology

2014

Determinants of Performance Across Domains and Within Bilingualism: Cognitive Abilities, Experiential Factors, and Predictability of the Task Environment

Macnamara, Brooke Noel

Princeton, NJ : Princeton University

<http://arks.princeton.edu/ark:/88435/dsp016t053g12z>

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. <https://www.copyright.gov/title17/title17.pdf>

**DETERMINANTS OF PERFORMANCE ACROSS DOMAINS AND
WITHIN BILINGUALISM:
COGNITIVE ABILITIES, EXPERIENTIAL FACTORS, AND
PREDICTABILITY OF THE TASK ENVIRONMENT**

Brooke Noel Macnamara

**A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY**

**RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF PSYCHOLOGY**

Adviser: Andrew R. A. Conway

June 2014

© Copyright 2014 by Brooke N. Macnamara. All rights reserved.

Acknowledgments

Michael Chow, Andrew Conway, Rae Drach, D. Zachary Hambrick, Judy Kegl, Adam Moore, and Frederick Oswald are second, third, or fourth authors of the studies presented in this document. Thank you to Jalisha Braxton, Eileen Forrestal, Robert Hills, LaTasha Holden, John Parvin, David Rivera, and Cynthia Williams for assisting with data collection or access to subject populations. Thank you to Neil Charness, Donald Chinn, Sarah Doumen, Joakim Ekstrand, Marije Elferink-Gemser, Mathias Haugaasen, Nicola Hodges, Rachelle Kamp, Thirayoot Limanond, Enrique Lopez, Gary McPherson, Joanne Ruthsatz, Igor Ryabov, Lydia Schaap, Steve Schultetus, Judy Sheard, Soren Svanum, and Marjorie Torenbeek for providing unpublished data, raw data associated with publications in order to calculate effect sizes, and/or clarification of methods or results for the meta-analysis. Thank you to Katherine Hansen, Deborah Holoien, and Kaite Yang for assisting with ratings. Thank you to John Yackel for measurement scoring, rating, and other assistance. Thank you to Andrew Conway for guidance with this research.

Abstract

How does one become an expert? The present research investigated relationships among general cognitive abilities, experiential factors, and performance variance. An initial study revealed that general cognitive abilities, and not experience, differentiated experts from non-experts in a complex bilingual task. Two longitudinal studies examining performance on the same task followed. The first longitudinal study investigated whether amount of training altered general cognitive abilities. Results indicated that general cognitive abilities recruited during performance in this task improved with experience. The purpose of the second longitudinal study was to investigate 1) individual differences in initial cognitive abilities as predictors of skill acquisition rates, 2) individual differences in cognitive ability change as predictors of skill acquisition rates, and 3) amount of training as a predictor of skill acquisition rates. The following studies were guided by a hypothesis that predictability of the task environment—the rate of change of the task environment while the performer is planning and executing an action, and the range of probable stimuli to which the performer may respond—moderates the relationship between performance variance and its determinants. That is, as predictability of the task environment decreases, the explanatory power of experiential factors decrease and the relative explanatory power of general cognitive abilities increase. For this reason, according to the hypothesis, experiential factors demonstrated a weak direct relationship with performance in the complex bilingual task. In order to test this hypothesis, a laboratory task was designed in which the two features of predictability—rate of change and range of probable stimuli—were manipulated. The results supported the predictability hypothesis: amount of practice was a stronger determinant of performance than general cognitive abilities when the task had low predictability; the opposite

pattern emerged when the task had high predictability. Finally, in order to test whether this hypothesis translated into real-world domains, a meta-analysis was conducted examining the relationship between practice and performance variance. The meta-analysis revealed that predictability of the task environment significantly moderated the relationship, such that the lower the predictability, the weaker the relationship between practice and performance. This work advances our understanding of skill acquisition and expertise determinants.

Contents

Acknowledgements	iii
Abstract	iv
Contents	vi
1 Individual differences: Cognitive abilities, experience, and performance	1
1.1 Chapter summary	1
1.2 Two classic views of expertise	2
1.3 Current theories of performance variance	3
1.4 Cognitive abilities as flexible variables	6
1.5 Overview of Chapters 2 – 7	7
2 Predictors of expertise in a complex bilingual processing task	12
2.1 Chapter summary	12
2.2 Introduction: The task of simultaneous interpreting	13
2.3 Methods	18
2.4 Results	24
2.5 Discussion	29
3 Cognitive ability change from complex bilingual processing experience	33
3.1 Chapter summary	33
3.2 Introduction: The bilingual advantage	34
3.3 Methods	38
3.4 Results	41
3.5 Discussion	44

4	Predictors of skill acquisition in a complex bilingual processing task	47
4.1	Chapter summary	47
4.2	Introduction: Longitudinal measurement of skill	48
4.3	Methods	49
4.4	Results	58
4.5	Discussion	73
5	A novel moderator variable: Predictability of the task environment	75
5.1	Chapter summary	75
5.2	Introduction: A laboratory measure of skill acquisition	76
5.3	Predictability Experiment 1	83
5.3.1	Methods	83
5.3.2	Results and Discussion	89
5.4	Predictability Experiment 2	92
5.4.1	Methods	92
5.4.2	Results and Discussion	95
5.5	General Discussion	107
6	The role of predictability of the task environment in real-world domains	109
6.1	Chapter summary	109
6.2	Introduction: A meta-analysis of deliberate practice and performance	110
6.3	Methods	114
6.4	Results	119
6.5	Discussion	124

7 Conclusion and future directions	129
7.1 Chapter summary	129
7.2 Introduction	130
7.3 Overview of empirical findings	130
7.4 Contributions and implications	134
7.5 Future directions and conclusion	137
References	141

Chapter 1

Individual differences: Cognitive abilities, experience, and performance

1.1 Chapter summary

This introductory chapter reviews a range of theories on predictors of expertise as well as the evidence for cognitive ability flexibility, and then provides an overview of the remaining document.¹

- Classic views of expertise focus either on hereditary predictors—“nature”—or experiential predictors—“nurture.”
- Current theories within cognitive psychology reflect, to an extent, this historical dichotomy.
- Recent research suggests that certain types of experience enhance general cognitive abilities—traits considered largely heritable.
- Directional relationships among cognitive abilities, experience, and performance remain controversial. An important next step is to investigate predictors in under-studied domains and test potential moderators of these relationships across domains to explain variance in findings.

¹ Parts of this introduction can be found in Macnamara, Hambrick, & Oswald, 2014.

1.2 Two classic views of expertise

Why do some individuals who practice the violin, or attempt to become fluent in French, or enter an academic setting succeed, while others do not? Scholars have been interested in answering this question for well over a century. Two classic and opposing views of human performance are a) experts are “born” and b) experts are “made.” The most frequent illustrations of these views come from Sir Francis Galton and John Watson.

Sir Francis Galton (1869), the founder of behavioral genetics who coined the terms *nature* and *nurture* to describe hereditary and environmental influences, theorized that experts were “born.” That is, that training and practice are necessary to reach a high level of performance, but that innate abilities limit the ultimate level of performance an individual can achieve. This conclusion followed from Galton’s examinations of the genealogical records of eminent scientists, musicians, writers, artists, and athletes and the observation that that they tended to be biologically related.

John Watson (1930), the founder of behaviorism, theorized that experts are “made.” That is, that both hereditary and environmental factors influence performance, but that training and practice were the best explanation for success across domains, including “genius.” This idea is captured in his famous quote

Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents... (p. 104).

Sir Francis Galton and John Watson assumed that both innate abilities and experiential factors contributed to variance in performance, but differed in which they felt was the more important predictor. Despite their relative similarity in perspective, Galton and Watson are

frequently referenced as examples of two opposing views, 1) that nature-based variables best explain performance variance, and 2) that nurture-based variables best explain performance variance.

1.3 Current theories of performance variance

The two classic views of expertise have continued to influence modern accounts of performance variance. Two branches of cognitive psychology a) individual differences, and b) skill acquisition and expertise, both revolve around investigations of performance variance, yet frequently yield little interaction and influence on one another as a whole. While there is some overlap and a few researchers who consider themselves scholars of both research areas, these two sub-disciplines generally reflect the extant nature-nurture/born-made dichotomy.

Individual differences researchers frequently examine general cognitive abilities such as working memory capacity and intelligence, and typically assume that these abilities reflect stable characteristics (Arbuckle, Maag, Pushkar, & Chaikelson, 1998; Davies et al., 2011; Hertzog & Schaie 1986; 1988; Wicherts et al. 2004; Waters & Caplan 2003; Zelinski & Stewart, 1998; however, see Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Klingberg, 2010, and Morrison & Chein, 2011 for recent studies on working memory training, and Macnamara, Chapter 3 this document for a longitudinal study on the bilingual advantage), often suggesting innateness since, in addition to stability, they emerge early in development (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Starkey 1992). Commonly used measures of cognitive abilities are designed to tap only general abilities, that is, most measures of this kind are created so that specific experience should have little influence on the results.

Underlying mechanisms of and relationships among general cognitive abilities are investigated in this area in order to elucidate the role of these abilities in a variety of functions; general cognitive abilities are considered important for predicting performance in any domain that recruits such abilities (Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Cowan, 1995; Engle, Tuholski, Laughlin, & Conway, 1999; Gottfredson, 1997; Hambrick & Engle, 2002; Lovett, Reder, & Lebiere, 1999). Most notably, individual differences in general cognitive abilities have demonstrated reliable and robust correlations with variance in academic achievement and occupational performance (Daneman & Carpenter, 1980; Deary, Strand, Smith, & Fernandes, 2007; S. B. Kaufman, Reynolds, Liu, A.S. Kaufman, & McGrew, 2012; Naglieri & Bornstein, 2003; Rohde & Thompson, 2007; Watkins, Lei, & Canivez, 2007).

In contrast to innate abilities as performance predictors, skill acquisition and expertise researchers frequently examine training protocols and practice regimens as predictors of performance across domains. Skill acquisition and expertise researchers typically assume that accumulated practice and other environmental factors best explain variance in performance. Most notably, differences in amount of accumulated practice have demonstrated reliable and robust correlations with variance in chess ratings and musical performance (Charness, 1976; Charness, Krampe, & Mayr, 1996; Chi, Glaser, & Farr, 1998; Ericsson, 2007; Ericsson, Krampe, & Tesch-Römer, 1993; Ericsson & Lehmann, 1996; Hoffman, 1992; Simon & Chase, 1973).

In a similar vein is the examination of the relationship between domain-specific experience and domain-specific cognitive abilities (that is, cognitive skills associated with performing a specific task). This research explores how extended domain-specific experience enhances perceptual speed, knowledge access efficiency, and memory capacity for domain-specific stimuli (Ericsson & Kintsch, 1995; Staszewski, 1988). These domain-specific cognitive

abilities are assumed to be independent from general cognitive abilities. For example, as a partial replication of earlier work (A. de Groot 1946/1978), Simon and Chase (1973) compared expert and non-expert chess players. They found that when chess pieces were arranged randomly on a chessboard, the two groups could remember similar numbers of chess piece locations; this memory for random positions was assumed to reflect general working memory capacity. However, when Simon and Chase arranged the chess pieces on the board representative of a game in progress, that is, the pieces were arranged to reflect the structure of the game, the experts could remember substantially more chess pieces than the non-experts. Findings such as this one are taken to support the notion that general cognitive abilities do not differentiate experts from non-experts, but that extended domain-specific experience improves performers' speed and efficiency in processing domain-relevant information, which improves their overall performance in the domain (Chase & Ericsson, 1981, 1982; Ericsson & Kintsch, 1995; Staszewski, 1988).

A major focus of this document is examining when general cognitive abilities are more predictive of performance than experiential factors and when experiential factors are more predictive of performance than general cognitive abilities. The hypothesis tested is that the predictability of the task environment—the rate of change of the task environment while the performer is planning and executing an action, and the range of probable stimuli to which the performer may respond—moderates these relationships (see Chapter 5). That is, when a task is highly predictable, experience with the task will be a stronger predictor of performance than general cognitive abilities. When the task has low predictability, cognitive abilities that allow one assess the novel demands of the task and adapt to the demands will be more predictive than experience with the task, as experience will not have prepared the performer for novel scenarios.

For example, experience cigar rolling or mirror tracing, highly predictable tasks, will improve cigar rolling performance (Crossman, 1959) and mirror tracing performance (Snoddy, 1926) respectively. Individual differences in cognitive abilities will have less impact on mirror tracing performance than experience, as demonstrated by patient HM's improved performance over time even without memory for performing the task (Milner, 1962). In contrast, reading performance, a semi-predictable task, is predicted by both genetic and environmental factors (Plomin, 2013). Tasks with very low predictability, such as simultaneous language interpreting, have been studied less frequently. The predictability of the task environment hypothesis suggests that experience after the initial learning phase will have very little impact on performance while individual differences in domain-general cognitive abilities will better explain performance variance. In other words, an interpreter with many years of experience could perform the task less well than an interpreter with a few years of experience.

1.4 Cognitive abilities as flexible variables

Another relevant area within cognitive psychology examines the relationship between a particular domain-specific experience—bilingual experience—and enhancement to general cognitive abilities. The *bilingual advantage* refers to the finding that bilinguals tend to outperform monolinguals on certain measures of cognitive control. The most common hypothesis for the source of this finding is that frequent engagement of executive control functions necessary for controlling target language selection and resolving conflict between

competing language alternatives enhances executive control (Bialystok, 1999; Bialystok & Craik, 2010; D. W. Green, 1998)².

Like individual differences researchers, investigators of the bilingual advantage study general cognitive abilities. However, unlike typical research conducted from an individual differences' perspective, that is, typical cognitive ability research, cognitive ability variation is investigated as an outcome variable and not as a predictor. The theory also necessitates the assumption that specific types of experience influence cognitive abilities.

In parallel, like skill acquisition and expertise researchers, investigators of the bilingual advantage study the relationship between domain-specific experience and cognitive processes. However, unlike typical research conducted from a skill acquisition and expertise perspective, the focus is on general cognitive abilities, not domain-specific cognitive processes. Also, unlike typical research conducted from an individual differences perspective, bilingual advantage research focuses on general cognitive abilities as the outcome variables, not the predictor variables.

1.5 Overview of Chapters 2 – 7

Individual differences research (i.e., general cognitive ability research), skill acquisition and expertise research, and bilingual advantage research each examine portions of the complex relationships among cognitive abilities, domain-specific experience, and performance variance

² This theory is not unique to bilingual research. For example, investigations comparing avid video game players to non-video game players reveal differences in cognitive abilities assumed to be recruited during video game play; avid players outperform non-players on general measures of attentional control and spatial reasoning (e.g., Feng, Spence, & Pratt, 2007; C. S. Green, & Bavelier, 2003).

(see Figure 1.1). The goal of this dissertation is to study these variables and their inter-relationships within the context of other related variables and relationships.

Chapter 2 presents a field study investigating predictors of performance in a complex bilingual task. Measures of general cognitive abilities and domain-specific experience reveal that there are no differences between experts and non-experts in years of experience or training. However, a statistical model that incorporates individuals' scores on general cognitive abilities and a personality trait reliably discriminates experts from non-experts. The results of this study support the notion that general cognitive abilities, but not domain-specific experience, are important predictors of performance in a complex bilingual task.

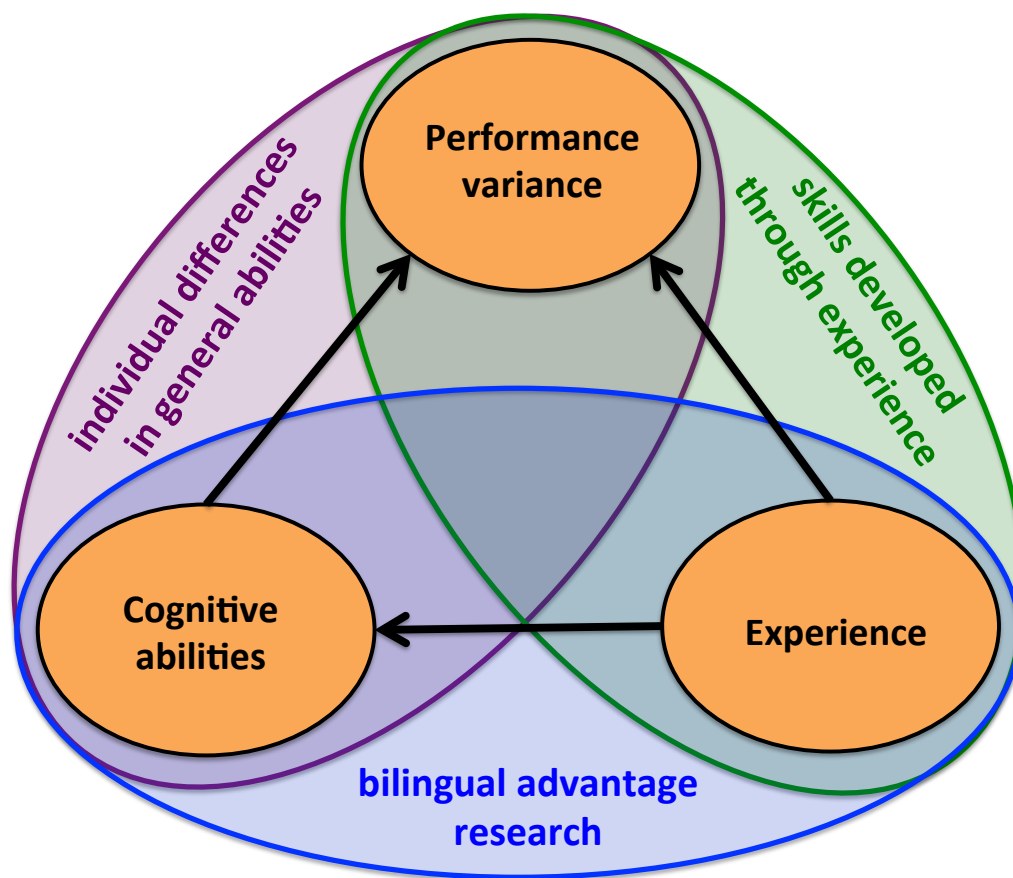


Figure 1.1. Schematic of three research areas, their variables of interest, and their inter-relationships.

Chapter 3 presents a longitudinal field study investigating the bilingual advantage in the same complex bilingual task of interest used in Chapter 2. Measures of general cognitive abilities thought to be recruited during performance of the complex bilingual task improved with domain-specific experience; measures of general cognitive abilities not thought to be recruited during performance of the task did not improve over time. The results of this study suggest that general cognitive abilities are typically stable, but that under conditions of high executive function recruitment and extended domain-specific experience, general cognitive abilities can improve.

Chapter 4 presents a longitudinal field study investigating skill acquisition in the same complex bilingual task used in Chapters 2 and 3. Measures of general cognitive abilities and task performance were administered at approximately 20%, 40%, 60%, and 80% completion through a 4-semester, domain-specific training program. Cognitive abilities became enhanced through training, corroborating the results of Chapter 3. General cognitive abilities also predicted interpreting performance, corroborating the results of Chapter 2. However, accumulated amount of domain-specific experience also predicted interpreting performance.

The study presented in Chapter 5 was guided by a hypothesis that predictability of the task environment—the rate of change of the task environment while the performer is planning and executing an action, and the range of probable stimuli to which the performer may respond—moderates the relationship between performance variance and its determinants. That is, as predictability of the task environment decreases, the explanatory power of experiential factors decrease and the relative explanatory power of general cognitive abilities increase. For this reason, according to the hypothesis, experiential factors demonstrated a weak direct relationship with high levels of performance in the complex bilingual task. In order to test this

hypothesis, a laboratory task was designed in which the two features of predictability—rate of change and range of probable stimuli—were manipulated. The results supported the predictability hypothesis: accumulated amount of practice was a strong determinant of performance when the task was highly predictable. Neither accumulated practice nor cognitive abilities determined performance when the task was not predictable. General cognitive abilities were significant determinants of performance in semi-predictable task environments.

Chapter 6 presents a meta-analysis on the explanatory power of deliberate practice (Ericsson, Krampe, & Tesch-Romer, 1993)—a highly popular account of skill acquisition and expertise—on performance variance. The meta-analysis also tests the hypothesis that predictability of the task environment moderates the relationship between practice and performance in real-world tasks across five domains—games, sports, music, education, and professions. A major finding was that deliberate practice explained only a small percentage of performance variance, leaving a much larger amount unexplained and potentially explainable by other factors. The explanatory power varied significantly by domain, as well as predictability of the task environment. On average, the lower the predictability of the task environment, the weaker the relationship between practice and performance.

Chapter 7 presents conclusions from the five studies and implications of these results on research investigating abilities as predictors of performance, research investigating skill, and bilingual advantage research. Following the conclusions, several future directions are discussed, including studies aimed at investigating directional relationships not previously studied in these contexts (e.g., the effect of [perceived] performance level on amount of deliberate practice).

This work represents a novel approach to studying general cognitive abilities, domain-specific experience, and variance in performance. Researchers who investigate the role of

abilities on performance (individual differences researchers) often consider domain-specific experience a confound; researchers who investigate domain-specific experience on variance in performance (skill acquisition and expertise researchers) often consider initial cognitive abilities noise; bilingual advantage researchers investigate domain-specific experience on general cognitive abilities and often do not consider levels of expertise. In contrast, the work presented in this document assumes that these variables should not be considered in isolation, and that a more promising approach is to examine the interrelationships of these variables and the contextual variables that moderate them.

Chapter 2

Predictors of expertise in a complex bilingual processing task

2.1 Chapter summary

A field study compares expert and non-expert professional American Sign Language-English simultaneous interpreters.³

- Simultaneous interpreting is a highly demanding bilingual task.
- Professional simultaneous interpreters were recruited, demographic/experiential variables were collected, and a battery of general cognitive ability and personality traits were measured.
- There was no difference between the groups in years of professional interpreting experience.
- A combination of general cognitive abilities reliably differentiated experts from non-experts.

³ This study can be found in a published report (Macnamara, Moore, Kegl, & Conway, 2011). Also presented at the Aptitude for Interpreting: Towards Reliable Admission Testing symposium (2014).

2.2 Introduction: The task of simultaneous interpreting

Simultaneous interpreting is an extremely demanding cognitive processing task (Christoffels, A. M. B. de Groot, Waldorp, 2003; Frauenfelder & Schriefers 1997; MacWhinney 1997). The interpreter must listen and comprehend the source message in one language while concurrently planning the closest equivalents and producing the message in another language with little to no control over the input rate or content. These tasks require alternately activating and suppressing production the two languages (Christoffels et al., 2003), analyzing the speaker's goals, inferences, and subtleties (Cowan, 2000; Moser 1978; Seal, 2004; Shlesinger 2003), and making rapid decisions on how to convey the meaning in a different language while taking into account cultural nuances and communication rules (Christoffels et al. 2003; Treisman, 1965).

Is the expert interpreter “born” or “made?” While it is generally accepted that domain-specific experience enhances one's performance in any domain, at least during the initial learning phase, the role of domain-general cognitive abilities as a predictor of expertise is less clear. Given the aforementioned interpreting demands—comprehension, planning, task switching, and reasoning—which presumably recruit general cognitive abilities, individual differences in general cognitive abilities should be highly predictive of overall interpreting performance.

Gerver, Longley, Long, & Lambert, (1989) studied differences between passing and failing interpreter students on a variety of discourse processing and verbal abilities. Their results indicated that passing interpreter students had better memory for texts, comprehension, and verbal generation than their failing counterparts. Discourse-processing abilities such as those

measured by Gerver and colleagues are mediated by working memory and more domain-general cognitive skills (Gernsbacher 1990; Just & Carpenter 1992; Kintsch 1988; Kintsch & van Dijk 1978). In other words, domain-general ability can predict performance differences in these language tasks over and above what can be explained by linguistic skill and domain-specific training alone.

Despite the plethora of evidence for the predictive power of domain-general cognitive abilities in psychological research, evidence as to the relationship between these types of measures and simultaneous interpreting performance has been mixed. Studies have shown positive correlations between general memory abilities and interpreter trainee performance (López Gómez, Bajo Molina, Padilla Benítez, & Santiago de Torres, 2007), that professional interpreters measure higher in working memory capacity and reasoning than non-interpreters (Christoffels, A. M. B. de Groot, & Kroll, 2006; P. Padilla, Bajo, Cañas, & F. Padilla, 1995; Rudser & Strong 1986; Seal, 2004), but also that reasoning ability does not significantly predict performance among professional interpreters (Rudser & Strong 1986; Seal, 2004), and a finding that novice interpreters measure higher in working memory capacity than expert interpreters (Köpke & Nespoulous, 2006).

The discrepancies observed in previous research may be due to several factors. First, the limited nature of the population studied may compromise the comparability of the experimental subjects in the respective studies and/or the statistical analysis may be underpowered to detect meaningful effects. Second, differences that are found between interpreters and non-interpreters in cross-sectional designs could be due either to changes in cognitive ability as a result of interpreting experience or self-selection of individuals with certain cognitive abilities into the interpreting field. Third, the tasks used to measure cognitive abilities of interest are not all

equally valid or reliable. Fourth, some cognitive abilities/traits will make no difference in one's future interpreting skill; some may relate to interpreting skill only up to a certain point (i.e. the first few months/years of training), while others may still correlate with interpreting performance regardless of experience or training. Finally, and perhaps most importantly, performance in complex real world interactions depends on the interplay of many cognitive abilities, so that even those studies that have measured multiple abilities/traits but only analyzed first-order relationships between each measure and population type or skill level will have missed the interactive contingencies that are certain to exist.

This is a long series of challenges, and no one study can meet them all. The goal of the present study is to focus centrally on the last point as a first step; through the use of tasks that have excellent psychometric validity and reliability coupled with advanced statistical analysis, in order to assess which combination of general cognitive abilities and emotion-cognition interaction traits predict interpreting expertise. Given the wide variety of standards for the level of skill required for graduation and professional practice among American Sign Language-English interpreters, it is possible to study groups of simultaneous interpreters who have similar interpreting experience and training, but who vary in interpreting skill.

Descriptive cognitive process models illustrating the complexity of simultaneous interpreting began to appear in the 1970s (e.g., Gerver, 1976). As both cognitive psychology and simultaneous interpreting research emerged, our understanding of the specific cognitive mechanisms engaged during simultaneous interpreting developed (cf. e.g., Moser, 1978) and today we can reasonably argue for the involvement of several key cognitive processing abilities and personality traits known to impact the constituent processes of simultaneous interpreting. This paper specifically addresses the following cognitive abilities: reasoning, working memory

capacity, processing speed, psychomotor speed, cognitive control, and mental flexibility.

Reasoning is essential for linguistic, environmental, and affective analysis of the source message for comprehension and prediction (Cokely 1992; Colonomos 1997, 2008; Macnamara 2009; Moser 1978) as well as planning the target message output (Cokely, 1992; Colonomos, 1997, 2008; Macnamara, 2009). Working memory capacity, one's limit of information that can be stored while simultaneously carrying out a processing task, is positively correlated with language comprehension (Daneman & Carpenter, 1980) and discourse processing ability (Just & Carpenter 1992; Kintsch, 1988; Kintsch & van Dijk, 1978; Gernsbacher, 1990), essential components of simultaneous interpreting. In addition to language processing, working memory is also critically involved in real-time problem-solving, reasoning, and planning, as well as manipulating or transforming incoming information — all of these abilities are certainly needed during simultaneous interpreting.

Simultaneous interpreting demands that information be processed rapidly. High working memory capacity and robust reasoning ability are not useful to interpreters if they cannot process the incoming information and execute decisions at a rate faster than or consistent with the incoming information. For this reason, both measures of processing (perceptual) speed and psychomotor speed (given that one of the interpreters' languages was manual) were measured. The rapid switching among subcomponents of interpreting—comprehending the source message, determining meaning equivalences, planning the production, and producing the target message (Cokely 1992; Colonomos 1997; Moser 1978)—engage cognitive control, which is necessary for switching among tasks and to manage task execution effectively (Monsell, 2003).

Interpreting, like other practice professions (e.g., medicine, teaching, counseling, law, and investigation) require technical knowledge and skills, but also require assessments of ever

changing situational and human interaction factors that impact how the technical knowledge and skills should be implemented in each situation (Dean & Pollard, 2005). Interpreters cannot perfectly predict the incoming message and will rarely interpret the same source message more than once. Decisions made while interpreting are adjusted for constantly changing situations as a function of this indeterminism. Interpreters rely on adaptive responses when handling the incoming message and other human interaction factors. One's capacity to adaptively coordinate actions in relation to others' actions in interpreting relies on mental flexibility, the final cognitive ability measured in this study.

Cognitive abilities, however, do not exist in a vacuum. This paper therefore also considers specific personality traits known to impact cognitive processing: willingness to engage in complex cognitive tasks, reward sensitivity, and risk sensitivity. One's general cognitive abilities may interact with one's willingness to employ mental resources during the task, affecting performance (Cacioppo & Petty, 1982). In addition to their willingness to engage in cognitive tasks, individuals vary in their motivation to engage cognitive control and make decisions based on sensitivities to the potential outcomes. To this end, individuals vary in their sensitivity to reward and risk (e.g., Gray 1982, 1987, 1990; Lopes 1987; Schneider & Lopes 1986). Individuals' sensitivity to reward motivates behavior toward subjectively positive outcomes, and increases the likelihood of decisions designed to approach desirable goals (Gray 1982). On the other hand, individuals with high risk sensitivity experience anxiety when presented with potential threat, non-reward, or novelty, and will structure their behavior around avoiding risk of aversive outcomes, as opposed to achieving positive outcomes (Gray 1982, 1987, 1990). Individuals with high anxiety will experience reduced working memory capacity (Eysenck, 1979, 1985; Leon & Revelle, 1985; Schmader & Johns, 2003; Wine, 1971) and are

likelier to respond more quickly to stimuli and produce more errors (Leon & Revelle 1985). Simultaneous interpreting demands that interpreters process novel, unrehearsed stimuli and adjust the time between comprehension and production based on the incoming message and interpreting abilities. An interpreter's risk sensitivity may therefore affect how quickly she or he responds to the incoming stimuli and the subsequent rate of errors.

The cognitive abilities and emotion cognition interaction traits relevant to interpreting included here are not meant to be exhaustive. Rather, the goal of this initial study is to test whether any of these abilities/traits can predict interpreter expertise, and if so, which combination of these abilities and traits best differentiates skill level.

2.3 Methods

Participants and the Rating Process

Twenty-nine professional ASL-English interpreters from two diverse locations in the United States participated in the study. Ages ranged from 20 to 60 ($M = 42$, $SD = 12$). Interpreting experience ranged from 6 months to 35 years ($M = 11$, $SD = 11$). Participants were not recruited on the basis of age or years of professional experience. Participants had similar interpreting training.

Five raters classified interpreters based on general interpreting competency. The raters were ASL-English interpreters who were familiar with the simultaneous interpreting work of the participants in a variety of settings within the prior year, and with a variety of consumers. These observations were the basis of their ratings. Participants did not undergo a standardized laboratory-based competency test, which would allow for a finer-grained level of performance analysis, for two reasons. First, given the high business demand for professional interpreters'

time (regardless of skill level as most hiring parties are not equipped to evaluate skill level), this method was used in order to recruit as many professional interpreters as possible: an additional test would increase the experimental session time and potentially decrease the number of interpreters willing to participate. Second, a laboratory-based competency test reduces ecological validity. For example, such a test does not easily measure how well one can adapt to the communication demands of the setting. All of the raters were trained interpreter evaluators. Competency was defined as performing well in skill-based aspects of the interpreting process including comprehension, ASL production, message equivalence, and the ability to perform appropriately to the situation (e.g. adapt to communication demands of the setting and consumers). Interpreters were not rated on ethical behavior, maturity, attitude toward consumers or colleagues, or other professional behavior.

The raters scored the interpreters on a three-point scale as (1) highly skilled, (2) somewhat skilled, and (3) less skilled. The two raters for the first location had very high inter-rater reliability, Cronbach's $\alpha = .978$. One rater was not familiar with two participants' interpreting skill and the other rater was not familiar with two different participants' interpreting skill. Because the inter-rater reliability was highly correlated between the two raters, the rating from the familiar rater was accepted. The two raters at the second location had very high inter-rater reliability, Cronbach's $\alpha = .935$. However, both raters were unfamiliar with one participant's interpreting skill. A third rater was secured who was familiar with the skill level of the participant in question. The third rater demonstrated high inter-rater reliability with both the other raters, Cronbach's $\alpha = .935$ and 1.0, respectively. The third rater's rating of the participant in question was accepted. (The inter-rater reliability for the three raters at the second location combined, was Cronbach's $\alpha = .958$.)

Since there were only two cases in which a rater rated a participant as somewhat skilled (the only two in which raters differed), and since inter-rater reliability was extremely high, the rating of the other rater was accepted in both cases. For the first instance in the first location, the participant was rated by one rater as in-between the two skill groups and by the second rater as less skilled — and was placed in the less-skilled group. For the second case in the second location, the participant was rated by one rater as in-between the two skill groups and by the other two raters as less skilled — and was placed in the less-skilled group. Due to the generally dichotomous ratings, as participants were generally rated as either highly skilled or less skilled and the few instances of somewhat skilled ratings were the only ones not agreed upon, the middle group was removed and the two groups were termed expert and non-expert. There were no instances of polar classifications (one rater placing a participant in the less skilled group and another placing the same participant in the highly skilled group). The expert group consisted of 15 interpreters and the non-expert group consisted of 14 interpreters.

Measurements

After recording age and years of professional ASL-English interpreting experience, seven cognitive ability measurements and three emotion-cognition interaction measurements (specific personality dimensions) were administered to participants. (See Table 2.1 for a summary of the measurements.)

Cognitive Ability Measurements

Raven's Advance Progressive Matrices. Raven's Advanced Progressive Matrices (referred to as "Ravens" in this document; Raven, 1962) is a multiple-choice abstract reasoning test designed to measure general fluid intelligence and reasoning in above-average intelligence individuals. Problems consist of a 3 X 3 matrix where each element is a pattern that shares some

features with adjacent elements. The lower right element is missing, and participants are asked to choose which of the available choices best completes the pattern. Participants had ten minutes to complete as many of the 18 increasingly difficult matrices as possible.⁴

Symmetry Span. Symmetry Span (Unsworth et al., 2005) is one of several “complex span tasks” that measure working memory capacity by measuring item recall in the face of interference. Complex span tasks are commonly used to measure working memory capacity. Participants are tasked with performing the processing components of the task as well as retaining as many of the memoranda as possible. In symmetry span, participants make judgments about the symmetry of abstract figures along the vertical axis that are interleaved with the presentation of a colored square on a 4 X 4 grid. Participants are asked to remember the position of the colored squares. Each trial consists of 2-6 symmetry decision-colored square presentations and then a recall screen in which individuals attempt to recall the location and order of the colored squares.

Letter Comparison. The Letter Comparison task (Salthouse & Babcock, 1991) assesses cognitive processing speed by measuring perceptual and decision-making speed through comparison of orthographic patterns. Participants are presented with 8 1/2 X 11-inch pages with pairs of non-lexical letter strings. Participants must compare the two items in the pair and mark whether they are the same or different. They complete as many pairs as possible in 30 seconds (per page.) Two pages with different letters are presented for each condition.

Pattern Comparison. The Pattern Comparison task (Salthouse & Babcock, 1991) is the same as the Letter Comparison task except instead of non-lexical letter strings Pattern Comparison uses abstract visual figures.

⁴ Only odd problems were used. Previous research has shown the odd-even split-half corrected reliability coefficient is .96 (Burke, 1972).

Connections Test A. Connections A (Salthouse et al., 2000) measures psychomotor speed with two conditions: numbers and letters. Participants are presented with 8 1/2 X 11-inch pages with 49 circles containing either all numbers or all letters. Participants connect the numbers or letters in sequence with a pen. The numbers or letters are not presented on the page in sequence, but a sequential letter or number is always adjacent (in any direction) to the former letter or number. Participants are allowed 20 seconds per page.

Connections B. Connections B (Salthouse et al., 2000) measures cognitive control, task switching, and psychomotor speed. Participants are presented with materials similar to those in Connections A, but must connect numbers *and* letters in alternating sequence. Connections B has two conditions: numbers-letters (i.e., 1 → A → 2 → B → 3 → C, etc.) and letters-numbers (i.e., A → 1 → B → 2 → C → 3, etc.). Scoring and time allowances for Connections A and B are the same.

Wisconsin Card Sorting Test. The Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948)) measures set shifting—the ability to adapt to changes and to shift among different sets of rules—i.e., mental flexibility. Participants view the images of four decks of cards face-up on the screen and are asked to sort new cards into one of the four piles based on one of three dimensional rules, that is, sorting based on the color, the number, or the shape of the symbol(s) on the card. Participants are not informed of the correct sorting rule, but are informed after each decision whether the sort choice was correct or incorrect. They must then infer the sorting rule from the feedback provided. Without warning, the rule changes during this task and the participants must discard the old rule and infer a new one based on the feedback.

Emotion-Cognition Interaction Measurements

BAS scale. The BAS scale (Carver & White, 1994) measures reward sensitivity, drive, and fun-seeking traits. BAS reflects approach orientation—when the goal is to move toward something desired. Individuals who score high on the BAS scale are more sensitive to positive rewards and personal enjoyment, and are more likely to actively pursue activities that yield such rewards than are those who score low on the scale. Participants respond on paper to a series of statements with a Likert-type scale, choosing whether each statement is “very true for me”, “somewhat true for me”, “somewhat false for me”, or “very false for me.”

BIS scale. The BIS scale (Carver & White, 1994) measures risk-taking sensitivity and anxiety surrounding aversive stimuli and novelty. Statements measuring BIS are interleaved statements measuring BAS. Participants respond to BIS scale statements exactly as they respond to BAS scale statements.

Need for Cognition Scale. The Need for Cognition scale (Cacioppo & Petty, 1982) measures an individual’s intrinsic motivation to engage in complex, cognitively demanding tasks. Participants respond on paper to a series of statements with a Likert-type scale, choosing whether in their view each statement is “completely true”, “mostly true”, “mostly false”, or “completely false.”

Table 2.1

Measurements

Task	Ability/trait measured
	Cognitive abilities
Ravens	fluid intelligence; reasoning
Symmetry Span	working memory capacity (maintenance + interference)
Letter Comparison	cognitive processing speed (perceptual speed)
Pattern Comparison	cognitive processing speed (perceptual speed)
Connections A	psychomotor speed
Connections B	psychomotor speed; cognitive control; task switching
WCST	mental flexibility; adaptability; task switching
	Emotion-cognition interaction (personality) traits
Need for Cognition scale	willingness to engage in complex cognitive tasks
BAS scale	sensitivity to reward / subjectively positive outcomes
BIS scale	sensitivity to risk / subjectively negative outcomes

Note. Ravens = Ravens Advanced Progressive Matrices. WCST = Wisconsin Card Sorting Test. BAS = Behavioral Approach System. BIS = Behavioral Inhibition Scale

Procedure

Participants were tested individually for approximately one hour and ten minutes and were paid for participation. Participants completed the tasks in the following order: 1) BIS/BAS scales, 2) Need for Cognition scale, 3) Letter Comparison, 4) Pattern Comparison, 5) Connections Tests (A and B alternated, the standard procedure for administering the Connections Tests), 6) Ravens, 7) Symmetry Span, and 8) the Wisconsin Card Sorting Test.

2.4 Results

Demographic Variables

There were no significant differences between the experts and non-experts in age (experts: $M = 42$, $SD = 13$; non-experts: $M = 42$, $SD = 10$), $F < 1$. There were also no significant

differences between the two groups in years of professional interpreting experience (experts $M=12$, $SD=11$, non-experts $M=10$, $SD=10$), $F < 1$.

Cognitive Ability and Emotion-Cognition Interaction Trait

Variables

Univariate analyses

Prior to statistical analysis, all measurement scores were examined for accuracy of data entry, missing values, and normality of distribution. Missing values occurred for two subjects for the questionnaires (BIS/BAS scales and the Need for Cognition scale) and one subject for the WCST (due to timing issues during data collection.) There were no univariate outliers and all distributions were normal. (See Table 2.2 for descriptive statistics.)

One-way analyses of variance (ANOVAs) were conducted for each of the measurements prior to multivariate analyses. The ANOVAs revealed that experts performed significantly better than non-experts on Connections A, $F(1, 27) = 5.17$, $p = .031$, $\eta^2 = .161$ and Connections B, $F(1, 27) = 5.84$, $p = .02$, $\eta^2 = .18$. ANOVAs also revealed that experts performed marginally significantly better on Pattern Comparison, $F(1, 27) = 3.64$, $p = .067$, $\eta^2 = .12$ and WCST, $F(1, 27) = 3.19$, $p = .085$, $\eta^2 = .11$, and scored lower on the measure of BIS, $F(1, 24) = 3.73$, $p = .065$, $\eta^2 = .14$. No other measures neared significance ($ps > .10$). The results of the ANOVAs and the effect sizes (see Table 2.2) suggest that faster psychomotor speed (Connections A), stronger cognitive control and task switching (Connections B), increased willingness to take risks (less inhibition surrounding risk-taking; BIS), faster cognitive processing speed (Pattern Comparisons), and more mental flexibility (WCST), are important for differentiating expert and non-expert interpreters.

Table 2.2

Descriptive statistics (N = 29)

Measurements	Experts (n = 15)		Non-experts (n = 14)		Cohen's d
	M	SD	M	SD	
Connections B	18.09	3.36	14.02	5.53	.93
BIS	20.46	2.21	23.15	3.76	-.92
Connections A	33.65	6.04	28.63	5.78	.88
WCST	76.92	9.06	68.85	12.12	.79
Pattern Comparison	.72	.14	.62	.13	.77
Letter Comparison	.56	.10	.51	.10	.52
Symmetry Span	.61	.10	.54	.18	.50
Need for Cognition	69.46	8.81	66.54	11.50	.30
Ravens	8.67	3.09	8.00	2.77	.24
BAS	13.46	1.84	13.61	1.53	.09

Note. Positive Cohen's *ds* represent the experts performing higher on the measures. The negative Cohen's *d* shows that experts scored lower on this measure.

Multivariate analyses

ANOVAs reveal differences between groups along a single dimension without taking into account other factors. Analyzing complex data sets in this manner can cause researchers to miss effects from contributing factors working in tandem with the variables being parceled out for univariate analysis. Discriminant function analysis reveals the best linear combination of predictors that differentiate two groups. Discriminant function analysis is similar to binary logistic regression, but is more powerful and efficient (Tabachnick & Fidell, 2007) and provides more accurate classifications and hypothesis testing as long as statistical assumptions are held (Grimm & Yarnold, 1995).

Prior to analysis, the cases were analyzed for multivariate outliers. Mahalanobis distance indicated two (one from each skill-level group) multivariate outliers (*z*-scores > 3.0). The two outliers were removed. Cases with missing values were also removed when measurements with the missing values were entered into the analyses in keeping with the requirement for discriminant function analysis. With the removal of the two multivariate outliers as well as cases

with missing values, and with no more predictors (number of measurements entered into the analysis) than cases in the smallest group, the statistical assumptions underlying discriminant function analysis were met, leaving 89% of the cases available for analysis.

Discriminant function analysis creates statistical models in order to predict group membership from a set of predictors (also known as classifiers). Various combinations of predictors were evaluated. A successful statistical model was revealed, Wilks' Lambda = .55, $\chi^2(5) = 11.62$, $p = .040$, correctly classified cases = 83.3% (76.9% of the expert interpreters, 90.9% of the non-expert interpreters). Like the results of the univariate analyses suggested, the measurements with the smallest p values and largest effect sizes created the significant discriminant function model: Connections A, BIS, Connections B, Pattern Comparison, and WCST.

In order to understand the relationships between the classifiers and group membership, one can calculate the structure coefficients and standardized canonical discriminant function coefficients. Structure coefficients are the correlations of each classifier to the discriminant functions (similar to factor loadings in factor analysis). Structure coefficients are used to assign meaningful labels to the discriminant functions. The structure matrix revealed the following coefficients: Connections B = .66, Connections A = .51, WCST = .37, Pattern Comparison = .35, and BIS = $-.19$. The expert group served as the reference group, therefore positive coefficients indicate the positive correlation of scores from the expert group to the discriminant functions while negative coefficients indicate the negative correlation of scores from the non-expert group to the discriminant functions. In other words, the higher an individual scored on Connections B, Connections A, WCST, and Pattern Comparison and the lower the score on BIS, the more the scores correlated to the discriminant functions and the likelier the model was to predict that the

individual belonged in the expert group. The following discriminant function labels were created from the result of the structure coefficient matrix: task switching ability (Connections B), psychomotor speed (Connections A), mental flexibility (WCST), cognitive processing speed (Pattern Comparison), and aversion to risk (BIS).

The standardized canonical discriminant function coefficients indicate the unique contribution of each classifier to the discriminant functions and are used to determine the relative importance of the classifiers in predicting group membership (similar to beta weights in multiple regression). Mental flexibility (.63) is the most important predictor relative to the other entered classifiers followed by cognitive processing speed (.61), aversion to risk (−.52), task-switching ability (.51) and, substantially less important, psychomotor speed (.22).

To further establish the validity of the classification, a permutation test was conducted. This examines the possibility that the discriminant function solution does not capture fundamental differences between experts and non-experts, but is simply a brute-force mathematical solution for separating cases into groups, regardless of what those groups might be. Put differently, the permutation test assumes that discriminant function analysis will find a solution predicting group membership for any arbitrary groups. To carry out this test, group membership is randomly reassigned for all the cases and the discriminant function analysis re-conducted with the original predictors. If this produces a statistically significant classification, then the original result is undermined and the discriminant function analysis has not captured true empirical differences between these groups.

However, if the discriminant function analysis is unable to produce a significant classification of the permuted data, then the original result does reflect true differences between these groups in the indicated abilities. Discriminant function analysis was performed using the

same five predictors with the cases now randomly assigned to the expert and non-expert groups. The results were not significant, Wilks' Lambda = .74, $\chi^2(5) = 5.77$, $p = .33$, indicating that the original discriminant function analysis model is tracking true and important differences between expert interpreters and those of lower skill level.

2.5 Discussion

The results from the ANOVAs, effect sizes, and discriminant function analysis strongly suggest that expert interpreters are more mentally flexible, have faster cognitive processing speed, are less anxious about risks, are faster and more accurate when task switching, and have faster psychomotor speed than non-expert interpreters, regardless of the fact that both groups have the same amount of professional experience. Additionally, results from the multivariate analysis indicate that mental flexibility and cognitive processing speed are the most important predictors closely followed by willingness to take risks and task switching ability and, to some extent, psychomotor speed. These five predictors were all significant and relatively powerful individual predictors, and together were the only combination of predictors to significantly and successfully predict group membership. The convergence of all the analytical results supports the hypothesis that certain general cognitive abilities and emotional-cognitive interaction traits are strongly related to simultaneous interpreting performance.

Additionally, two other measures, Letter Comparison and Symmetry Span, exhibited medium effect sizes, but were not statistically significant predictors. These results, however, imply that variations of the tasks measuring the same or similar constructs may contribute significantly in a future study. (A variant of Letter Comparison, Pattern Comparison, already provided significant statistical results in the current study.)

The medium effect size demonstrated by Symmetry Span suggests that working memory capacity differences may be predictive of interpreter performance, but further studies should explore various tools that better capture the processing and storage capacities needed during simultaneous interpreting. Oberauer (2004) defines three distinct types of working memory capacity measurements that load onto a single working memory factor. They are 1) storage of briefly presented material concurrent with interfering processing, 2) formation of new structures and relationships (storage while transforming maintained information), and 3) some executive functioning, such as updating. Symmetry Span primarily taps the first factor. Simultaneous interpreting on the other hand, demands storage along with concurrent and continuous restructuring. Therefore, future studies utilizing working memory measures of maintenance and transformation ability are likely to prove more auspicious when assessing variation among interpreters or students of interpreting than the measurement used in the current study. Indeed, the current results identify several psychological constructs for which multiple measures have been developed. In future work, structural equation modeling would potentially provide much greater resolution on the issue of exactly which abilities and traits combine, and in what fashion, to predict simultaneous interpreting performance. Furthermore, such techniques are naturally suited to address issues of causality, provided that the relevant longitudinal data are available.

While many of the cognitive ability measures appear promising in predicting group membership, years of professional experience demonstrated no relationship with group membership. This finding suggests that other variables (e.g., domain-general cognitive abilities) better explain variance in interpreter performance than years of experience. If interpreting was a repetitive, predictable task, in which pacing was controlled by the performer (e.g., mirror tracing or cigar rolling), experience would likely correlate with performance. However, interpreting

demands that the performer assess and adapt to a different setting each time she performs and efficiently process ephemeral information never processed previously. If an individual does not adapt well to the demands of the environment, does not efficiently process ephemeral information, and is anxious about missing information and therefore does not restructure the message appropriately, experience will have a limited positive effect and could potentially have a negative effect as repeatedly performing the task poorly will enforce performing the task poorly. In contrast, if an individual's executive control processes support her ability to adapt to the demands of the environment, efficiently process ephemeral information, and restructures the message into the target language appropriately, she will either improve with experience or will perform at a higher rate than an interpreter whose executive control processes fail to support the task demands regardless of years of experience. Further studies are necessary for determining whether training during the initial learning phase is more predictive of interpreting performance than years of professional experience and whether cognitive abilities will still explain relatively more performance variance compared to experiential factors. If this is the case, a next step is identifying which features of simultaneous interpreting lend themselves to being predicted by cognitive abilities rather than experience in order to generalize and understand skill acquisition across domains.

The cross-sectional design of this study does not speak to whether the cognitive profiles of the expert and non-expert interpreters have remained stable since before interpreter training to the time of this study, or whether these abilities changed from interpreting experience, perhaps varying in their amount of change. Given that simultaneous interpreting is a complex bilingual task, this question is relevant not only to individual differences and skill acquisition and

expertise research, but to research conducted on the bilingual advantage. The goal of Chapter 3 is to investigate cognitive ability change over the course of simultaneous interpreter training, within the context of the bilingual advantage literature.

Chapter 3

Cognitive ability change from complex bilingual processing experience

3.1 Chapter summary

Cognitive abilities of participants with low inter-language conflict and high bilingual demands are measured longitudinally.⁵

- The source of the bilingual advantage—cognitive enhancement in bilinguals relative to monolinguals—remains open to debate.
- A major hypothesis is that experience managing inter-language conflict produces the result.
- The hypothesis driving this study is that the bilingual advantage depends on the cognitive demands of the bilingual task and experience with those demands and not degree of inter-language conflict.
- Participants with low inter-language conflict and high bilingual management demands were recruited.
- The bilingual advantage emerged with experience within this sample.

⁵ This study can be found in a published report (Macnamara & Conway, 2013). Also previously presented at a Psychonomic Society meeting (2012), an Association for Psychological Science convention (2013), and the International Symposium on Signed Language Interpretation and Translation Research (2014).

3.2 Introduction: The bilingual advantage

The bilingual advantage refers to enhanced cognitive control in bilinguals relative to monolinguals, and has been found in infants, children, and adults (Bialystok & Craik, 2010). Bialystok and colleagues suggest that the bilingual advantage occurs from bilinguals' experience switching between languages and the languages' respective grammatical properties. In other words, in order to appropriately switch between languages, bilinguals must (a) control the selection of the target language and (b) resolve conflict between competing alternatives; these ongoing demands within bilinguals increase the amount of practice with engaging processes supported by domain-general mechanisms (Bialystok, 1999; Green, 1998). According to this interpretation, the bilingual advantage should be strongest among bilinguals who speak two languages with a high degree of conflict, and weakest among bilinguals who speak two languages with less conflict. A few studies have investigated the effect of varying degrees of inter-language conflict on cognitive enhancement. For example, Emmorey, Luk, Pyers, and Bialystok (2008) found that bimodal (signed language–spoken language) bilinguals did not outperform monolinguals on tasks of cognitive control. Emmorey, Luk, et al. explained that the lack of a bilingual advantage in this population was likely due to the two languages' distinct motor and perceptual pathways reducing the needs for both conflict resolution (Emmorey, Borinstein, Thompson, & Gollan, 2008) and target language selection (Pyers & Emmorey, 2008)—that is, low inter-language conflict. In contrast, research with unimodal (spoken language–spoken language) bilinguals of various language pairs with varying amounts of grammatical overlap—and therefore varying amounts of inter-language conflict—has revealed uniform bilingual advantages in cognitive abilities for children (Barac & Bialystok, 2012). These

results suggest that varying levels of inter-language conflict do not necessarily vary the amount of executive control enhancement.

Other studies have also contributed to the debate surrounding the bilingual advantage: Some researchers have failed to observe a bilingual advantage at all (Morton & Harper, 2007; Paap & Greenberg, 2013), and in cases in which differences are observed, the effect may be confounded with various group differences (e.g., Morton & Harper, 2007). However, other studies have controlled for such differences—for example, socio-economic status—and still show a bilingual advantage (Bialystok, 2009). In a domain as complex as bilingualism, many differences among sample groups are likely to emerge, thus making generalizations to all bilinguals difficult.

The goal of the present study is to test a novel hypothesis: enhanced cognitive control does not occur simply from being bilingual or necessarily from the degree of inter-language conflict, but on how cognitively demanding the management of the two languages are for the individual and the amount of experience with these bilingual management demands (BMD). For example, Paap and Greenberg's (2013) finding of no advantage in executive functions among college student bilinguals may have been due to specific features of their sample. College students may have relatively low BMD; they are generally surrounded by one dominant environmental language, and therefore receive continuous target language cues. Thus, they may not need to engage as much executive control to manage the bilingual processes of inhibiting the incorrect target language and selecting among competing alternatives as would a bilingual who frequently switches languages or who frequently functions in a multilingual environment. Although evidence suggests that both languages in bilinguals are activated even in monolingual environments (Kroll, Bobb, Misra, & Guo, 2008), recent investigations have shown that those

who frequently switch between their languages have enhanced cognitive control as compared with bilinguals who do not frequently switch between their languages (Calabria, Hernández, Branzi, & Costa, 2011; Festman, Rodriguez-Fornells, & Münte, 2010; Prior & Gollan, 2011; Soveri, Rodriguez-Fornells, & Laine, 2011). These findings support the idea that cognitive enhancements among (some) bilinguals may not just be a result of being bilingual, or of the degree of similarity between languages, but from how much one needs to engage cognitive control in order to manage the two languages.

The hypothesis to be tested is that the degree to which cognitive control mechanisms are recruited—and therefore, the degree to which a bilingual advantage would emerge—depends on two variables: (1) the magnitude of the BMD and (2) the amount of experience managing those demands. More specifically, that heightened BMD and experience managing those demands are necessary to produce enhanced cognitive functioning. This hypothesis yields clear predictions: 1) If heightened bilingual management demands—the degree to which one must engage cognitive control in order to appropriately use the two languages—enhances cognitive control and working memory (WM), then individuals with high bilingual management demands, even with low inter-language conflict, should demonstrate enhanced cognitive abilities with experience.

In order to test this hypothesis, American Sign Language (ASL)–English simultaneous interpreter students were recruited. ASL–English interpreter students are bimodal bilinguals, and thus have low inter-language conflict, but should have high BMD—they need to switch frequently between languages, coordinate the two languages simultaneously, comprehend the source message in one language while producing the target message in the other language, and use executive control to select the appropriate target language in environments in which both languages are cued (Christoffels & de Groot, 2005; Gerver, 1976; Gile, 1997; Lambert, 1992;

Padilla, Bajo, Cañas, & Padilla, 1995; Yudes, Macizo, & Bajo, 2012). Experience managing these heightened bilingual demands was varied by recruiting beginning (first-semester) ASL–English interpreter students and comparing these same students to themselves later as advanced (fourth-semester) ASL–English interpreter students. Following the hypothesis, ASL–English interpreter students should outperform themselves 2 years earlier on measures of general cognitive abilities engaged during simultaneous interpreting.

Other studies have shown that professional unimodal simultaneous interpreters outperform non-interpreter unimodal bilinguals in measures of mental flexibility and cognitive control (Yudes, Macizo, & Bajo, 2011), and that professional unimodal interpreters outperform unimodal interpreter students, who in turn outperform non-interpreter unimodal bilinguals in WM processes (Yudes et al., 2012). These studies suggest that experience interpreting increases cognitive abilities, but given the cross-sectional designs, one cannot know whether those who succeed as interpreters already have enhanced cognitive abilities or whether experience interpreting causes enhanced cognitive abilities. The present study therefore investigated interpreter students longitudinally, in order to determine whether cognitive abilities—namely, cognitive control and working memory capacity—become enhanced with interpreter training.

With respect to cognitive control, measures of task switching and mental flexibility—two cognitive abilities often associated with the bilingual advantage—should be enhanced among advanced interpreter students relative to themselves as beginning interpreter students, because simultaneous interpreting demands rapid and flexible switching between sets of grammatical and cultural rules. Simultaneous interpreting also demands maintaining a portion of the source message in memory while transforming another portion of the message into the target language. Therefore, with respect to WM, enhancement among advanced interpreter students in tasks that

required the “coordination and transformation” of information (Oberauer, 2004) should be observed if the present hypothesis is correct. In contrast, WM, as measured by tasks that require “maintenance and interference” (e.g., complex span tasks; Unsworth, Heitz, Schrock, & Engle, 2005), should not differentiate advanced and beginning interpreter students. Simultaneous interpreting does not demand the maintenance of information while processing an unrelated task; rather, interpreters are trained to find relationships among pieces of information in order to determine the gestalt meaning of the communication. Finally, although these data were not directly related to the hypothesis, measures of psychomotor speed, perceptual speed, and general fluid intelligence were also administered, in order to gauge other potential differences across groups.

3.3 Methods

Participants

The participants were 21 ASL–English simultaneous interpreting students. The interpreter students were recruited in their first semester and were retested in their fourth (final) semester. All participants were recruited from two institutions with similar curricula. Demographic information is presented in Table 3.1.

Table 3.1

<i>Demographic information</i>	
Sample Size	
Institution I	13
Institution II	8
Age	
<i>M (SD)</i>	26.24 (5.08)
Median (range)	24 (21-40)
CODAs	3
Female	18
Semesters of College	
<i>M (SD)</i>	8.81 (3.78)
Median (range)	8 (2-17)
Semesters of ASL	
<i>M (SD)</i>	4.90 (2.28)
Median (range)	5 (1-12)

Note. CODAs = Children of Deaf Adults (i.e., participants exposed to ASL since birth)

Procedure and Materials

Measures of seven cognitive constructs were included. Two of these constructs were representative of those common to the bilingual advantage literature: task switching and mental flexibility. Two types of cognitive processing speed—perceptual speed and psychomotor speed—and also two tests of nonverbal intelligence/reasoning ability were included. Finally, working memory capacity was measured in two ways: 1) one's ability to coordinate and transform maintained information (coordination and transformation), and 2) one's ability to maintain information in the face of interference (maintenance and interference). (See Table 3.2.)

Table 3.2

Measurement descriptions

Construct	Tasks	Reference	Participants Are Asked To...
Task-switching	- Connections B: Letters-Numbers - Connections B: Numbers-Letters	Salthouse et al. (2000)	rapidly alternate between numerical and alphabetical sequences.
Mental flexibility	- Wisconsin Card Sorting Test	Berg (1948)	learn a rule based on feedback and adapt to new rule when feedback changes.
Perceptual speed	- Pattern Comparison - Letters Comparison	Salthouse & Babcock (1991)	determine whether two figures or letter strings are identical.
Psychomotor speed	- Connections A: Letters - Connections A: Numbers	Salthouse et al. (2000)	draw lines in alphabetical or numerical sequence as quickly as possible.
Nonverbal intelligence	- Raven's Advanced Progressive Matrices - Cattell's Culture Fair Test	Raven (1962); Cattell, Cattell, & Institute for Personality and Ability Testing (1960)	recognize patterns and solve problems on the basis of non-verbal patterns.
Working memory: Coordination & transformation	- Backward digit span - Letter-number sequencing	no reference associated; adapted from Wechsler (1997)	mentally transform the presented stimuli into a different order for recall.
Working memory: Maintenance & interference	- Automatic Reading Span - Automatic Operation Span	Unsworth et al. (2005)	recall memoranda in sequence while performing a secondary task.

3.4 Results

The predicted pattern emerged: The advanced interpreter students outperformed themselves as beginning interpreter students on multiple measures, including the two measures representative of those common to the bilingual advantage literature: task switching, $t(20) = 4.54, p < .001, d = 0.99$ and mental flexibility, $t(20) = 3.05, p < .01, d = 0.67$, on the two measures of cognitive processing speed—perceptual speed, $t(20) = 2.41, p < .05, d = .46$, and psychomotor speed, $t(20) = 8.74, p < .001, d = 1.91$ —and the two tests of nonverbal intelligence/reasoning ability: Ravens, $t(20) = 2.99, p < .01, d = 0.65$ and Cattell's, $t(20) = 2.90, p < .01, d = 0.63$. Advanced interpreter students also outperformed themselves as beginning interpreter students on measures of one's ability to coordinate and transform maintained information (coordination and transformation measures of working memory capacity): letter-number sequencing, $t(20) = 3.47, p < .01, d = 0.76$ and backward digit span, $t(20) = 3.91, p < .001, d = .85$, but not on measures of one's ability to maintain information in the face of interference (maintenance and interference measures of working memory capacity): automated reading span, $t(20) = 0.40, ns, d = .18$ and automated operation span, $t(20) = 0.21, ns, d = .05$, see Table 3.3.

Table 3.3

<i>Within-group contrasts: Interpreter students</i>					
Measures	Beginning Students		Advanced Students		Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Cognitive Control, Speed, and Reasoning					
Task Switching					
Connections B [†]	18.64	3.96	21.31	5.14	.99
Mental Flexibility					
Wisconsin Card	76.93	10.72	83.26	4.55	.67
Sorting Test*					
Perceptual Speed					
Pattern/Letter	.63	.08	.67	.10	.46
Comparisons					
Psychomotor Speed					
Connections A**	33.64	5.63	39.55	5.60	1.91
Intelligence/ Reasoning Ability					
Raven's Matrices [†]	.50	.16	.59	.17	.65
Cattell's Culture	.53	.11	.58	.13	.63
Fair Test					
Working Memory Capacity					
Coordination & Transformation					
Letter-Number	.86	.06	.91	.06	.76
Sequencing*					
Backward Digit	.88	.06	.93	.06	.85
Span*					
Maintenance & Interference					
Automated	.68	.19	.72	.17	.18
Reading Span					
Automated	.67	.26	.68	.27	.05
Operation Span					

Note. Measures of speed and task switching are scored either as numbers or proportions of correct responses within a time limit. [†] $p < .08$, * $p < .05$, ** $p < .01$.

Additional Analyses and Discussion

Overall, the results are consistent with the notion that the magnitude of bilingual management demands and the amount of experience managing those demands predict enhancement in cognitive control and WM. It is important to consider, however, potential confounding variables—that is, other variables varying across the groups that might also

contribute to the observed effects.

First, being raised as a bilingual could be driving the effects: Three of the 21 interpreter students had learned ASL since birth. Therefore these individuals were excluded and the analyses were repeated. The results were consistent with the original findings, suggesting that these individuals were not driving any of the observed effects. Two p values changed slightly, but this was to be expected, given the reduction in degrees of freedom; all of the effect sizes remained relatively stable (task switching, 0.96; mental flexibility, 0.82; perceptual speed, 0.48; psychomotor speed, 1.85; intelligence/ reasoning ability, 0.63/0.61; coordination and transformation, 0.96 and 0.74; maintenance and interference, 0.07 and 0.02).

Second, it is possible that the program selection and progression criteria could differ between the interpreter programs and that these criteria could play a role in the effects observed here. Institution was entered as a covariate for all of the analyses, and no significant differences were observed on any of the measures between participants in the two institutions (all $ps > .05$).

Relatedly, the effects could be driven by self-selection or progression criteria at the individual level. In other words, interpreter students with higher cognitive abilities might be more likely to enter and to progress through an interpreter program, and therefore our findings could be the result of initial cognitive abilities and not of the interpreter training. Teachers and the nine interpreter students who participated in the study as beginning students but did not participate as advanced students were contacted. Of the nine who left the study, three left the study but progressed through the program with their peers, and one temporarily left the program for personal reasons and came back to finish. The remaining five (one of whom had been learning ASL since birth) were compared to the other beginning interpreters who progressed in both the interpreter program and the present study. Although firm conclusions should not be

drawn from this limited sample, pairwise comparisons revealed that this group of five did not differ in any of the cognitive ability measures from the beginning interpreters who had continued in the study, tentatively suggesting that initial cognitive control, WM capacity, and general intelligence did not differentiate those who complete the program from those who did not.

Finally, to further test whether the increase in cognitive abilities for the interpreter students could be explained by preexisting cognitive abilities, we conducted regression analyses with initial (beginning student) scores as the predictor variable and the gain scores (advanced student scores minus beginning student scores) as the outcome variable. The initial score was predictive of the gain score in three measures: the Wisconsin Card Sorting Test, the backward digit span task, and the letter–number sequencing task. However, in these cases, initial score negatively predicted gain score. Inspecting scatterplots revealed a ceiling effect. Those initially near the highest possible score on the tasks had lower gain scores, since there was not room for improvement: The farther from ceiling the initial score, the more that these participants improved from the beginning to the end of the program (Wisconsin Card Sorting Test, $\beta = -.906$; letter–number sequencing, $\beta = -.585$; backward digit span, $\beta = -.587$). Taken together, the evidence does not support the idea that the participants in this study experienced cognitive enhancements due to institutional or participant characteristics (i.e., learning ASL in childhood, higher initial cognitive abilities.)

3.5 Discussion

Given that previous findings with bimodal bilinguals (with far more bilingual experience than our sample had) have not demonstrated enhanced cognitive functions (Emmorey, Luk, et al., 2008), and that participants in this sample who had learned ASL since birth did not seem to

be driving the effects, the results suggest that experience with ASL alone does not produce cognitive enhancements. Instead, the results support the hypothesis that the mechanism responsible for cognitive enhancements among bilinguals is the interplay between the magnitude of BMD and the amount of experience managing those demands. Among bilinguals with high BMD, experience enhanced cognitive control and WM: The advanced interpreter students demonstrated a bilingual advantage. This finding is consistent with previous results in similar populations, namely Yudes et al.'s (2011, 2012) findings that simultaneous (unimodal) interpreters—bilinguals who have higher target language selection demands and inter-language coordination demands—outperformed bilinguals with less interpreter experience and bilinguals with lower BMD. Our study was able to demonstrate longitudinally that simultaneous interpreters' cognitive control and WM improve with experience.

Cognitive enhancements—sometimes of a magnitude near or over a standard deviation—were observed among bilinguals with high BMD and experience managing those demands, relative to themselves before they had accumulated experience managing these demands. These results also suggest that the specific mechanisms engaged while managing bilingual demands depend on the specific performance demands of the bilingual processing task. For example, ASL–English interpreters have high hand–eye coordination demands—demands that unimodal spoken-language bilinguals presumably would not have. Enhancements specific to this demand—that is, enhanced psychomotor speed, were observed. Another BMD-specific enhancement—that is, one specific to the cognitive demands of simultaneous interpreting—was also observed in the dissociation between WM task types: only enhancements to WM were observed when the task engaged processes assumed to be active during simultaneous interpreting.

The bilingual advantage has generated many years of interesting research, but the source has remained controversial. The present hypothesis assumed that the bilingual advantage does not occur from maintaining two languages within an individual per se, but occurs from experience managing bilingual demands. When BMD are high, such as when an individual rapidly switches between the languages and must coordinate the comprehension and production of the two languages simultaneously, the individual heavily recruits cognitive control mechanisms in order to manage this language task, leading to cognitive control and WM enhancements.

These results suggest a possible explanation of why researchers working with different bilingual populations find discrepant results: Bilingual populations differ in their experiences with, magnitude of, and type of BMD (e.g., college students habituated to the majority language differ from those who frequently switch languages), and therefore are likely to differ in their executive control enhancement. These results also provide a potential framework for understanding the relationship between performance demands and cognitive enhancements, not only among bilingual populations, but also in general as individuals acquire and perform complex skills with varying demands.

Chapter 4

Predictors of skill acquisition in a complex bilingual processing task

4.1 Chapter summary

The roles of initial cognitive abilities, training, and change in cognitive abilities from training on skill acquisition are measured over time among American Sign Language-English simultaneous interpreter students⁶.

- Within a complex bilingual processing task—simultaneous interpreting—it is unclear how cognitive abilities, training, and cognitive ability enhancement effect skill acquisition rates.
- This study measured cognitive abilities and affective traits of simultaneous interpreter students four times over two years of interpreter training, and simultaneous interpreting performance three times over the same two years.
- Initial cognitive abilities predicted final interpreting performance.
- Time in the program predicted interpreting performance.
- Several cognitive abilities became enhanced over the course of interpreter training.

⁶ This study is in preparation for publication (Macnamara & Conway, in prep). Also, partially presented at the International Symposium on Signed Language Interpretation and Translation Research (2014).

4.2 Introduction: Longitudinal measurement of skill

Chapter 2 presented a study investigating predictors of expertise among simultaneous interpreters. The primary finding was that amount of domain-specific experience did not predict expertise and individual differences in cognitive abilities did predict expertise. Chapter 3 presented a study investigating the role of domain-specific simultaneous interpreting experience on general cognitive abilities. The primary finding was that domain-specific experience predicted change in general cognitive abilities. Taken together, these results suggest a complex relationship among domain-specific experience, general cognitive abilities, and performance among simultaneous interpreters that warrants further investigation.

The longitudinal design of the present study offers a framework for testing and comparing multiple hypotheses regarding the complex relationships among experience, cognitive abilities, and performance. One hypothesis—likely to be of interest to proponents of the experts-are-made/nurture-over-nature view—is that experience is compensatory. That is, cognitive abilities and other factors are likely to predict initial performance, but that with accumulated training, performance will converge. A second hypothesis—likely to be of interest to proponents of the experts-are-born/nature-over-nurture view—is termed “the rich get richer.” That is, cognitive abilities are likely to predict initial performance, and that with accumulated training, performance variance will diverge so that those who begin with higher initial performance will have the steepest learning curves while those with the lowest initial performance will produce the shallowest learning curves. A third hypothesis—likely to be of interest to researchers of the bilingual advantage—is that change in cognitive abilities will best predict skill acquisition rates. That is, individuals whose enhancement in cognitive abilities is the greatest, will have the

steepest learning curves while individual whose cognitive abilities remain relatively stable over time will produce the shallowest learning curves.

This study presents two goals. One goal is to examine the relative contributions of these predictors by asking how well each hypothesis—compensatory, “rich get richer”, and change in abilities—fits the data in comparison to the others. The second goal is to develop hypotheses about predictors of skill acquisition across performance domains. In other words, how do the results from the studies on simultaneous interpreting generalize to other performance domains?

4.3 Methods

Participants

Fifty ASL-English simultaneous interpreting students were recruited in the middle of their first semester and again mid-semester for the next three semesters for a total of four data collection waves (approximately 20%, 40%, 60% and 80% completion through the interpreter training program). Of the 50 participants, 34 provided data at all four time points. Of the sixteen who did not provide data at all four time points, nine did not pass coursework in the training program and were asked by the faculty to either repeat the coursework or leave the program, three left the program for personal reasons but came back to the program or planned to come back to the program, and four finished the program on pace with their peers, but chose not to continue participating in the study. Demographic information is presented in Table 4.1 and 4.2.

Table 4.1

Demographic information: Entire targeted sample

Sample Size	
Institution I	27
Institution II	23
Age	
<i>M (SD)</i>	27.90 (7.49)
Median (range)	29 (20-56)
CODAs	7
Female	82%
Semesters of College	
<i>M (SD)</i>	8.54 (3.62)
Median (range)	8 (2-17)
Semesters of ASL	
<i>M (SD)</i>	5.24 (2.34)
Median (range)	5 (1-12)

Note. CODAs = Children of Deaf Adults (i.e., participants exposed to ASL since birth). All variables refer to data collected during the first semester of interpreter training.

Table 4.2

Demographic information: Sample providing data under analysis

Sample Size		
Institution I		19
Institution II		15
Age		
<i>M (SD)</i>		26.68 (6.88)
Median (range)		24 (20-56)
CODAs		3
Female		85%
Semesters of College		
<i>M (SD)</i>		8.67 (3.65)
Median (range)		8 (2-17)
Semesters of ASL		
<i>M (SD)</i>		5.24 (2.34)
Median (range)		5 (1-12)

Note. CODAs = Children of Deaf Adults (i.e., participants exposed to ASL since birth). All variables refer to data collected during the first semester of interpreter training.

Procedure and Materials

Interpreter students were recruited mid-semester during their first semester of training to complete a battery of affective and cognitive assessments. Participants completed the same battery of measures mid-semester of each training semester for a total of four times per participant. Participants also completed a performance measure of simultaneous English to ASL interpreting during their second, third, and fourth semesters. (See Table 4.3 for measurement order.) The performance measure was not included during the first wave of data collection because the participants had not had more than a few weeks of training, would presumably be unable to complete the task, and posed a threat that participants would find the task too stressful at that stage and choose to leave the study.

The length of time between data collection waves coupled with the generally high test-retest reliability of the cognitive ability and affective trait measures (c.f., e.g., Carver & White, 1994; Conway et al., 2005; LeMay, Be'dard, Rouleau, & Gilbert Tremblay, 2004; Unsworth et al., 2005) suggests that changes in performance on these measures is likely due to change in cognitive abilities and not practice effects. Likewise, change in performance on the simultaneous interpreting measure is likely due to change in interpreting skill and not measurement practice effects: the three presentations of the simultaneous interpreting measure, presented approximately 6 months apart, follow the mandatory wait time rules imposed by the National Council of Interpreting between interpreting tests (that is, an individual must wait 6 months before retaking the National Interpreting Certification test; NAD-RID National Interpreter Certification Candidate Handbook, 2013).

Non-verbal Reasoning / Fluid Intelligence

Two non-verbal reasoning / fluid intelligence tasks were administered and scores from the tasks were combined to create a composite non-verbal reasoning / fluid intelligence measure.

Raven's Advance Progressive Matrices. Raven's Advanced Progressive Matrices (Raven, 1962) is a multiple-choice abstract reasoning test designed to measure general fluid intelligence and reasoning in above-average intelligence individuals. This measure is described in more detail in Chapters 2 and 3 of this document.

Cattell's Culture Fair Test. Cattell's Culture Fair Test (Cattell, Cattell, & Institute for Personality and Ability Testing, 1960) is a multiple-choice reasoning test designed to measure general fluid intelligence and reasoning. Participants attempt to complete patterns and solve non-verbal problems within time limits. This measure is described in more detail in Chapter 3 of this document.

Working Memory Capacity (Maintenance & Interference)

Two working memory (maintenance and interference) tasks were administered and scores from the tasks were combined to create a composite working memory (maintenance and interference) measure.

Automated Reading Span. Automated Reading Span (Unsworth et al., 2005) is one of several “complex span tasks” that measure working memory capacity by measuring item recall in the face of interference. This measure is described in more detail in Chapter 3 of this document.

Automated Operation Span. Automated Operation Span (Unsworth et al., 2005) is also a “complex span task” designed to measure working memory capacity by measuring item recall in the face of interference. This measure is described in more detail in Chapter 3 of this document.

Working Memory Capacity (Coordination & Transformation)

Two working memory (coordination and transformation) tasks were administered and scores from the tasks were combined to create a composite working memory capacity (coordination and transformation) measure.

Letter-Number Sequencing. Letter-Number Sequencing (adapted from Wechsler, 1997) measures one’s ability to manipulate information in one’s mind. Participants view a combination of 3-7 digits and letters presented one at a time on the screen. Participants are tasked with recalling the numbers and letters in numerical and alphabetical sequence (and not the order of presentation). This measure is described in more detail in Chapter 3 of this document.

Backward Digit Span. Backward Digit Span measures one’s ability to manipulate information in one’s mind. Participants view 3-7 digits presented one at a time on the screen.

Participants are tasked with recalling the numbers in the reverse order of presentation. This measure is described in more detail in Chapter 3 of this document.

Perceptual Speed

Two perceptual speed tasks were administered and scores from the tasks were combined to create a composite perceptual speed measure.

Letter Comparison. The Letter Comparison task (Salthouse & Babcock, 1991) assesses cognitive processing speed by measuring perceptual and decision-making speed through comparison of orthographic patterns within a time limit. This measure is described in more detail in Chapters 2 and 3 of this document.

Pattern Comparison. The Pattern Comparison task (Salthouse & Babcock, 1991) is the same as the Letter Comparison task except instead of non-lexical letter strings Pattern Comparison uses abstract visual figures. This measure is described in more detail in Chapters 2 and 3 of this document.

Psychomotor Speed

Two psychomotor speed tasks were administered and scores from the tasks were combined to create a composite psychomotor speed measure.

Connections Test A: Numbers. Connections A: Numbers (Salthouse et al., 2000) measures psychomotor speed by asking participants to connect a sequence of numbers as quickly as possible. This measure is described in more detail in Chapters 2 and 3 of this document.

Connections Test A: Letters. Connections A: Letters (Salthouse et al., 2000) measures psychomotor speed by asking participants to connect a sequence of letters as quickly as possible. This measure is described in more detail in Chapters 2 and 3 of this document.

Task-Switching Speed

Two task-switching speed tasks were administered and scores from the task were combined to create a composite task-switching speed measure.

Connections B: Numbers-Letters. Connections B: Numbers-Letters (Salthouse et al., 2000) measures cognitive control, task switching speed, and psychomotor speed by asking participants to connect alternating sequences as quickly as possible. This measure is described in more detail in Chapters 2 and 3 of this document.

Connections B: Letters-Numbers. Connections B: Letters-Numbers (Salthouse et al., 2000) is the same as Connections B: Numbers-Letters except the order of the sequencing is reversed. This measure is described in more detail in Chapters 2 and 3 of this document.

Mental Flexibility

Wisconsin Card Sorting Test. The Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948)) measures set-shifting—the ability to adapt to changes and to shift among different sets of rules—i.e., mental flexibility. This measure is described in more detail in Chapters 2 and 3 of this document.

Emotion-Cognition Interaction

BAS scale. The BAS scale (Carver & White, 1994) measures reward sensitivity, drive, and fun-seeking traits. This measure is described in more detail in Chapter 2 of this document.

BIS scale. The BIS scale (Carver & White, 1994) measures risk-taking sensitivity and anxiety surrounding aversive stimuli and novelty. This measure is described in more detail in Chapter 2 of this document.

Simultaneous Interpreting Performance

Participants listened to a short story presented in spoken English and attempted to simultaneously interpret the story into ASL. Before listening to the story participants were read the following background: “You are going to listen to a short, ten-sentence story about two middle-school friends named Joe and Brian who are talking over lunch. Joe and Brian are the only two people in the scene, although others may be mentioned, and Joe and Brian are both present from the start. Please simultaneously interpret the story. Use as much processing time as you would like. While the story is short and simple, it is not necessarily easy to interpret, so just do the best you can. I won’t watch so just let me know when you are done.” The experimenter did not observe the participants interpreting as they performed so that there was not additional performance pressure of an audience.

Participants then listened to the following story, read by a male actor:

So Joe, looked down at the food on his tray and he made—he made a face, and he said to his friend Brian “Ugh! It’s macaroni and cheese again! They never serve anything good for lunch!” Brian looked down at his lunch and he’s, he’s frowning, and he says “You think that’s bad?! I’ve got peanut butter and jelly again. It’s the third time this week. Maybe I’ll tell Mom, I’m, like, allergic so she’ll pack me something different.” And Joe told him a little white lie will come back to haunt him. And saying that made Joe feel, all like, morally superior, and so he added “I should totally run for a position on the middle school student government. They have room on the board and that way I could have a say about the lunches around here!”

The story was read as though naturally telling the story, slight errors (e.g., he made—he made a face), normal in speech were not removed. The story was designed to be easy to

comprehend (readability average =grade 6; Flesch-Kincaid Grade Level [Kincaid, Fishburne, Rogers, & Chissom, 1975] = 4.9, Gunning-Fog Score [Gunning, 1952] = 8.5, Coleman-Liau Index [Coleman & Liau, 1975] = 7.3, SMOG Index [McLaughlin, 1969] = 6, Automated Readability Index [Senter & Smith, 1967] = 4.3) but challenging to interpret. In addition to restructuring the sentences into ASL word order, which varies between Subject-Verb-Object and Object-Subject-Verb, places adjectives after the nouns, and explicitly indicates spatial information that is generally implicit in English narration, the text included 1) a garden path sentence: “And Joe told him a little white lie will come back to haunt him.”, 2) words that are homonyms in English but not in ASL (“run,” “room”) and in which the most frequent meaning in English would produce the incorrect interpretation in ASL (e.g., “room” in the story means availability, not a space in a building with walls), and 3) terms in which there is no direct translation (“morally superior,” “white lie,” “haunt,” “have a say”). All participants listened to the recording from an audio file played on a Macbook laptop and were video recorded with a Flip UltraHD video recorder.

The story was divided into 22 idea units. Each performance was scored on 5 areas: 1) idea unit presence, where each idea unit was scored from 0 (completely absent from the interpretation) to 4 (completely presented during the interpretation), 2) idea unit sequence, where the overall interpretation was scored on a scale of 0 (completely inappropriate to the target language idea unit sequencing) to 4 (completely appropriate to the target language idea unit sequencing), 3) target language structure, where each idea unit was scored on a scale of 0 (incomprehensible in ASL) to 4 (native/near native structure production), 4) accent, where the overall interpretation was scored on a scale of 0 (incomprehensible in ASL) to 4 (native/near native fluency), and 5) lexical errors, where a count of lexical errors (e.g., producing the sign for

the concept of a space in a building with walls instead of the sign for the concept of availability when interpreting “They have room on the board....” The first 4 scores were summed to create a composite simultaneous interpreting score.

Table 4.3

Order of Tasks

Order	Task	Administration time
Session I		
1	BIS scale (concurrent with BAS scale)	~ 2.5 minutes
1	BAS scale (concurrent with BIS scale)	~ 2.5 minutes
2	Pattern Comparison	practice + 1 minute
3	Letter Comparison	practice + 1 minute
4	Connections A: Numbers (concurrent)	practice + 40 seconds
4	Connections A: Letters (concurrent)	practice + 40 seconds
4	Connections B: Numbers-Letters (concurrent)	practice + 40 seconds
4	Connections B: Letters-Numbers (concurrent)	practice + 40 seconds
5	Raven's Advanced Progressive Matrices	practice + 10 minutes
6	Cattell's Culture Fair Test	practice + 12.5 minutes
Session II		
1	Automated Reading Span	practice + ~ 10 minutes
2	Automated Operation Span	practice + ~ 10 minutes
3	Letter-Number Sequencing	practice + ~ 7 minutes
4	Backward Digit Span	practice + ~ 7 minutes
5	Wisconsin Card Sorting test	~ 6 minutes
6	Simultaneous Interpreting Performance Task	~ 5 minutes

4.4 Results

Table 4.4 shows the means and standard deviations for each measure across the data collection waves. Tables 4.5 through 4.8 show the correlations among cognitive ability and affective trait measures for each wave. Figure 4.1 shows the individual simultaneous interpreting performance growth curves.

Further correlation analyses revealed that initial interpreting performance generally predicted final interpreting performance: idea unit presence (initial and final), $r = .66, p < .001$; idea unit structure (initial and final), $r = .61, p < .001$; target accent (initial and final), $r = .26, ns$; and lexical errors (initial and final), $r = .24, ns$. See Figure 4.1 for individual growth curves.

Table 4.4

Descriptive Statistics

Measure	Cognitive Ability / Affective Trait Measures							
	Wave 1		Wave 2		Wave 3		Wave 4	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
BIS	3.07	.45	3.12	.41	3.11	.42	3.08	.39
BAS	3.18	.28	3.17	.32	3.27	.30	3.25	.32
Perceptual								
Speed	.63	.08	.65	.11	.65	.09	.67	.10
ConnectA	.69	.14	.76	.14	.78	.13	.83	.12
ConnectB	.38	.09	.42	.11	.42	.10	.44	.11
Intelligence	.53	.13	.55	.14	.56	.15	.60	.14
WMC.Span	.72	.20	.74	.20	.74	.21	.74	.20
WMC.CT	.87	.06	.90	.06	.92	.04	.93	.05
WCST	.79	.09	.79	.10	.83	.07	.83	.05
Measure	Interpreting Performance Measures							
Idea Unit Presence	---	---	1.95	.59	2.25	.50	2.29	.45
Idea Unit Sequence	---	---	2.89	.57	3.06	.34	3.00	.00
Target Structure	---	---	1.49	.52	1.79	.54	1.79	.38
Target Accent	---	---	1.89	.79	2.21	.88	2.85	.70
Lexical Errors	---	---	0.41	.79	0.65	.95	.32	.53
Composite Measure	---	---	8.22	2.03	9.31	1.83	9.93	1.20

Note: BIS = Behavioral Inhibition System Sensitivity Scale. BAS = Behavioral Approach System Sensitivity Scale. Connect.A = Connections A Test. Connect.B = Connections B Test. WMC.Span = Working Memory Capacity: Maintenance + Interference. WMC.CT = Working Memory Capacity: Coordination + Transformation. WCST = Wisconsin Card Sorting Test.

Table 4.5

Correlations Among Cognitive Ability and Affective Trait Measures: Wave 1

	1	2	3	4	5	6	7	8
1. BIS	---							
2. BAS	-.01	---						
3. Perceptual Speed	-.08	-.18	---					
4. Connect.A	-.11	-.03	.69	---				
5. Connect.B	-.17	.06	.53	.81	---			
6. Intelligence	.08	.10	.49	.59	.50	---		
7. WMC.Span	-.08	-.22	.25	.33	.33	.40	---	
8. WMC.CT	-.12	-.09	.26	.36	.40	.45	.77	---
9. WCST	.14	.07	.19	.27	.30	.23	.34	.25

Note: BIS = Behavioral Inhibition System Sensitivity Scale. BAS = Behavioral Approach System Sensitivity Scale. Connect.A = Connections A Test. Connect.B = Connections B Test. WMC.Span = Working Memory Capacity: Maintenance + Interference. WMC.CT = Working Memory Capacity: Coordination + Transformation. WCST = Wisconsin Card Sorting Test. **Bold** = significant at the $p < .05$ level.

Table 4.6

Correlations Among Cognitive Ability and Affective Trait Measures: Wave 2

	1	2	3	4	5	6	7	8
1. BIS	---							
2. BAS	-.18	---						
3. Perceptual Speed	.02	-.07	---					
4. Connect.A	-.06	-.05	.57	---				
5. Connect.B	.11	-.05	.59	.81	---			
6. Intelligence	.00	.07	.54	.63	.72	---		
7. WMC.Span	-.11	-.01	.23	.45	.39	.46	---	
8. WMC.CT	-.04	.13	.38	.38	.36	.50	.69	---
9. WCST	.15	.29	.20	.19	.26	.48	.44	.39

Note: Same abbreviations as in Table 4.4. **Bold** = significant at the $p < .05$ level.

Table 4.7

Correlations Among Cognitive Ability and Affective Trait Measures: Wave 3

	1	2	3	4	5	6	7	8
1. BIS	---							
2. BAS	.11	---						
3. Perceptual Speed	.09	.07	---					
4. Connect.A	-.09	.05	.62	---				
5. Connect.B	.08	.22	.69	.84	---			
6. Intelligence	.23	.27	.46	.55	.57	---		
7. WMC.Span	-.06	.02	.31	.45	.38	.67	---	
8. WMC.CT	-.05	.03	.25	.37	.38	.60	.77	---
9. WCST	-.01	-.07	.23	.43	.31	.30	.47	.28

Note: = Same abbreviations as in Table 4.4. **Bold** = significant at the $p < .05$ level.

Table 4.8

Correlations Among Cognitive Ability and Affective Trait Measures: Wave 4

	1	2	3	4	5	6	7	8
1. BIS	---							
2. BAS	-.06	---						
3. Perceptual Speed	-.27	.07	---					
4. Connect.A	-.35	.03	.74	---				
5. Connect.B	-.13	.12	.69	.79	---			
6. Intelligence	-.12	.26	.54	.51	.34	---		
7. WMC.Span	-.10	.15	.27	.33	.27	.52	---	
8. WMC.CT	-.05	-.19	.37	.34	.38	.16	.56	---
9. WCST	-.12	-.21	.23	.24	.18	.25	.43	.37

Note: Same abbreviations as in Table 4.4. **Bold** = significant at the $p < .05$ level.

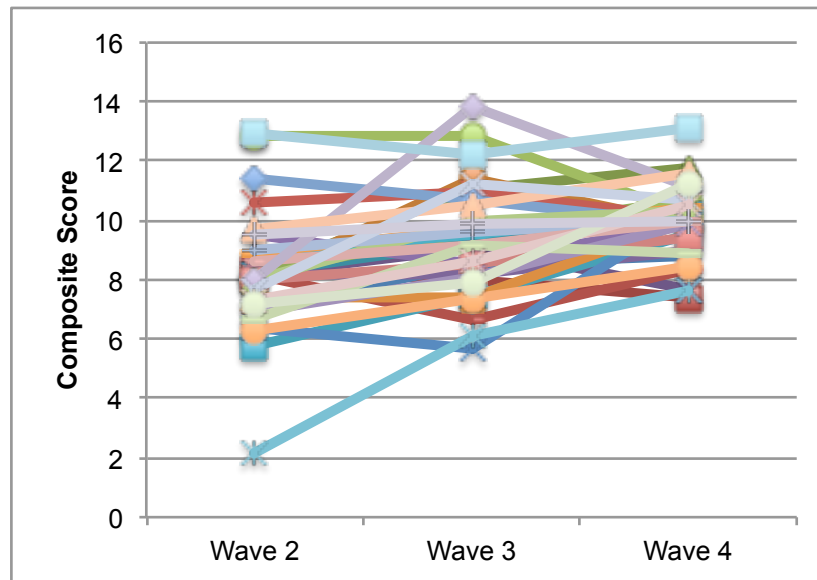


Figure 4.1. Interpreter performance growth curves.

Analyses

The bilingual advantage. In order to test change in cognitive abilities occurring from interpreter training with a larger sample than was used in Chapter 3 of this document, a series of dependent sample t-tests were conducted in order to determine if changes from the initial to the final wave were significant. Neither affective measure—behavioral inhibition system sensitivity and behavioral approach system sensitivity—changed significantly over the course of interpreter training, $ps > .08$. In contrast, changes in the cognitive ability measures were similar to the results reported in Chapter 3. Advanced interpreter students significantly outperformed themselves as beginning interpreters on measures of perceptual speed, $t(33) = 3.66, p = .001, d = 0.63$; psychomotor speed, $t(33) = 10.38, p < .001, d = 1.78$; task switching speed, $t(33) = 5.83, p < .001, d = 1.00$; mental flexibility, $t(33) = 2.66, p < .05, d = .46$; and fluid intelligence, $t(33) = 4.99, p < .001, d = 0.86$, see Figure 4.2. The pattern of results regarding measures of working memory capacity were also consistent with the findings reported in Chapter 3: advanced

interpreter students outperformed themselves as beginning interpreter students on measures of one's ability to coordinate and transform maintained information (coordination and transformation measures of working memory capacity), $t(33) = 5.99, p < .001, d = 1.03$, but not on measures of one's ability to maintain information in the face of interference (maintenance and interference measures of working memory capacity) $t(33) < 1.00, ns, d = 0.15$, see Figure 4.3.

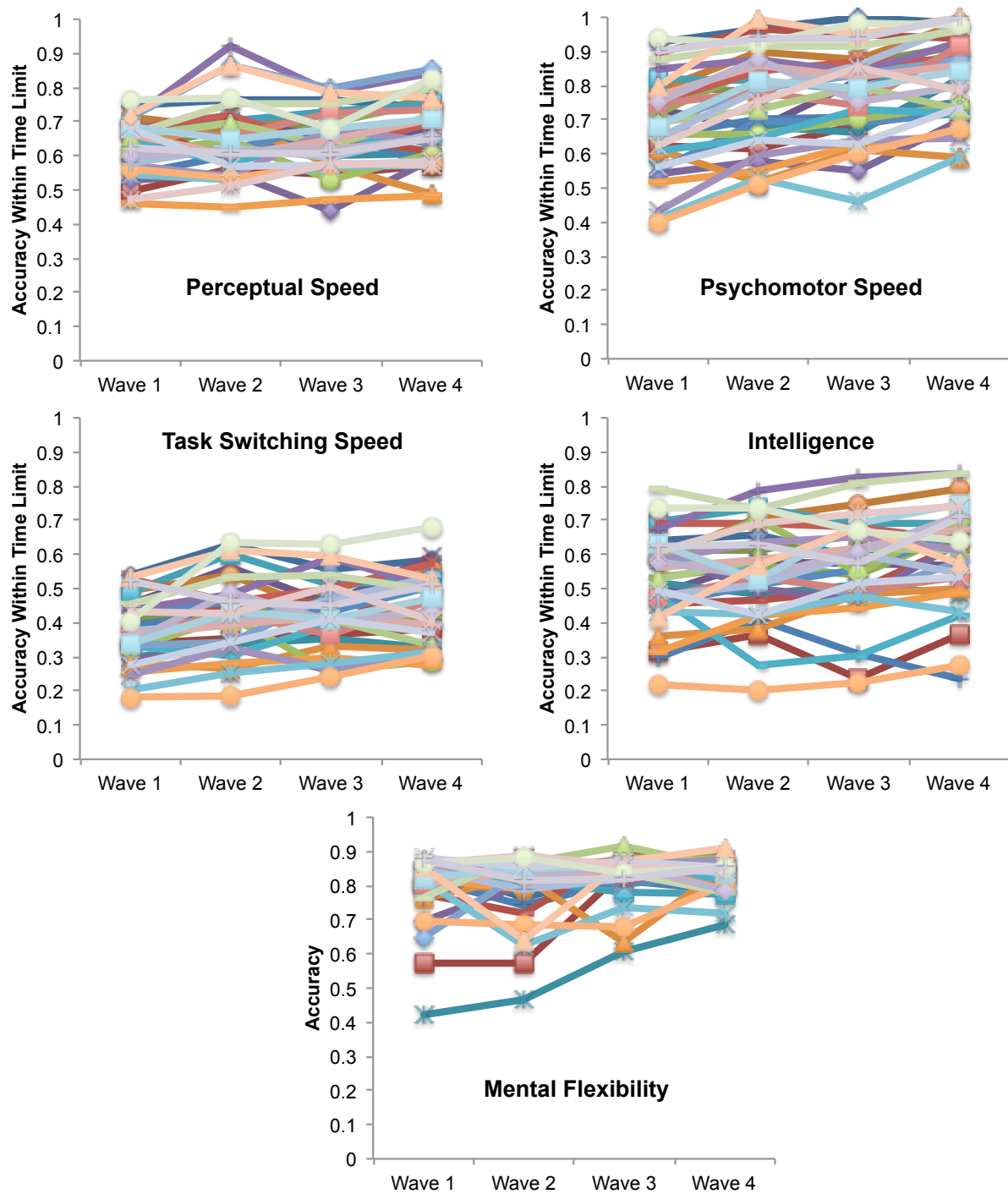


Figure 4.2. Cognitive ability growth curves.

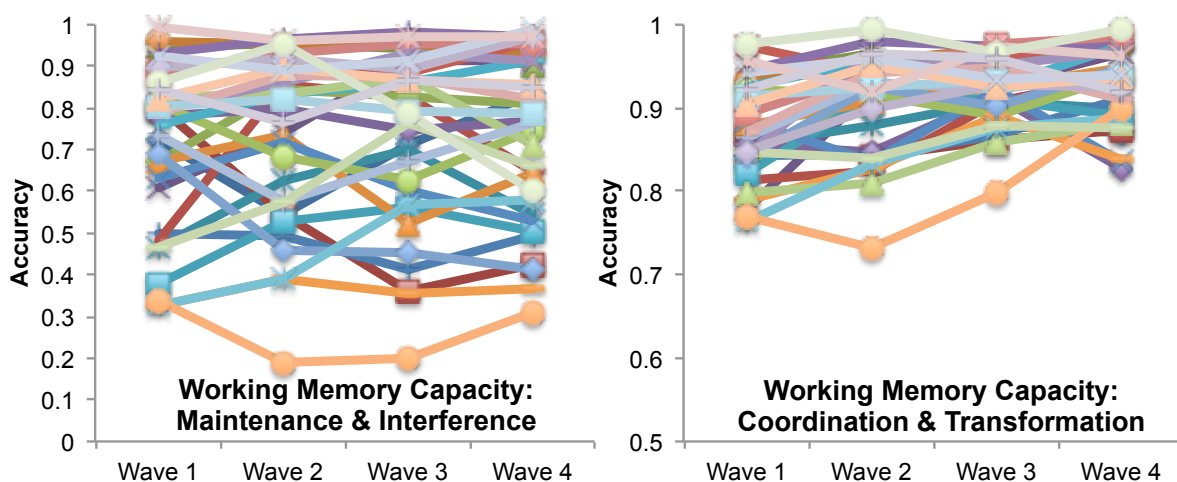


Figure 4.3. Working memory capacity growth curves.

Initial cognitive abilities as predictors of initial and final performance. In order to test whether initial cognitive abilities predicted interpreter performance, correlation and regression analyses were run. Table 4.9 shows the correlation coefficients between initial cognitive abilities and initial interpreting performance. Table 4.10 shows the correlation coefficients between initial cognitive abilities and final interpreting performance.

The correlations revealed that several cognitive abilities and affective traits were important for initial and final interpreting performance. Specifically, for idea unit presence one's psychomotor speed was important initially but was less important by the final measure, whereas intelligence and both types of working memory capacity were important predictors of initial idea unit presence performance and even stronger predictors of final idea unit presence performance.

For the general target language sequence, both types of working memory capacity were important initially, but variance in performance on this measure converged for the final performance to the point where there was no variance.

Table 4.9

Correlations Between Initial Cognitive Abilities/Affective Traits and Initial Interpreting Performance

	Idea Presence	Target Order	Idea Structure	Target Accent	Lexical Errors	Composite Score
BIS	.195	-.033	.120	-.079	.116	.049
BAS	-.016	-.048	-.170	.079	-.374*	-.090
Perceptual						
Speed	.277	.208	.225	.190	-.314 [†]	.254
ConnectA	.349*	.253	.233	.116	-.092	.277
ConnectB	.213	.193	.112	.183	.042	.218
Intelligence	.334*	.268	.193	.014	-.162	.216
WMC.Span	.472**	.409*	.467**	.198	.074	.442**
WMC.CT	.395*	.359*	.337*	.302 [†]	.110	.415*
WCST	.083	-.009	-.001	-.104	-.273	-.020

Note. BIS = Behavioral Inhibition System Sensitivity Scale, BAS = Behavioral Approach System Sensitivity Scale, ConnectA = Connection A Tests, ConnectB = Connection B Tests, WMC.Span = Working memory capacity: maintenance and interference tasks. WMC.MM = Working memory capacity: coordination and transformation tasks. WCST = Wisconsin Card Sorting Test. Bold = significant at $p \leq .05$. * = $p \leq .05$. ** = $p \leq .01$.

Table 4.10

Correlations Between Initial Cognitive Abilities/Affective Traits and Final Interpreting Performance

	Idea Presence	Target Order	Idea Structure	Target Accent	Lexical Errors	Composite Score
BIS	.117	---	.291	-.036	.126	.116
BAS	.062	---	-.099	.440**	.056	.250
Perceptual						
Speed	.291	---	.238	.177	-.136	.290
ConnectA	.331 [†]	---	.195	.078	.027	.233
ConnectB	.175	---	-.029	.189	.079	.167
Intelligence	.507**	---	.393*	.083	.085	.366*
WMC.Span	.526**	---	.398*	.269	-.042	.484**
WMC.CT	.524**	---	.328 [†]	.398*	.148	.536**
WCST	.119	---	.221	.429**	-.101	.366*

Note. BIS = Behavioral Inhibition System Sensitivity Scale, BAS = Behavioral Approach System Sensitivity Scale, ConnectA = Connection A Tests, ConnectB = Connection B Tests, WMC.Span = Working memory capacity: maintenance and interference tasks. WMC.MM = Working memory capacity: coordination and transformation tasks. WCST = Wisconsin Card Sorting Test. Bold = significant at $p \leq .05$. * = $p \leq .05$. ** = $p \leq .01$. [†] = $p \leq .08$. All interpreter students scored the same in their target order performance, thus there was no variance and correlation coefficients could not be computed.

When producing the target language structure, intelligence did not predict performance initially, while both measures of working memory capacity did. However, by the final measure of performance, intelligence was a strong predictor of target language structure performance,

while the relationship between both types of working memory capacity and target language structure performance weakened from initial to final performance.

Predictors of the interpreters' accent fluency changed the most from initial performance to final performance. No cognitive abilities or affective traits were significant predictors at initial performance (although working memory capacity: coordination and transformation was a marginally significant predictor). However initial BAS, working memory: coordination and transformation, and mental flexibility were all significant predictors of one's accent fluency in ASL during the final performance.

BAS (significantly) and perceptual speed (marginally significantly) negatively predicted lexical errors, that is, the higher one's behavioral approach system sensitivity and the better one's perceptual speed, the fewer the lexical errors. However, many interpreter students did not produce lexical errors at any point.

Finally, the composite measure—the sum of all the interpreter measures except the error count—was initially predicted by both types of working memory capacity. At final performance, intelligence, both types of working memory capacity, and mental flexibility predicted the final composite measure.

Given that several of these cognitive abilities correlate with one another, the question remains whether each of the initial cognitive abilities predicting interpreter performance are contributing unique variance or are accounting for the same variance as other cognitive abilities. In order to answer this question, hierarchical regression analyses were conducted. First, a full model (all initial cognitive and affective measures predicting final performance) was compared to a partial model (all initial cognitive and affective measures with significant or marginally significant correlations with final performance predicting final performance). All initial cognitive

abilities and affective measures combined accounted for 55% of the variance in final interpreting performance, $F(9, 24) = 3.27, p = .010, R^2 = .551$. These regression models were compared to models that only included the measures that significantly or marginally significantly correlated with final interpreting performance. The change in R^2 was not significant, $\Delta R^2 = .192, F(5, 24) = 2.05, p = .108$, indicating that the four initial measures that significantly or marginally significantly correlated with final interpreting performance accounted for 36% of the 55% of the transfer performance variance explained.

Multiple hierarchical regression models was constructed comparing a model in which all the initial cognitive ability measures that correlated with performance were taken into account (partial model), with models where a single one of these variables was excluded in turn in order to test whether the difference between the models was significant, that is, whether the excluded measure accounted for significant unique variance. See Table 4.11.

Fluid intelligence (“Intelligence”), both measures of working memory capacity (“WMC.Span” “WMC.CT”) and mental flexibility (“mental flexibility”) accounted for about 36% of the variance in final interpreting performance. However, the predictive power of these measures was shared, that is, none of these measures accounted for a significant amount of unique variance.

Table 4.11

Summary of Regression Analyses, Final Interpreting Performance

Outcome = Final Interpreting Performance						
Partial Model						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
Intelligence	1.104	1.536	0.72	.121	.478	
WMC.Span	0.494	1.448	0.34	.081	.736	
WMC.CT	7.455	4.856	1.54	.364	.136	
WCST	2.796	2.028	1.38	.219	.179	
						.359
Excluding Intelligence						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
WMC.Span	0.576	1.432	0.40	.095	.690	
WMC.CT	8.281	4.680	1.77	.405	.087	
WCST	2.955	1.999	1.48	.232	.150	
						.322
$\Delta R^2 = .011, F(1, 29) = .52, p = .478$						
Excluding WMC.Span						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
Intelligence	1.145	1.508	0.76	.125	.454	
WMC.CT	8.621	3.398	2.54	.421	.017	
WCST	2.954	1.945	2.151	.232	.139	
						.357
$\Delta R^2 = .002, F(1, 29) = .116, p = .736$						
Excluding WMC.CT						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
Intelligence	1.662	1.525	1.09	.182	.285	
WMC.Span	2.059	1.052	1.96	.339	.060	
WCST	2.668	2.071	1.29	.209	.208	
						.307
$\Delta R^2 = .052, F(1, 29) = 2.36, p = .136$						
Excluding WCST						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
Intelligence	1.336	1.549	0.86	.146	.395	
WMC.Span	0.950	1.431	0.66	.156	.512	
WMC.CT	7.181	4.924	1.46	.351	.155	
						.317
$\Delta R^2 = .042, F(1, 29) = 1.901, p = .179$						

Notes. Partial Model = all initial measures whose correlation coefficients were significant or marginally significant with final interpreting transfer performance.. Intelligence = Composite measure of fluid intelligence. WMC.Span = Composite measure of working memory capacity: Maintenance and interference. WMC.CT = Composite measure of working memory capacity: coordination and transformation. WCST = Wisconsin Card Sorting Test.

Change in cognitive abilities as predictors of performance. In order to test whether change in cognitive abilities predicted performance, three sets of correlation analyses were conducted. The first set of correlations tested whether early change—change in cognitive abilities from Wave 1 to Wave 2—predicted early interpreter performance in Wave 2. The second set of correlations tested whether change in cognitive abilities from Wave 1 to Wave 3 predicted interpreter performance in Wave 3. The third set of correlations tested whether change in cognitive abilities from Wave 1 to Wave 4 predicted final interpreter performance.

No significant relationships emerged in the first or second set of correlations. The only significant correlation was between change in working memory capacity: coordination and transformation from Wave 1 to Wave 4 and final interpreting performance, $r = -.395$. Inspection of scatterplots revealed a ceiling effect. Those who started close to 100% on the initial measure of working memory capacity had the least amount of change in working memory capacity from Wave 1 to Wave 4 and that initial working memory capacity positively predicted both working memory capacity at Wave 4 and interpreting performance at Wave 4.

However, as shown in Tables 4.10 and 4.12, final cognitive ability measures predicted interpreting performance in some cases that initial cognitive abilities did not and were not predictive in some cases where initial cognitive abilities were. Differences in perceptual speed measured in the final wave was predictive of both idea presence performance and target language idea unit structure performance although initial perceptual speed did not correlate with any measure of final interpreter performance.

Table 4.12

Correlations Between Final Cognitive Abilities/Affective Traits and Final Interpreting Performance

	Idea Presence	Target Order	Idea Structure	Target Accent	Lexical Errors	Composite Score
BIS	-.114	---	.022	-.037	.149	-.058
BAS	.246	---	.022	.237	.007	.239
Perceptual						
Speed	.395*	---	.435	.028	-.080	.304 [†]
ConnectA	.360*	---	.226	.069	.039	.248
ConnectB	.193	---	.141	.171	.012	.218
Intelligence	.483**	---	.385*	.177	-.153	.409*
WMC.Span	.362**	---	.206	.279	-.019	.366*
WMC.CT	.200	---	.234	.175	.059	.253
WCST	.112	---	.243	.143	-.227	.204

Note. IUP = Idea Unit Presence. IUS = Idea Unit Sequence. TLS = Target Language Structure. AF = Accent Fluency. LE = Lexical Errors. BIS = Behavioral Inhibition System Sensitivity Scale. CFS = Composite Fluency Score. BAS = Behavioral Approach System Sensitivity Scale. ConnectA = Connection A Tests, ConnectB = Connection B Tests, WMC.Span = Working memory capacity: maintenance and interference tasks. WMC.MM = Working memory capacity: coordination and transformation tasks. WCST = Wisconsin Card Sorting Test. **Bold** = significant at $p \leq .05$. [†] = $p \leq .08$. * = $p \leq .05$. ** = $p \leq .01$.

These findings are somewhat consistent with Ackerman's (1988) theory of complex skill acquisition: in the early phase, general cognitive abilities are most predictive, in the middle phase, general cognitive abilities become less predictive and perceptual speed becomes highly predictive of performance, and in the final phase, only non-cognitive psychomotor ability (i.e., "muscle memory") is predictive of performance. Ackerman (1988) also suggests that consistency moderates this relationship such that when performing an inconsistent task, general cognitive abilities will remain predictive and neither perceptual speed nor non-cognitive psychomotor ability will become predictive, that is, skill acquisition will remain in the first stage. The current results and the results of Chapter 2 do not fully support this aspect of Ackerman's (1988) theory. In Chapter 2, the measure of performance was real-world observations of simultaneous interpreting across contexts and is therefore not highly consistent. Both general cognitive abilities and perceptual speed differentiated experts from non-experts at the final phase of skill acquisition. The results in the current study show that 1) Wave 4 measures of perceptual speed is

predictive in some aspects of performance though Wave 1 measures of perceptual speed were not, 2) the explanatory power of working memory seems to be reduced from initial measures to final measures on final performance, and 3) the role of intelligence remains predictive from Wave 1 to Wave 4 on final performance.

In the current study, an argument could be made that the performance task was consistent, given that the measure was repeated four times over two years, if one assumes that the measure is consistent, and that advanced students are in the middle stage of skill acquisition (that is, not yet at the final expert level, but beyond the initial stages), the results should show that the explanatory power of general cognitive abilities asymptotes while perceptual speed is highly predictive. Working memory capacity: coordination and transformation reveals a ceiling effect so higher end variance is lost and the actual explanatory power of working memory capacity: coordination and transformation is unknown. Also, if the task is indeed consistent, then the explanatory power of intelligence should reduce by Wave 4, but it does not (Wave 1 correlation with transfer performance, $r = .366$; Wave 4 correlation with transfer performance, $r = .409$). If one assumes that simultaneous interpreting in general is not consistent and that the measure reflected general simultaneous interpreting performance, then general cognitive abilities should remain consistent and perceptual speed should not become predictive. The results suggest that the role of consistency may indeed be a moderator of the relationships between predictors of skill acquisition and performance, but might not fully explain changes to the relationship.

Effect of training on performance. In order to test whether amount of accumulated time in the training program predicted interpreter performance, a MANCOVA was conducted. Wave (i.e., 20%, 40%, 60% or 80% program completion) was entered as the predictor variable and subject was entered as the covariate. The MANCOVA revealed that subject was a marginally

significant covariate, $F(6, 93) = 2.12, p = .06$, partial $\eta^2 = .12$. Wave was a significant predictor of performance, $F(12, 188) = 3.22, p < .001$, partial $\eta^2 = .17$.

4.5 Discussion

Similar to the results of the study described in Chapter 2, cognitive abilities were important predictors of simultaneous interpreting performance. However, unlike the study in Chapter 2, which investigated predictors of simultaneous interpreting expertise, accumulated amount of domain-specific experience also predicted simultaneous interpreting performance. These results suggest that initial training and experience are likely more important early in skill acquisition as an individual develops basic performance competency but is less likely to predict performance variance at high levels of performance.

The results of the current study corroborated the results described in Chapter 3: accumulated experience in a task with high bilingual management demands—simultaneous interpreting—enhanced cognitive abilities. Both initial cognitive abilities and cognitive abilities that had become enhanced predicted interpreter performance. The amount of change in one's cognitive abilities did not predict interpreting performance. Rather, the absolute ability was a better predictor, especially after one had accumulated training. Working memory predicted initial interpreting performance, while working memory, intelligence, and mental flexibility strongly predicted interpreting performance at the end of training.

Several cognitive abilities implicated in discriminating experts from non-experts revealed in Chapter 2 were not implicated in beginners' performance. However, the cognitive abilities that discriminated experts from non-experts were the same cognitive abilities that became enhanced with training. The two sets of results suggest a complex pattern of performance attainment: 1)

certain cognitive abilities are important for initial learning of simultaneous interpreting, 2) cognitive abilities recruited during simultaneous interpreting become enhanced with domain-specific training and experience, and 3) the predictive power of cognitive abilities change over the stages of skill acquisition such that predictors of learning or not necessarily the same predictors of expert performance.

The hypothesis that change in cognitive abilities would predict variance in interpreting performance was not supported. Neither the compensatory hypothesis nor the “rich get richer” hypothesis were completely supported. Training predicted interpreting performance and performance variance converged somewhat with training, however, cognitive abilities predicted performance and initial performance predicted final performance. These results suggest that both training and cognitive abilities are important in early skill acquisition of simultaneous interpreting, and that while training compensates for initial performance variance, one’s interpreting performance relative to others remains somewhat stable with training.

Chapter 5

A novel moderator variable: Predictability of the task environment

5.1 Chapter summary

Two laboratory studies manipulated predictability of the task environment and measured predictors of performance⁷.

- Participants were randomly assigned to a high, moderate, or low level of predictability of the task environment.
- Data from cognitive ability measures and accumulated domain-specific practice was collected.
- Growth curves and learning transfer were investigated.
- In the highly predictable condition, participants demonstrated the steepest growth curves and highest final performance.
- Individual differences in cognitive abilities correlated with performance in moderate predictability conditions.

⁷ This study is being prepared for publication (Macnamara, Drach, Chow, & Conway, in prep).

5.2 Introduction: A laboratory measure of skill acquisition

The studies presented thus far have focused on a single performance domain—simultaneous ASL-English interpreting. Simultaneous interpreting offers a rich task environment for testing multiple hypotheses on the inter-relationships among domain-specific experience, general cognitive abilities, and performance variance. The results of these studies demonstrate the importance of general cognitive abilities predicting skill acquisition and expertise. However, research in other domains, such as chess, has demonstrated that general cognitive abilities, at least intelligence, are not highly predictive of performance variance (e.g., Bilalić & McLeod, 2007). This finding may be due in part to researchers only testing intelligence instead of a range of cognitive abilities, and/or this finding may be due to qualitative differences in the task environment of chess compared to simultaneous interpreting and the associated demands of these task environments on cognition.

Predictability of the task environment

This section describes a new classification system for traits of the task environment. The traits of the task environment will explain the task demands. Therefore, the traits of the task environment may explain variance in predictors of performance.

Lability. Some tasks' stimuli to which a performer responds are static. That is, there is not strong time pressure to respond to the stimuli and therefore the performer can take time to analyze the stimuli and consider multiple strategies for responding to the stimuli. In these cases, the performer can apply learned strategies when planning an appropriate response. Examples of

tasks with static stimuli at the time of response planning and execution are chess playing, darts playing, reading a book, bowling, and mirror tracing.

Some tasks' stimuli to which a performer responds are spatially or temporally dynamic. That is, the performer must continuously update the contents of working memory as the stimuli are changing, whether due to change in location or motion (spatially dynamic), or due to ephemeral stimuli (temporally dynamic). The performer must also coordinate the timing of the inputs with the changing inputs and decide when to initiate the execution keeping in mind that the inputs may change from the beginning of the executed response to the end of the performer's response action. In these cases, the dynamic stimuli place more demands on executive functions than when the task is static: the performer must update the contents of working memory and coordinate these updates with execution planning. Examples of tasks with highly dynamic stimuli at the time of response planning and execution are hockey (spatial), a pass by a quarterback as players on the field are in motion (spatial), sight-reading music (temporal), and simultaneous interpreting (temporal).

Given that, all else being equal, a task with dynamic stimuli places more demands on executive functions compared to a task with static stimuli, individual differences in executive functions should more strongly predict performance in the task with dynamic stimuli during performer planning and execution than in the task with static stimuli at the time of performer planning and execution.

Constraints. Some tasks' stimuli are drawn from a pool with distinct boundaries. That is, with experience, one can encounter all the potential stimuli of the task and all the appropriate responses to those stimuli. In these cases, the more experience a performer has with the stimuli, the more likely she is to have encountered a given stimuli and to prepare and execute an

appropriate/skilled response. Examples of tasks with highly constrained sets of stimuli are chess (at least beginning and end game), darts, bowling, and mirror tracing.

Some tasks' stimuli are drawn from a pool without distinct boundaries. That is, there are an unknown number of potential stimuli and associated responses. Only a limited number of stimuli are repeated and thus performance depends not on rehearsed responses, but on assessing the present stimuli and determining an appropriate response. In these cases, the task demands more reasoning and problem solving than if the responses could be automatically executed from previous experience with the same stimuli. Examples of tasks with less constrained set of stimuli are reading a book, police investigation, and simultaneous interpretation.

Given that, all else being equal, a task whose pool of stimuli are drawn from less constrained set places more demands on executive functions compared to a task whose pool of stimuli are drawn from a highly constrained set, individual differences in executive functions should more strongly predict performance in the task with a less-constrained stimuli set than in the task with a more-constrained stimuli set.

Predictability. Both lability of the stimuli and constraints of the stimuli set can be thought of as aspects of predictability. If stimuli are static during performer response, then the performer can perfectly predict the stimuli's arrangement during response execution. That is, the performer can plan his response with the arrangement of the present stimuli. For example, a chess player can execute an action based on the present configuration of the board as he reaches for the chess piece he will move. The configuration of the other pieces on the board will not rearrange as he is moving his chess piece.

If the stimuli are highly dynamic during performer response, then the performer cannot perfectly predict the stimuli's arrangement during response execution; the arrangement of the

stimuli will have changed from when the performer begins to execute the response to when the execution of the response is complete. That is, the performer must plan his response based on reasoning about the arrangement of the stimuli at the completion of the executed response. For example, a quarterback attempting a pass in football cannot choose an appropriate pass based on the present arrangement of the players on the field as he begins his throwing motion. He must reason about where his eligible receivers and the defensive coverage will be as the ball reaches its final position—his choice in pass timing, speed, and trajectory cannot be based on the present configuration of the players on the field and is thus not perfectly predictable from start of execution to completion of execution.

If stimuli are drawn from a limited pool, then the upcoming stimuli can be learned and are therefore predictable. In this case, the performer can practice the best response for any given stimuli and know if it will be encountered again. For example, a master chess player can predict, within a very limited set, the likely upcoming board configurations. There are a limited number of possible legal configurations at any given time and an even more limited number of likely configurations. The chess player can practice recognizing the board configuration and executing the best move based on the configuration. The numerable nature of the stimuli allows a computer to calculate the best move by predicting the upcoming configurations. A computer can outperform a human at chess.

If stimuli are drawn from a pool without clear boundaries, then not every potential stimulus can be practiced prior to performance and is therefore less predictable than if the stimuli were only drawn from a limited set. For example, in a police investigation an individual's means, motive, and opportunity must be constructed from witness statements, alibies, political and social relationships, and planned and chance circumstances. There are an almost unlimited number of

potential factors that must be considered when planning the next step of the investigation. The factors are innumerable and thus a computer cannot solve a crime better than a human detective.

A task that is both static at the time of execution and whose stimuli are drawn from a limited set is highly predictable. Individual differences in experiential factors, such as practice, should best explain individual differences in performance: individuals should improve on a log-linear scale and thus more skilled individuals should have more experience with the task than individuals who are less-skilled performers. In contrast, a task that is both highly dynamic at the time of execution and whose stimuli are drawn from a set with low constraints has little predictability. Individual differences in executive functions, such as updating, reasoning, and problem solving, should best explain individual differences in performance: individuals with higher cognitive abilities from which performance on the task demands, should perform the task better (after an initial phase of task learning) than individuals with lower cognitive abilities, even if those with lower cognitive abilities have more experience with the task.

Using this framework, one can examine the task of simultaneous interpreting. The task environment of simultaneous interpreting is highly dynamic. The source language message is ephemeral, that is, the interpreter must continuously update the contents of working memory in order to perform the task of interpreting, and has little control over the pace at which the information is presented. The task environment also has low constraints, that is, the content and context of the source language message could be of any content within any context in which human communication can occur. The interpreter is often working in novel environments and interpreting novel utterances. The simultaneous interpreting task environment is not highly predictable for these two reasons. That is, the interpreter cannot easily predict an upcoming utterance from moment to moment, nor can she predict how the communication will change

from the time she begins to plan her response from the time she has executed her response (attempted to produce the target message). Her initial understanding of the topic under discussion might be altered while she is performing. Neither future nor current stimuli are highly predictable. Thus, simultaneous interpreting is a task with low predictability. Individual differences in cognitive abilities recruited by performing the task should explain a substantial amount of performance variance; individual differences in amount of experience should not explain much performance variance (after the initial phase of task learning).

The predictability of the simultaneous interpreter's task environment can be contrasted with the task environment of the spoken language translator. Translators work with texts: While a translator is not likely to be able to predict the upcoming words on the page, the source language message is not temporally dynamic and she can close her eyes and perfectly predict words just read as they will not have changed from when she closed her eyes to when she opened them. The task environment of translation is more predictable than the task environment of simultaneous interpreting. Thus, translation is a task with moderate predictability. Individual differences in cognitive abilities recruited by performing the task should explain a moderate amount of performance variance; individual differences in amount of experience should also explain a moderate amount performance variance (after the initial phase of task learning).

The predictability of the task environment of chess differs from the predictability of the task environments of both simultaneous interpreting and translation. The task environment of chess is completely static at the time of performer response, that is, the board configuration does not change while the player is planning and executing an action, he could close his eyes on his turn and perfectly predict the board configuration when he opens his eyes as it will not have changed (barring a cheating opponent!) The task environment in chess is also highly constrained,

that is, there are numerable immediate future moves the chess player's opponent can play at any time and with accumulated knowledge the player can predict, within a small range, the probable upcoming chess move of his opponent. Thus, chess is a task with high predictability. Individual differences in cognitive abilities should not explain much performance variance; individual differences in amount of experience should explain a substantial amount of performance variance.

The purpose of the present study was to test whether predictability of the task environment moderates the relationships between 1) individual differences in cognitive abilities and skill acquisition, and 2) accumulated amount of domain-specific experience and skill acquisition. More specifically, the predictability of the task environment hypothesis consists of three sub-hypotheses: 1) In highly predictable task environments, that is, when the task environment is static at the time of performer-response and the range of potential future stimuli are delimited (i.e., when the problem space is highly constrained⁸), accumulated domain-specific practice will more strongly account for performance variance compared with general cognitive abilities. 2) In less predictable task environments, that is, when the task environment is highly dynamic at the time of performer-response and the range of potential future stimuli are not easily numerable (i.e., the problem space is not highly constrained), individual differences in general cognitive abilities will more strongly account for performance variance compared with accumulated domain-specific experience.

⁸ Predictability of the task environment is a novel moderating variable, although studies of skill acquisition have investigated the role of stimuli consistency (e.g., Ackerman, 1988; Schneider & Fisk, 1982), which overlaps with the constraints component of predictability.

5.3 Predictability Experiment 1

5.3.1 Methods

Participants

Eighty-six undergraduate students participated in exchange for partial course credit. Participants were randomly assigned to one of four between-groups conditions and then completed two within-groups conditions.

Procedure and Materials

Zombie Apocalypse. All participants completed two variants of the experimental task—“Zombie Apocalypse.” Participants were given a booklet with a backstory and instructions. The backstory told participants that they would be completing a task and in the world of the task scientists were testing cancer vaccines on rabbits when something went terribly wrong—and they accidentally created an army of zombie bunnies (see Figure 5.1, panel A). The participant is told that he (his avatar) is a cute, innocent, unaffected bunny trying not to become a zombie bunny (see Figure 5.1, panel B) and that the virus is airborne, so the participant’s goal is to try to keep the greatest average distance from all four zombie bunnies by using direction keys on the keyboard.

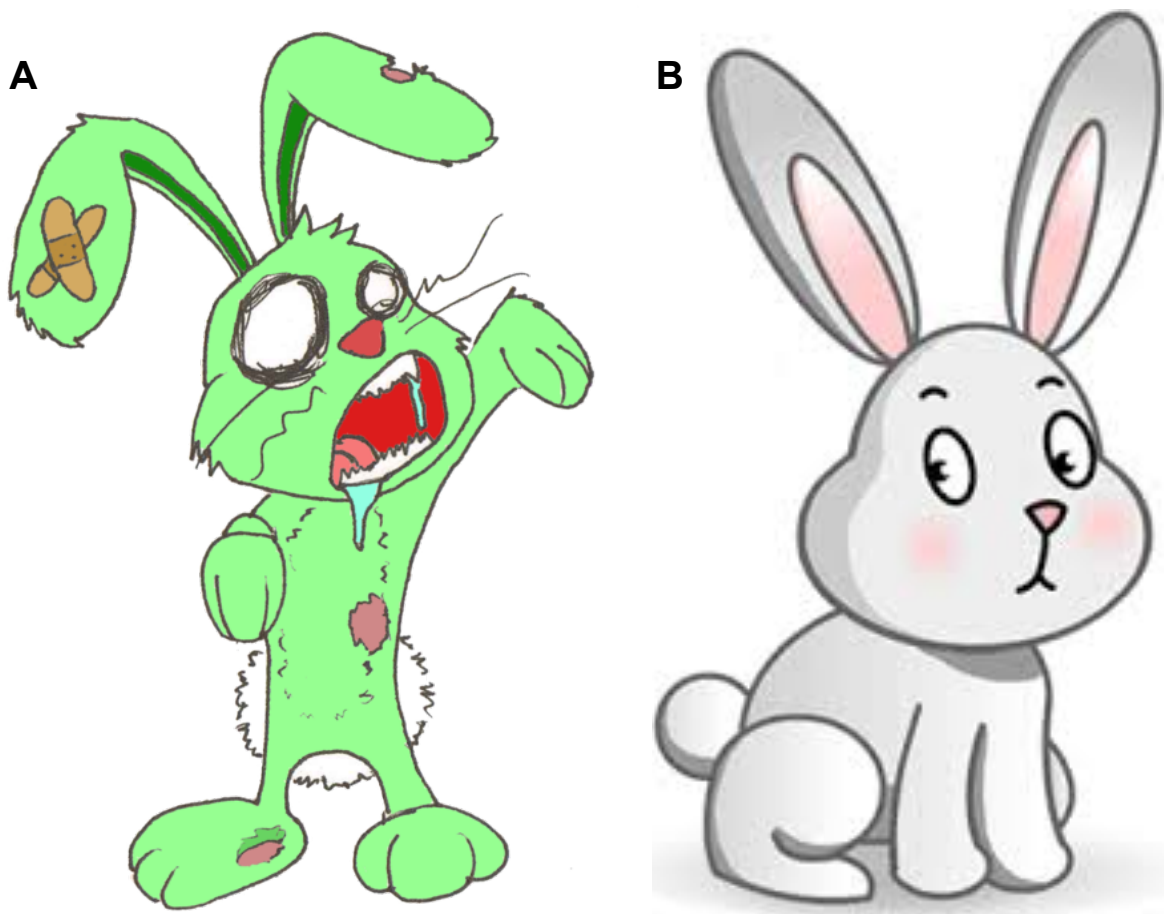


Figure 5.1. Stimuli for experimental task “Zombie Apocalypse.” Participants controlled an avatar (panel B) and attempted to avoid “zombie bunnies” (e.g., panel A).

All participants completed five performance blocks with four interleaved deliberate practice blocks (see Figure 5.2). In each performance block, the participant was tasked with trying to keep the greatest distance from all zombie bunnies at a time, and to keep his avatar moving. The participant accumulated points based on the average distance from all the zombie bunnies for each move taken. The participant’s goal was to accumulate as many points as

possible. The points per move were shown directly below the accumulated points so that the participant received immediate feedback on his performance at all times.

In the first deliberate practice block, between the first and second performance blocks, participants were told how *Zombie #1* moves (e.g., left to right across screen, up one, right to left across screen, up one, and so on; randomly) and the participant practiced keeping the greatest distance from *Zombie #1*. In the second deliberate practice block, between the second and third performance blocks, participants were told how *Zombie #2* moves (e.g., bottom to top of screen, one to the left, top to bottom across screen, one to the left, and so on; randomly) and the participant practiced keeping the greatest distance from *Zombie #2*, then practiced avoiding both *Zombie #1* and *Zombie #2*. In the third deliberate practice block, the third zombie was introduced, participants practiced avoiding *Zombie #3*, and then *Zombies #1, #2, and #3*, and in the fourth deliberate practice block, the fourth zombie was introduced, participants practiced avoiding *Zombie #4*, and then practiced avoiding all the zombies before the final performance block. See Figure 5.2.

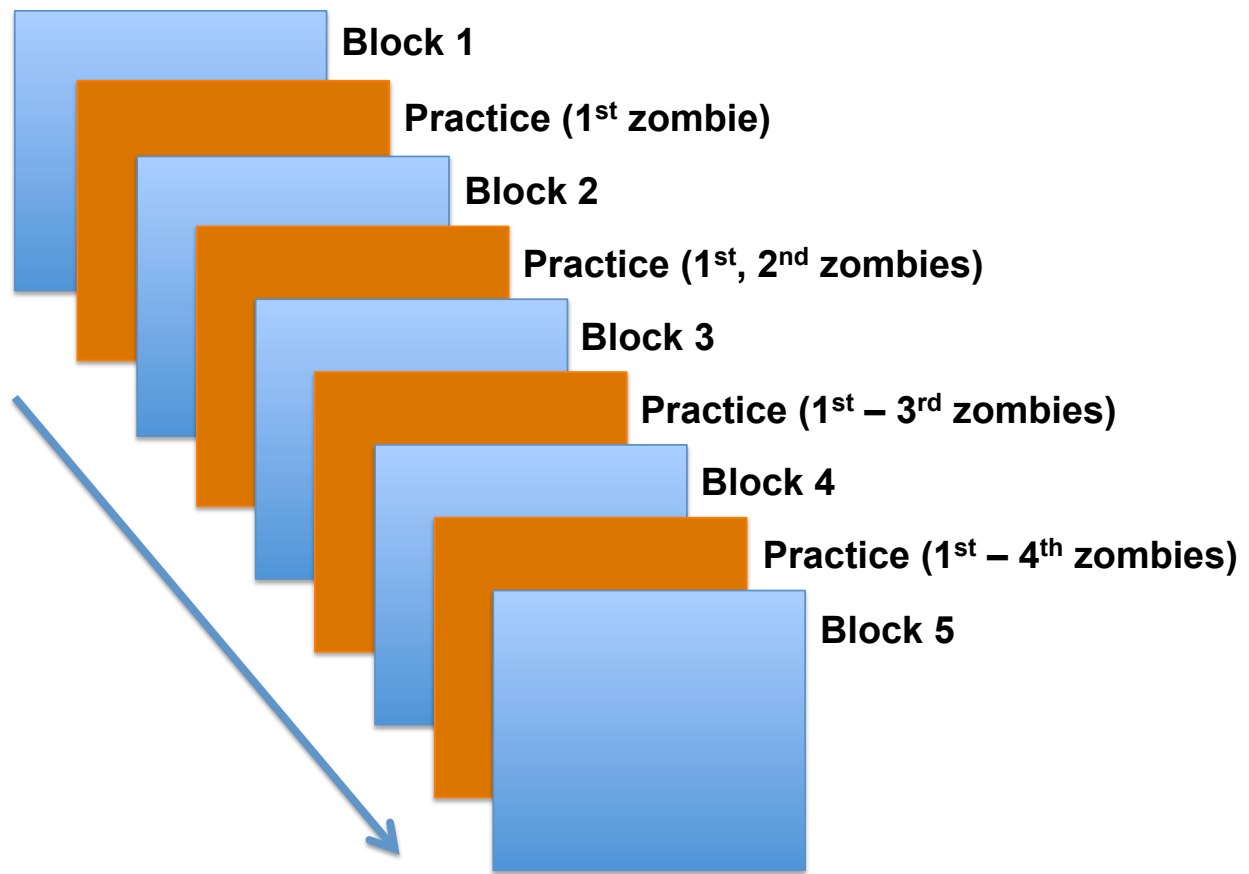


Figure 5.2. Components and sequence of the experimental task “Zombie Apocalypse.”

Four versions of the experimental task were created in order to manipulate 1) lability, that is, the rate of change of the task environment (stimuli were static at the time of response, “static”, or in constant motion, “dynamic”), and 2) constraints, that is, the range of potential stimuli to which a participant is expected to respond (stimuli moved in patterns, “pattern”, or moved pseudo-randomly, “random”). Therefore, the four versions of the task were 1) pattern-static, 2) pattern-dynamic, 3) random-static, and 4) random-dynamic. See Figure 5.3.

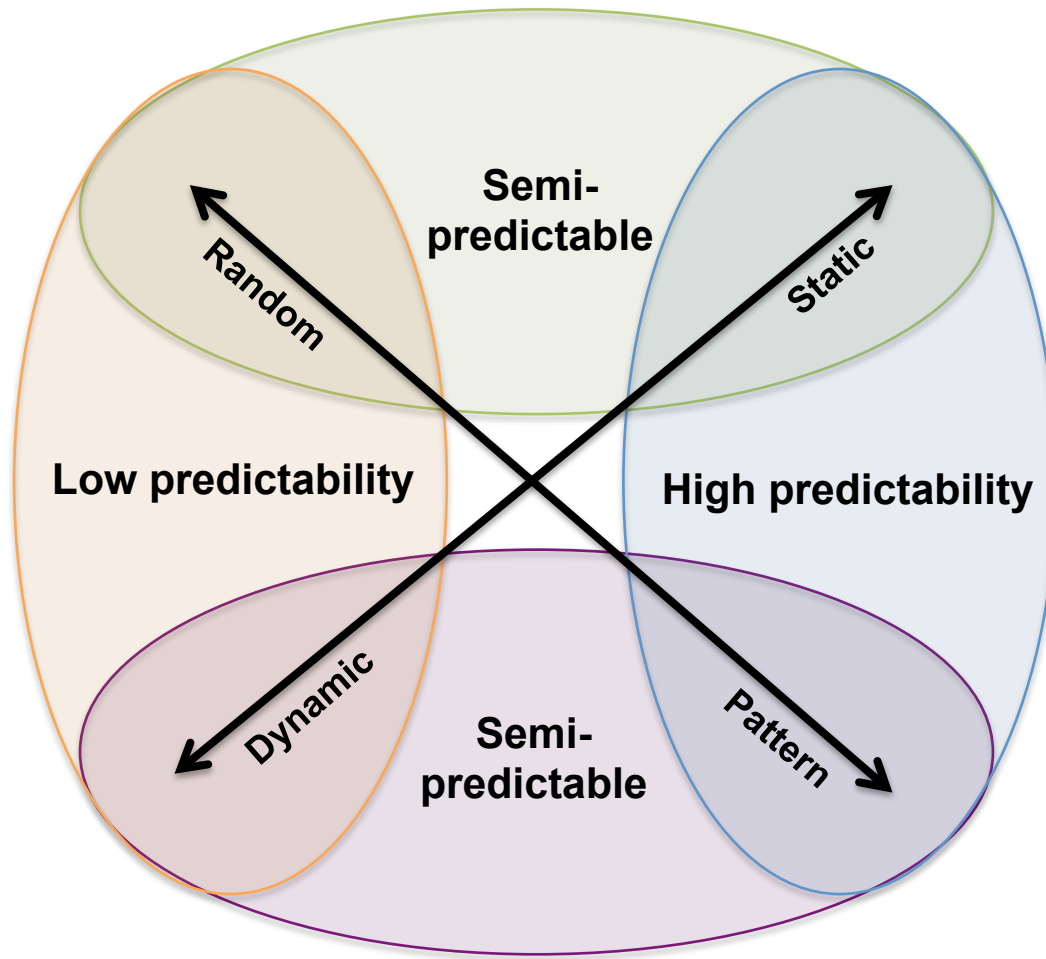


Figure 5.3. The four predictability conditions and their components.

Highly predictable condition: Pattern-Static. In this condition, the zombie bunnies moved in learnable patterns. For example, Zombie #1 might move from left to right to left and so on shifting its position slightly upward each time it reaches the edge of the screen. Also in this condition, the zombie bunnies were static at the time of performer response, that is, the zombie bunnies did not move until the participant had moved—each zombie moves one space for every space moved by the participant’s avatar.

Semi-predictable condition: Pattern-Dynamic. In this condition, the zombie bunnies moved in learnable patterns (see description of “Pattern” in previous paragraph). Also in this

condition, the zombie bunnies moved continuously; they did not stop and wait for the participant to move and thus the stimuli on the screen were in constant flux.

Semi-predictable condition: Random-Static. In this condition, the zombie bunnies moved pseudo-randomly. Also in this condition, the zombie bunnies were static at the time of performer response, that is, the zombie bunnies did not move until the participant has moved—each zombie moved one space for every space moved by the participant’s avatar.

Less predictable condition: Random-Dynamic. In this condition, the zombie bunnies moved pseudo-randomly. Also in this condition, the zombie bunnies moved continuously; they do not stop and wait for the participant to move and thus the stimuli on the screen were in constant flux.

Participants completed two out of four versions of a task. Of the 12 possible combinations and orders, four combinations and orders were chosen in order to investigate performance and learning transfer when constraints of the task environment vary and when lability is constant within participants. That is, participants completed one of these four combinations and orders: 1) pattern-static (highly predictable) then random-static (semi-predictable) ($n = 23$), 2) random-static (semi-predictable) then pattern-static ($n = 21$), 3) pattern-dynamic (semi-predictable) then random-dynamic (least predictable) ($n = 20$), or 4) random-dynamic (least predictable) then pattern dynamic (semi-predictable) ($n = 22$), see Table 5.1.

Table 5.1

Ordering and Combination of Tasks Completed

Completed First	Completed Second			
	Pattern-Static	Pattern-Dynamic	Random-Static	Random-Dynamic
Pattern-Static	---		X	
Pattern-Dynamic		---		X
Random-Static	X		---	
Random-Dynamic		X		---

5.3.2 Results and Discussion

Participants' performance in the final block was the highest in the most predictable condition (pattern-static), regardless of whether or not the condition was completed first or second. Participants' performance in the final block was the lowest in the least predictable condition (random-dynamic) regardless of whether or not the condition was completed first or second. Performance was measured in the average distance in pixels from all the zombies per move, thus the higher the number of pixels in distance, the better the performance. See Table 5.2.

Table 5.2

Performance in the Final Block of Each Condition

	First Condition Completed		Second Condition Completed	
	M	SD	M	SD
Pattern-Static	538.48	80.08	547.26	98.06
Random-Static	518.75	101.02	498.37	96.31
Pattern-Dynamic	461.95	53.16	498.59	87.06
Random-Dynamic	456.47	56.57	492.49	83.48

Independent samples-t-tests, rather than an ANOVA, were conducted in order to compare performance in the final block across conditions because Levene's test indicated a significant effect of task condition on subject variability for the condition completed first, $F(3, 82) = 4.69$, $p < .05$, thus violating the homogeneity of variance assumption of ANOVA. Likewise, Welch's t-tests, rather than Student's t-tests, were conducted for pairwise comparisons that still violated the homogeneity of variance assumption.

When comparing performance in conditions completed first, performance in the pattern-static (highly predictable) condition was significantly better than performance in the pattern-dynamic (semi-predictable) condition, $t(38.50) = 3.73$, $p = .001$, $d = 1.14$, and significantly better than performance in the random-dynamic (least predictable) condition, $t(39.64) = 3.98$, $p < .001$, $d = 1.21$. Performance in the random-static (semi-predictable) condition was significantly better than performance in the pattern-dynamic (semi-predictable) condition, $t(30.60) = 2.27$, $p = .031$, $d = .72$ and the random-dynamic (least-predictable) condition, $t(31.11) = 2.48$, $p = .019$, $d = .78$. When comparing performance in the final block in conditions completed second, performance in the pattern-static (highly predictable) condition was marginally significantly better than performance in the final block of the random-dynamic (least predictable) condition, $t(39) = 1.92$, $p = .062$, $d = .62$.

Paired-samples t-tests between the final block of the first condition and the initial block of the second condition were conducted in order to investigate learning transfer. For participants who completed the two static conditions (pattern-static then random-static or random-static then pattern-static), initial performance in the second condition was significantly lower than final performance in the first condition—pattern-static then random-static, $t(22) = -2.87$, $p = .009$, $d =$

-.60; random-static then pattern-static, $t(20) = -3.18$, $p = .005$, $d = -.69$ —suggesting that learning did not transfer across these conditions. However, for participants in the dynamic conditions (pattern-dynamic then random-dynamic or random-dynamic then pattern-dynamic) there were no significant differences in performance on the final block of the first condition and initial block of the second condition for either order—pattern-dynamic then random-dynamic, $t(19) = -.17$, ns , $d = -.04$; random-dynamic then pattern-dynamic, $t(21) = .16$, ns , $d = .03$ —suggesting that learning did transfer across these conditions. The differences between static conditions and dynamic conditions may be due to participants developing strategies to manage the dynamic stimuli, which transferred regardless of the constraint type, while participants responding to the static stimuli did not need to develop strategies for handling dynamic stimuli, and therefore participant’s learning may have occurred from focusing on the constraint type, which varied from the first condition to the second condition. See Figure 5.4 for individual growth curves across conditions.

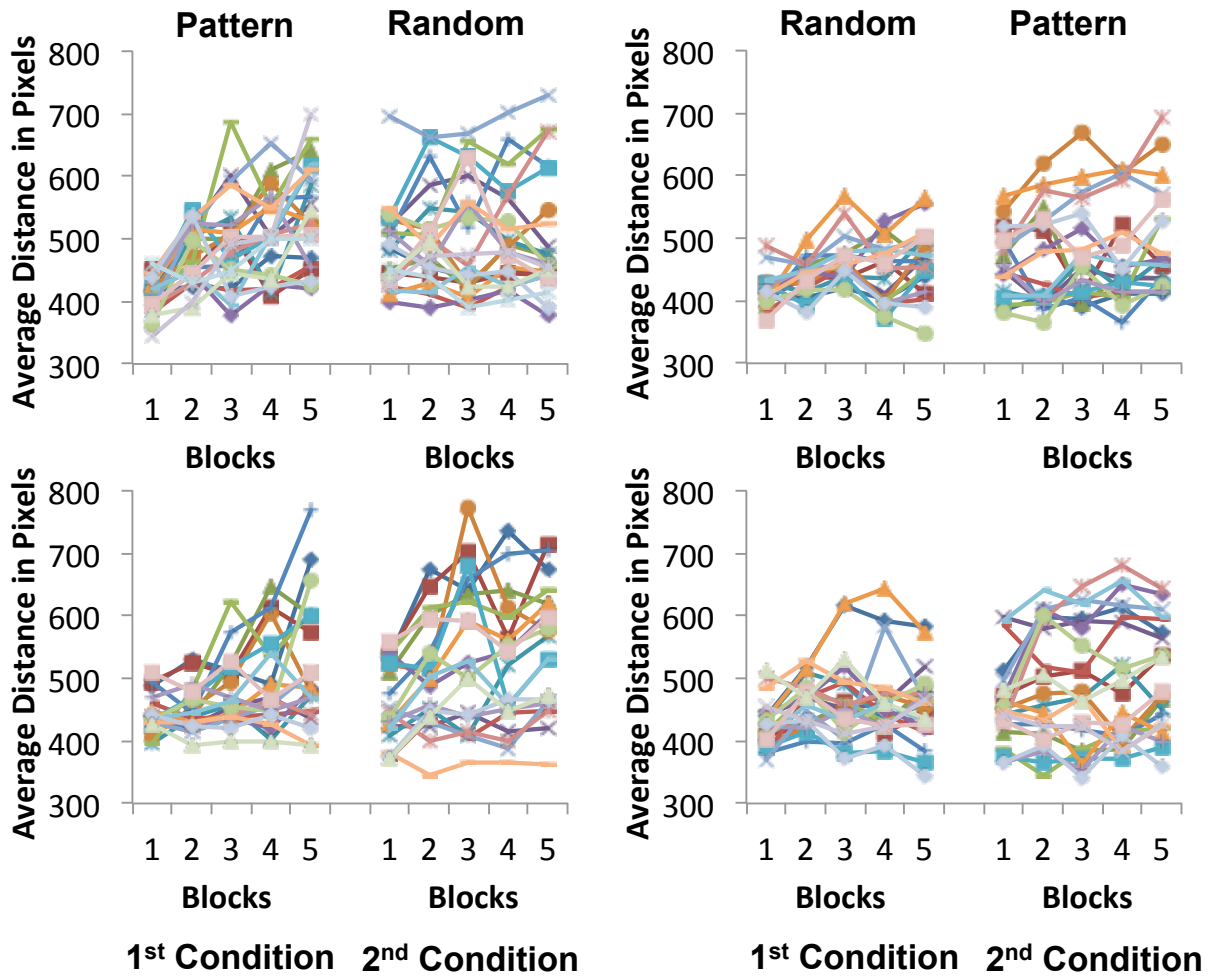


Figure 5.4. Individual growth curves across conditions.

5.4 Predictability Experiment 2

5.4.1 Methods

Participants

One hundred thirty-two undergraduate students participated in exchange for partial course credit. Participants were randomly assigned to one of two between-groups conditions and then completed two within-groups conditions.

Procedure and Materials

All participants completed the following cognitive ability and affective trait measures prior to completing the experimental task.

Emotion-Cognition Interaction

BIS/BAS scales. All participants completed the BIS and BAS scales (Carver & White, 1994), previously described in Chapter 2.

Perceptual Speed

Pattern/Letter Comparison. All participants completed Pattern Comparison and Letter Comparison (Salthouse & Babcock, 1991) previously described in Chapter 2. The scores were averaged to create a composite perceptual speed score.

Psychomotor Speed

Connections A tests. All participants completed Connections A: Numbers and Connections A: Letters (Salthouse et al., 2000), previously described in Chapter 2.

Task Switching Speed

Connections B tests. All participants completed Connections B: Numbers-Letters and Connections B: Letters-Numbers (Salthouse et al., 2000), previously described in Chapter 2.

Non-verbal Reasoning / Fluid Intelligence

Ravens Advanced Progressive Matrices. All participants completed the Raven's Advanced Progressive Matrices (Raven, 1962), previously described in Chapter 2.

Working Memory Capacity (Maintenance & Interference)

Automated Symmetry Span. All participants completed the Automated Symmetry Span (Unsworth et al., 2005), previously described in Chapter 2.

Working Memory Capacity (Coordination & Transformation)

Letter-Number Sequencing. All participants completed the Letter-Number sequencing task (adapted from Wechsler, 1997), previously described in Chapter 2.

Associative Learning Measurement

Three-term contingency learning task. All participants completed the three-term contingency learning task (Williams & Pearlberg, 2006). In the three-term contingency learning task, participants completed four learning blocks and four test blocks presented immediately after each learning block. In a learning block trial, a participant viewed a single word on the screen with three selection choices, A, B, and C. When the participant selected A, she would view the word in the A position associated with the primary word on the screen, when the participant selected B, she would view the word in the B position associated with the primary word on the screen, and when the participant selected C, she would view the word in the C position associated with the primary word. For example, in one trial the primary word was “LAB”. When participants selected A, the word “PUN” appeared, when the participant selected B, the word “TRY” appeared, and when the participant selected C, the word “EGG” appeared. The participant was tasked with learning the word associations.

During each learning block, the participant was tasked with learning ten words and their three associated words for a total of thirty associations. A test block followed each learning block. In the test block, participants viewed one of the primary words and was tasked with typing in the associated word in the correct position (A, B, or C). Immediate feedback and the correct response were provided to the participant on the screen and the participant attempted to recall the appropriate associative words in their locations for the other nine primary words. The remaining

learning blocks presented the same primary words and the same associative words, but the stimuli were presented in a different order. The possible overall scores ranged from 0 to 120.

Experimental Task

Zombie Apocalypse. All participants completed two variants of the experimental task used in Experiment 1. Unlike Experiment 1, where the two conditions completed by participants did not vary in their lability (i.e., were both static or both dynamic) and did vary in the constraints of the stimuli (i.e., participants completed one pattern condition and one random condition), participants in E2 completed two conditions that did not vary in their constraints (i.e., were both pattern or both random) and did vary in their lability (i.e., participants completed one static condition and one dynamic condition). In order to obtain sample sizes large enough to calculate reliable correlations and perform regression analyses, only two between-groups conditions were chosen: all participants completed the static condition (pattern or random) followed by the dynamic condition of the same constraint type (pattern or random), that is participants either completed 1) pattern-static (highly predictable), then pattern-dynamic (semi-dynamic) ($n = 64$), or 2) random-static (semi-dynamic), then random-dynamic (least predictable) ($n = 64$).

5.4.2 Results and Discussion

One participant did not respond for the first block of the experiment and was excluded from analyses that included task performance. Similar to Experiment 1, performance in the final block was highest in the pattern-static (highly predictable) condition and lowest in the random-dynamic (least predictable) condition. See Table 5.3 for descriptive statistics of the experimental task. In the final block, pattern-static (highly predictable) performance was significantly higher than performance in the pattern-dynamic (semi-predictable) condition, $t(64) = 4.78$, $p < .001$, $d =$

.59, the random-static (semi-predictable) condition, $t(129) = 2.99, p = .003, d = .53$, and the random-dynamic (least predictable) condition, $t(129) = 3.74, p < .001, d = .66$.

Table 5.3

Final Block Performance Across Conditions

	Static (First) Condition		Dynamic (Second) Condition	
	M	SD	M	SD
Pattern	517.74	75.83	478.79	64.34
Random	478.15	75.54	468.24	75.50

Pattern conditions. In the pattern-static (highly predictable) condition, initial performance predicted performance in the final block ($r = .311, p = .012$) and the difference in performance from the initial block to the final block was strongly correlated with final performance ($r = .956, p < .001$). Performance in the final block also predicted initial ($r = .510, p < .001$) and final performance ($r = .571, p < .001$) performance in the pattern-dynamic (semi-predictable) condition that followed the pattern-static condition. That is, how well one learned the constraint patterns in the less-dynamic environment predicted performance on the same constraint patterns in the dynamic environment. Amount of performance change from the initial block to the final block of the pattern-static (highly predictable) condition predicted initial ($r = .506, p < .001$) and final ($r = .604, p < .001$) performance in the following pattern-dynamic (semi-predictable) condition and final performance.

Random conditions. In the random-static (semi-predictable) condition, initial performance negatively predicted amount of performance change from the initial block to the final block ($r = -.451, p < .001$). The difference in performance from the initial block to the final

block was strongly correlated with final performance ($r = .859, p < .001$). Performance in the final block predicted initial ($r = .665, p < .001$) and final ($r = .569, p < .001$) performance in the following random-dynamic (least predictable) condition. Amount of performance change from the initial block to the final block in the first random-static condition also predicted initial performance in the following dynamic condition ($r = .507, p < .001$) and the final block ($r = .419, p < .001$). Performance in the initial block of the second random-continuous condition predicted performance in the final block ($r = .520, p < .001$) of the same condition and negatively predicted change in performance from the initial block to the final block ($r = -.355, p = .003$). Change in performance from the initial block to the final block predicted final performance ($r = .614, p < .001$). See Table 5.4 for difference scores and Figure 5.5 for individual growth curves across conditions.

Table 5.4

Difference Scores (Final Block Performance Minus Initial Block Performance)

	Static (First) Condition		Dynamic (Second) Condition	
	M	SD	M	SD
Pattern	114.63	87.74	-14.32	51.02
Random	21.27	84.43	16.85	68.98

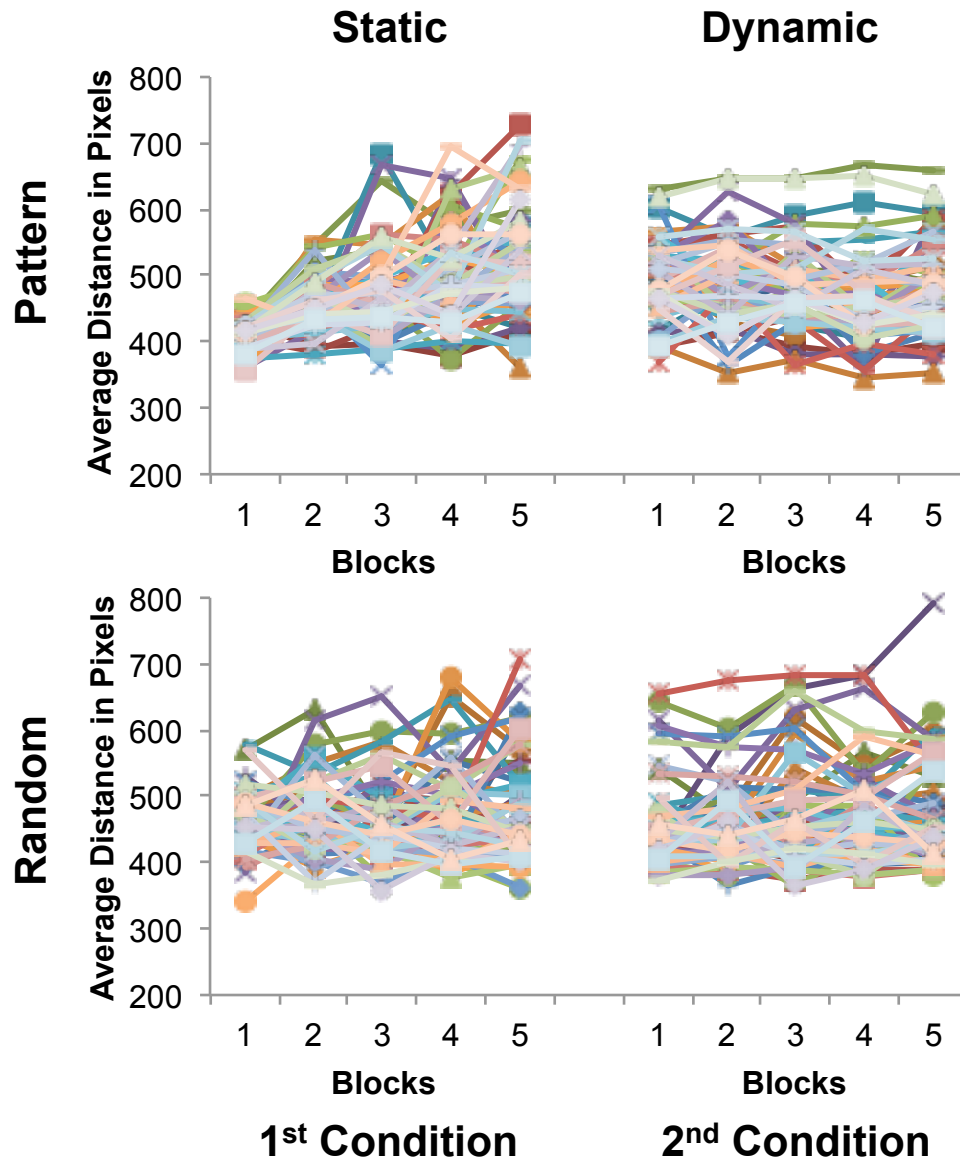


Figure 5.5. Individual growth curves across conditions.

Cognitive ability measures. See Table 5.5 for correlations among the cognitive ability and affective traits measured prior to participants completing the experimental task. Missing values (14 Symmetry Span scores due to technical error and 1 LNS score due to the participant not following instructions) were replaced with the imputed mean.

Table 5.5

Cognitive Ability and Affective Trait Inter-correlations

	1	2	3	4	5	6	7	8
1. BIS	---							
2. BAS	-.03	---						
3. Perceptual Speed	-.06	-.08	---					
4. ConnectA	-.15	-.09	.57	---				
5. ConnectB	-.07	-.13	.42	.70	---			
6. Ravens	.01	-.13	.27	.32	.40	---		
7. SSpan	-.04	.07	.30	.36	.34	.31	---	
8. LNS	-.09	-.02	.09	.23	.31	.27	.16 [†]	---
9. 3-Term	.22	.08	-.05	.03	.17	.11	.22	.12

Note. BIS = Behavioral Inhibition System Sensitivity. BAS = Behavioral Approach System Sensitivity. Perceptual Speed = Composite perceptual speed score. ConnectA = Connections A Test. ConnectB = Connections B Test. Ravens = Ravens Advanced Progressive Matrices. SSpan = Automated Symmetry Span/ LNS = Letter-Number Sequencing. 3-Term = Three-Term Contingency Learning Task. **Bold** = significant at the $p < .05$ level. [†] = $p < .08$.

In order to test whether a combination of the cognitive abilities measured predicted performance in the experimental task, correlation and regression analyses were conducted. See Table 5.6 for correlations between cognitive abilities and initial performance in the experimental task and Table 5.7 for correlations between cognitive abilities and final performance in the experimental task.

Table 5.6

Correlations Between Cognitive Abilities and Initial Block Performance

	Pattern-Static	Pattern-Dynamic	Random-Static	Random-Dynamic
BIS	.007	-.207	.032	-.005
BAS	-.128	-.255*	.014	-.011
Perceptual				
Speed	.145	.082	.091	.072
ConnectA	.035	.154	.094	.109
ConnectB	.118	.292*	.149	.174
Ravens	.162	.262*	.035	.014
SSpan	.170	.316*	.065	-.058
LNS	.061	.454**	-.028	.065
3-Term	-.032	-.086	.009	-.001

Notes. BIS = Behavioral Inhibition System Sensitivity. BAS = Behavioral Approach System Sensitivity. Pattern = Pattern Comparison. Letter = Letter Comparison. ConnectA = Connections A Test. Connect B = Connections B Test. Ravens = Raven's Advanced progressive Matrices. SSpan = Automated Symmetry Span. LNS = Letter-Number Sequencing. 3-Term = Three-term Contingency Task. **Bold** = significant at $p < .05$. ** = $p < .01$. * = $p < .05$. † = $p < .08$.

Table 5.7

Correlations Between Cognitive Abilities and Final Block Performance

	Pattern-Static	Pattern-Dynamic	Random-Static	Random-Dynamic
BIS	-.184	-.154	.070	.047
BAS	-.169	-.285*	.050	.097
Perceptual				
Speed	.112	.113	.078	-.109
ConnectA	.050	.093	.082	.066
ConnectB	.087	.132	.158	.152
Ravens	.184	.224 [†]	.120	-.087
SSpan	.103	.251*	.074	.030
LNS	.172	.207	.268*	.209
3-Term	-.125	-.220 [†]	.241 [†]	.204

Notes. BIS = Behavioral Inhibition System Sensitivity. BAS = Behavioral Approach System Sensitivity. Pattern = Pattern Comparison. Letter = Letter Comparison. ConnectA = Connections A Test. Connect B = Connections B Test. Ravens = Raven's Advanced progressive Matrices. SSpan = Automated Symmetry Span. LNS = Letter-Number Sequencing. 3-Term = Three-term Contingency Task. **Bold** = significant at $p < .05$. ** = $p < .01$. * = $p < .05$. † = $p < .08$.

Cognitive ability measures correlated significantly and marginally significantly with performance in the semi-predictable conditions. Performance on the initial block of the pattern-continuous (semi-predictable) condition was predicted by performance on BAS ($r = -.255$, $p =$

.040), Connections B ($r = .292, p = .018$), Ravens ($r = .262, p = .035$), and Symmetry Span ($r = .316, p = .010$), and LNS ($r = .454, p < .001$). Performance on the final block of the pattern-continuous condition was predicted by BAS ($r = -.285, p = .021$), Ravens ($r = .224, p = .073$), Symmetry Span ($r = .251, p = .044$) and 3-Term Contingency Task ($r = -.220, p = .078$). Performance on the final block of the random-static (semi-predictable) condition as well as the change in performance between the initial and final blocks was predicted by performance on the letter-number sequencing task ($rs = .268$, and $.254, ps < .05$) and 3-Term Contingency Task predicted final random-static performance ($r = .241, p = .051$).

Given that several of these cognitive abilities correlate with one another, the question remains whether each of the cognitive abilities predicting experimental task performance are contributing unique variance or are accounting for the same variance as other cognitive abilities. In order to answer this question, hierarchical regression analyses were conducted. First, a full model (all cognitive and affective measures predicting initial and final performance) was compared to a partial model (all cognitive and affective measures with significant or marginally significant correlations with final block performance predicting initial and final block performance).

Pattern-Static (highly predictable) condition. All cognitive abilities and affective measures combined accounted for only 08% of the variance in initial pattern-static block performance, $F(9, 55) = .56, ns, R^2 = .084$ and 13% of the variance in final pattern-static block performance, $F(9, 55) = .90, ns, R^2 = .128$. No further regression analyses were conducted with pattern-static performance as the outcome variable.

Pattern-Dynamic (semi-predictable) condition. All cognitive abilities and affective measures combined accounted for 40% of the variance in initial pattern-dynamic block

performance, $F(9, 55) = 4.01, p = .001, R^2 = .396$ and 28% of the variance in final pattern-dynamic block performance, $F(9, 55) = 2.37, p = .024, R^2 = .279$. These regression models were compared to models that only included the measures that significantly or marginally significantly correlated with initial or final block performance. The change in R^2 was not significant when the outcome measure was initial pattern-dynamic block performance, $\Delta R^2 = .043, F(4, 55) = .98, p = .427$, indicating that the measures that significantly or marginally significantly correlated with initial performance accounted for 35% of the 40% of the initial pattern-dynamic performance variance explained. The change in R^2 was also not significant when the outcome measure was final pattern-dynamic block performance, $\Delta R^2 = .095, F(6, 55) = 1.12, p = .318$, indicating that the measures that significantly or marginally significantly correlated with final transfer performance accounted for 19% of the 28% of the final pattern-dynamic performance variance explained.

For each cognitive ability that correlated with initial pattern-dynamic block performance or final pattern-dynamic block performance, a regression model was constructed comparing a model in which all the measures that correlated with performance were taken into account, and a model where a single variable was excluded in order to test whether the difference between the models was significant, that is, whether the excluded measure accounted for significant unique variance. See Tables 5.8 and 5.9.

Table 5.8

Summary of Regression Analyses, Initial Pattern-Continuous Block Performance

Outcome = Initial Pattern-Continuous Block Performance						
Partial Model						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-41.888	15.76	-2.66	-.295	.010	
ConnectB	-.48	1.73	-0.28	-.037	.781	
Ravens	7.85	51.51	0.15	.019	.879	
SSpan	122.65	44.91	2.73	.324	.008	
LNS	158.62	46.04	3.45	.399	.001	
						.353
Excluding BAS						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
ConnectB	-0.02	1.80	-0.01	-.002	.990	
Ravens	34.88	52.99	0.66	.084	.513	
SSpan	89.94	45.32	1.99	.238	.052	
LNS	156.82	48.30	3.25	.394	.002	
						.276
$\Delta R^2 = .077, F(1, 59) = 7.06, p = .010$						
Excluding ConnectB						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-41.44	15.56	-2.66	-.292	.010	
Ravens	2.90	48.00	0.06	.007	.952	
SSpan	118.58	42.15	2.81	.313	.007	
LNS	154.74	43.56	3.55	.389	.001	
						.352
$\Delta R^2 = .001, F(1, 59) = .08, p = .781$						
Excluding Ravens						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-42.35	15.33	-2.76	-.299	.008	
ConnectB	-0.39	1.61	-0.24	-.030	.808	
SSpan	123.51	44.19	2.80	.336	.007	
LNS	159.57	45.24	3.53	.401	.001	
						.353
$\Delta R^2 = .000, F(1, 59) = .02, p = .879$						
Excluding SSpan						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-30.08	15.96	-1.89	-.212	.064	
ConnectB	1.048	1.72	0.61	.081	.544	
Ravens	25.61	53.78	0.48	.062	.636	
LNS	155.34	48.44	3.21	.390	.002	
						.271
$\Delta R^2 = .082, F(1, 59) = 7.46, p = .008$						

Excluding LNS						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-41.08	17.13	-2.40	-.290	.020	
ConnectB	1.31	1.79	0.73	.101	.467	
Ravens	31.81	55.47	0.57	.076	.569	
SSpan	118.61	48.79	2.43	.313	.018	
						.223
$\Delta R^2 = .130, F(1, 59) = 11.87, p = .001$						

Notes. Partial Model = all measures whose correlation coefficients were significant or marginally significant with final interpreting transfer performance. BAS = Behavioral Approach System Sensitivity. ConnectB = Connections B Test. Intelligence = Composite measure of fluid intelligence. Ravens = Ravens Advanced Progressive Matrices. SSpan = Automated Symmetry Span. LNS = Letter-Number Sequencing.

Table 5.9

Summary of Regression Analyses, Final Pattern-Continuous Block Performance

Outcome = Final Pattern-Continuous Block Performance						
Partial Model						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-52.72	19.87	-2.65	-.324	.010	
Ravens	40.86	59.08	0.69	.086	.492	
SSpan	126.75	53.65	2.36	.292	.021	
						.185
Excluding BAS						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
Ravens	81.94	59.73	1.37	.172	.175	
SSpan	90.27	54.33	1.66	.208	.102	
						.091
$\Delta R^2 = .094, F(1, 61) = 7.04, p = .010$						
Excluding Ravens						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-56.32	19.10	-2.95	-.346	.009	
SSpan	137.91	50.95	2.71	.318	.004	
						.178
$\Delta R^2 = .006, F(1, 61) = .478, p = .492$						
Excluding SSpan						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	<i>R</i> ²
BAS	-40.69	19.90	-2.04	-.250	.045	
Ravens	82.82	58.39	1.42	.173	.061	
						.110
$\Delta R^2 = .075, F(1, 61) = 5.58, p = .021$						

Notes. Partial Model = all measures whose correlation coefficients were significant or marginally significant with final interpreting fluency performance. BAS = Behavioral Inhibition System Sensitivity. Ravens = Ravens Advanced Progressive Matrices. SSpan = Automated Symmetry Span.

Behavioral approach system sensitivity (“BAS”), task switching speed (“ConnectB”), fluid intelligence (“Ravens”), working memory capacity: maintenance and interference (“SSpan”), and working memory capacity: coordination and transformation (“LNS”) accounted for about 35% of the 40% variance in initial pattern-continuous block performance. BAS accounted 08% of the initial block performance variance. In final pattern-continuous block performance, both BAS and working memory capacity: maintenance and interference accounted for unique amounts of variance in performance. BAS accounted for 09% of the final performance variance and SSpan accounted for 07.5% of the final performance variance.

Random-Static (semi-predictable) condition. All cognitive abilities and affective measures combined accounted for only 03% of the variance in initial random-static block performance, $F(9, 55) = .56, ns, R^2 = .034$ and 13% of the variance in final random-static block performance, $F(9, 55) = .90, ns, R^2 = .126$. No cognitive or affective measures correlated with initial random-continuous block performance. Both working memory span: coordination and transformation and associative learning correlated with final random-static block performance. The regression model in which final performance was the outcome variable was compared to a model that only included the measures that significantly or marginally significantly correlated with initial or final block performance. The change in R^2 was not significant, $\Delta R^2 = .023, F(7, 55) = .21, p = .636$, indicating that the measures that significantly or marginally significantly correlated with final performance accounted for 10% of the 13% of the final random-static performance variance explained.

For each cognitive ability that correlated with final random-static block performance, a regression model was constructed comparing a model in which all the measures that correlated

with performance were taken into account, and a model where a single variable was excluded in order to test whether the difference between the models was significant, that is, whether the excluded measure accounted for significant unique variance. See Table 5.10.

Table 5.10

Summary of Regression Analyses, Final Random-Static Block Performance

Outcome = Final Random-Static Block Performance						
Partial Model						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	R^2
LNS	156.00	88.25	1.77	.219	.082	
3-Term	0.452	.31	1.47	.182	.146	
						.103
Excluding LNS						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	R^2
3-Term	-.344	.345	-1.00	-.125	.323	
						.058
$\Delta R^2 = .044, F(1, 63) = 3.12, p = .082$						
Excluding 3-Term						
Predictors	<i>B</i>	<i>SE</i>	<i>t</i>	β	<i>p</i>	R^2
LNS	190.95	85.78	2.23	.268	.030	
						.072
$\Delta R^2 = .031, F(1, 63) = 2.17, p = .146$						

Notes. Partial Model = all measures whose correlation coefficients were significant or marginally significant with final interpreting fluency performance. LNS = Letter-Number Sequencing Task. 3-Term = Three term contingency learning task.

Random-Dynamic (least predictable) condition. All cognitive abilities and affective measures combined accounted for only 05% of the variance in initial random-continuous block performance, $F(9, 56) = .34, ns, R^2 = .051$ and 13% of the variance in final pattern-static block performance, $F(9, 56) = .92, ns, R^2 = .129$. No cognitive abilities or affective trait measures significantly or marginally significantly correlated with either initial or final random-dynamic

block performance. No further regression analyses were conducted with random-dynamic performance as the outcome variable.

5.5 General Discussion

This chapter presented two experiments in which predictability of the task environment was manipulated and complexity—operationalized as the number of inputs, that is, stimuli and rules—was kept consistent by maintaining the same number of stimuli and rules within each task environment. The results of the two experiments revealed that in the most predictable condition—the pattern-static condition—initial performance was similar across individuals, however variance diverged over time. In addition, performance in the final block was the highest compared with the other conditions and change from initial performance to final performance was the greatest compared with the other conditions. Initial performance also predicted final performance.

Overall, the most learning occurred in the predictable condition and the least amount of learning occurred in the least-predictable condition following the stated hypotheses. Cognitive abilities did not predict performance in the most predictable condition nor in the least predictable condition. Instead, cognitive abilities predicted performance in the semi-predictable conditions, primarily the pattern-dynamic condition.

Learning transfer occurred when both task environments were dynamic, regardless of whether the constraint type was predictable or not. The lack of difference between whether the pattern or random condition is completed first in this scenario suggests that the participant is not able to learn movement patterns when the environment is highly dynamic and that only strategies for managing a dynamic environment were developed. However, learning transfer also occurred

when the first task environment's stimuli were highly predictable (pattern-static) and the second task environment differed only in the lability of the stimuli (pattern dynamic). This result suggests that learning patterns in the absence of a highly dynamic environment not only increases pattern learning (shown from the higher rates of learning compared with other conditions), but also allows the participant to transfer this learning to a highly dynamic environment. Supporting this idea is the finding that learning transfer does not take place between a task environment whose stimuli are predictable (pattern) and static and a task environment whose stimuli are not predictable (random) and static, suggesting again, that the participant is focused on learning the patterns, which does not transfer to stimuli that do not follow the same patterns.

While the results support the idea that practice better predicts performance in the most predictable condition, and that cognitive abilities predict performance in less (semi-) predictable conditions, cognitive abilities did not predict performance in the least predictable condition. This is likely due to the un-learnable components of the task: there is not time to employ strategies for responding to the stimuli and the movement of stimuli cannot be learned.

Future research should attempt to vary predictability to mirror varying predictability in real-world tasks, that is, even tasks with low predictability have some predictability if participants can improve their performance on them. In this way, a computer task may have stimuli moving in patterns that can change at any time and must either be learned or recognized as previously learned patterns. Changing patterns such as this would have similarity to real-world, low-predictability tasks such as simultaneous interpreting.

Chapter 6

The role of predictability of the task environment in real-world domains

6.1 Chapter summary

The relationship between deliberate practice and performance variance is investigated through a meta-analysis⁹.

- Proponents of the deliberate practice view claim that deliberate practice is a sufficient account of performance variance across domains.
- The first formal meta-analysis of deliberate practice and performance is presented, which covers all the major domains in which deliberate practice has been studied.
- On average, deliberate practice accounted for 12% of the variance in performance.
- Both domain and predictability of the task environment significantly moderated the relationship between deliberate practice and performance, such that as predictability of the task environment decreased, so did the strength of the relationship between deliberate practice and performance.

⁹ This study has been accepted for publication (Macnamara, Hambrick, & Oswald, 2014). Also, a previous version was presented at a Psychonomic Society meeting (2013).

6.2 Introduction: A meta-analysis of deliberate practice and performance

Conclusions drawn from a meta-analysis are statistically stronger than any one study can produce. Therefore, a meta-analysis was conducted with the primary aim of estimating the true magnitude of the relationship between deliberate practice—a highly popular account of skill acquisition and expertise that supports of nurture-over-nature/made-over-born view—and performance variance, and a secondary aim of investigating potential moderators of the relationship between deliberate practice and performance variance in real-world tasks, such as predictability of the task environment.

The deliberate practice view

The deliberate practice view, first proposed by Ericsson, Krampe, & Tesch-Römer in 1993, holds that expert performance largely reflects accumulated amount of *deliberate practice*, which Ericsson et al. defined as engagement in structured activities created specifically to improve performance in a domain. In two studies, Ericsson et al. recruited musicians representing different levels of accomplishment and asked them to estimate the amount of deliberate practice they had engaged in per week for each year of their musical careers. Cumulative averages were much higher for the most accomplished groups of musicians than for the less accomplished groups. For example, at age 20, the average for the “best” violinists was over 10,000 hours, compared with about 7,800 hours for the “good” violinists, and about 4,600 hours for the least accomplished group.

Ericsson et al. (1993) concluded that “high levels of deliberate practice are necessary to attain expert level performance” and added: “Our theoretical framework can *also provide a*

sufficient account of the major facts about the nature and scarcity of exceptional performance. Our account does not depend on scarcity of innate ability (talent)..." (p. 392, emphasis added). Ericsson et al. continued, "we argue that the differences between expert performers and normal adults reflect a life-long period of deliberate effort to improve performance in a specific domain" (p. 400). Ericsson (2007) reiterated this perspective when he claimed "the distinctive characteristics of elite performers are adaptations to extended and intense practice activities that selectively activate dormant genes that all healthy children's DNA contain" (p. 4).

The deliberate practice view has inspired a great deal of interest in expert performance. The Ericsson et al. (1993) article has been cited over 4,200 times (Google Scholar, April, 2014) and Ericsson and colleagues' research is discussed in a number of popular books, including Malcolm Gladwell's (2008) *Outliers*, Steven Levitt and Stephen Dunbar's (2009) *SuperFreakonomics*, and Geoff Colvin's (2008) *Talent is Overrated*. Ericsson and colleagues' research was also the stimulus for what Gladwell termed the "10,000-hour rule" (Ericsson, 2013)—the idea that it takes 10,000 hours of practice to become an expert.

At the same time, the deliberate practice view has been sharply criticized in the scientific literature. Gardner (1995) commented that the view requires a "blindness...to decades of psychological theorizing" (p. 802), and Sternberg (1996) observed that "[d]eliberate practice may be correlated with success because it is a proxy for ability: We stop what we do not do well and feel unrewarded for" (p. 350). Anderson (2000) added that "Ericsson and Krampe's research does not really establish the case that a great deal of practice is sufficient for great talent" (p. 324), and Marcus (2012) concluded "it would be a logical error to infer from the importance of practice that talent is somehow irrelevant, as if the two were in mutual opposition" (p. 94).

Furthermore, there is growing evidence that although deliberate practice is important, it is considerably less important than Ericsson and colleagues have argued. Gobet and Campitelli (2007) found a large amount of variability in total amount of deliberate practice even among master-level chess players—from just over 3,000 hours to over 23,000 hours. More recently, in a reanalysis of previous findings, Hambrick, Oswald, Altmann, Mainz, Gobet, and Campitelli (2014) found that deliberate practice accounted for only about one third of the reliable variance in performance in chess and music, leaving the rest potentially explainable by other factors.

The Current Meta-analysis

The current meta-analysis is a broad investigation of studies relevant to the deliberate practice view. It is the first formal meta-analysis of the relationship between deliberate practice and human performance, and covers all major domains in which this relationship has been studied: music, games, sports, professions, and education.

The first goal of the meta-analysis is to estimate the overall correlation between amount of deliberate practice and performance. Ericsson and colleagues have based their conclusions about the importance of deliberate practice on findings with measures reflecting accumulated amount (i.e., number of hours) of deliberate practice (e.g., Duffy, Baluch, & Ericsson, 2004; Ericsson et al., 1993; Lehmann & Ericsson, 1996; Tuffiash, Roring, & Ericsson, 2007). Thus, the specific question asked is how much of the total variance in performance is explained by accumulated amount of deliberate practice.

The second goal is to investigate factors that might moderate the relationship between deliberate practice and performance. The first set of factors, which we term *theoretical*

moderators, includes (1) *Domain*: music, games, sports, professions, or education^{10,11,12}; and (2) *Predictability of task environment*—the degree to which the task environment can change while the performer is planning and executing an action, and the range of possible actions that may be executed: low predictability (e.g., handling aviation emergencies), moderate predictability (e.g., fencing), or high predictability (e.g., running). There are no predictions about how the strength of the relationship between deliberate practice and performance varies across domains, but according to the predictability of the task environment hypothesis, this relationship will generally be more positive for activities higher in predictability than for activities lower in predictability, based on findings that effects of training on performance are stronger when the task environment is more predictable than when less predictable (e.g., Ackerman, Kanfer, & Goff, 1995; Schneider, & Fisk, 1982).

The second set of factors, which we term *methodological moderators*, includes (1) *Method used to assess deliberate practice*: retrospective questionnaire, retrospective interview, or log (diary or computer); and (2) *Method used to assess performance*: expert ratings of performance, standardized objective measures of performance (e.g., chess rating), group membership (e.g., amateur vs. professional), or performance on an objectively scored laboratory task. When a retrospective method is used to assess deliberate practice (questionnaire or interview), participants are asked to recall and estimate their past engagement in deliberate practice. By contrast, when the log method is used, deliberate practice is recorded on an ongoing

¹⁰The *professions* category did not include studies of professional athletes (included in Sports) or of professional musicians (included in Music). Rather, this category included professions not captured by the other domains: computer programming, military aircraft piloting, soccer refereeing, and insurance selling.

¹¹The *education* category primarily included studies of university students in which the achievement outcome was a course grade or semester GPA. See Open Data notes for details.

¹²We classified ballet as a sport because it is a highly physical activity and has similarities to sports such as gymnastics and figure skating.

basis, either by the participant in a diary or by a computer. Given that people do not have perfect memory for the past, particularly the distant past, the log method presumably yields more accurate (valid) estimates of deliberate practice than retrospective methods. Therefore, we ask whether the relationship between deliberate practice and performance differs for the log method versus the retrospective methods.

6.3 Methods

The meta-analysis and results were conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (www.prisma-statement.org).

Inclusion Criteria, Literature Search, and Coding

There were five criteria for including a study in the meta-analysis: (1) a measure of accumulated amount (e.g., number of hours)¹³ of one or more activities interpretable as deliberate practice¹⁴ (henceforth, *deliberate practice*) was collected, with reference in the study report to at least one publication on deliberate practice by Ericsson and colleagues, (2) a measure of performance reflecting level of skill in some domain was collected, (3) an effect size reflecting the relationship between accumulated amount of deliberate practice and performance, or information needed to compute this effect size, was reported or obtained, (4) the methods and

¹³For studies in which the total amount of time that participants had to accumulate deliberate practice was a constant (e.g., a college semester), we were able to use weekly amount of deliberate practice as a measure of accumulated amount of deliberate practice, given that this variable and accumulated amount of deliberate practice would necessarily correlate the same with performance.

¹⁴The focus of this meta-analysis is on the relationship between individual difference in accumulated deliberate practice and performance. We did not include studies that experimentally manipulated training and then compared trained and untrained individuals.

results were in English, and (5) the participants were human. Studies were not excluded based on participant age or skill level.

To identify studies meeting these criteria, relevant published and unpublished articles in psychology, education, sports science, medicine, and other disciplines were systematically searched through March 24th, 2014 (see PRISMA flowchart, Figure 6.1). An e-mail was also sent to authors of articles on deliberate practice requesting information relevant to our meta-analysis not accessible (e.g., unpublished data), and asking that they forward the email to colleagues.

The search and e-mail request yielded 9,331 potentially relevant articles. After examining these articles and discarding irrelevant ones (e.g., literature reviews, commentaries), 88 studies were identified that met all the inclusion criteria. Each study was coded and the measures were collected in it for reference information, methodological characteristics, and results. These studies included 111 independent samples, with 157 effect sizes and a total sample size of 11,135 participants (see Table 6.1 for additional characteristics of the meta-analysis).

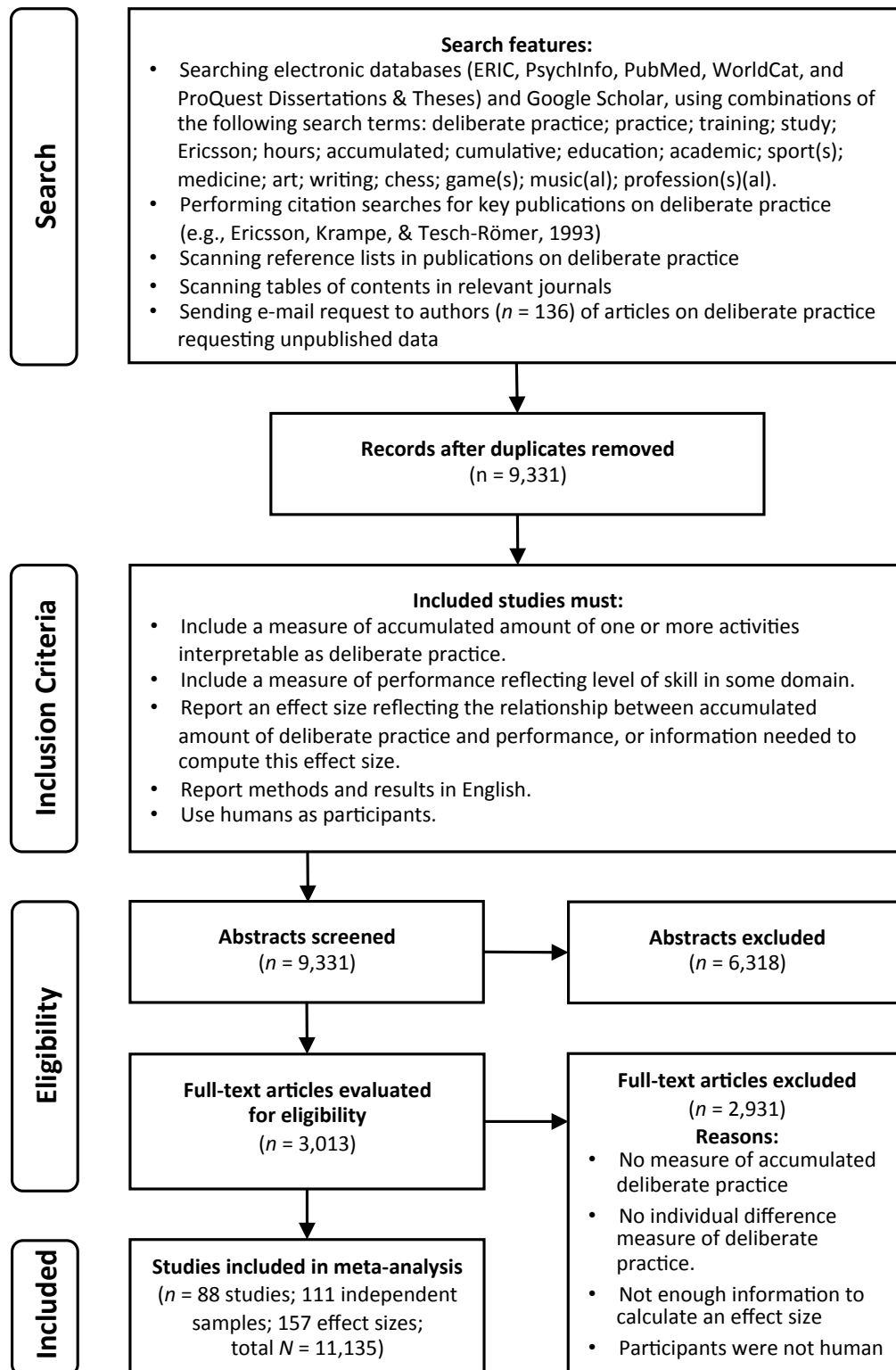


Figure 6.1. Flow Diagram of the Literature Search and Study Coding.

Table 6.1
Descriptive Characteristics of Meta-Analysis

	Frequency	
Documents	82	
Studies	88	
Independent samples	111	
Effect Sizes		
	Frequency	Cumulative <i>N</i>
Total Number	157	11,135
Domain		
Music	28	1,259
Games	11	1,291
Sports	60	2,633
Professions	7	321
Education	51	5,631
Method used to estimate accumulated practice		
Interviews	36	1,238
Questionnaires	96	8,233
Logs	25	1,664
Method used to estimate performance		
Standardized objective measures	73	7,275
Laboratory measures	20	473
Group memberships	35	2,266
Expert ratings	29	1,307
Publication status		
Published	128	10,155
Unpublished	29	980

Note. Subtotaled *N*s do not always add up to Cumulative *N* for the meta-analysis due to some samples contributing to multiple types of effects.

For each moderator variable, two researchers on this project worked together to code the effect sizes, and then two individuals who had no knowledge of the effect sizes provided separate sets of coding so that inter-rater reliability could be assessed. As indexed by Cohen's kappa (k) for the categorical variables and Spearman's rho (ρ) for the quantitative variable, inter-rater agreement was generally high for all variables, both in terms of agreement between the independent raters, and agreement between these individuals' ratings and the authors' ratings:

domain ($ks=.99-1.00$), predictability of the task environment ($\rho s=.89-.96$), method used to assess deliberate practice ($ks=.91-.98$), and method used to assess performance ($ks=.78-.83$).

The researchers resolved any discrepancies.

Effect Sizes

The meta-analysis used the correlation between accumulated amount of deliberate practice and performance as the measure of effect size. Most studies reported a correlation coefficient¹⁵; for studies that reported group-level comparisons (e.g., professional vs. amateur musicians), we converted standardized mean differences (Cohen's ds) to biserial correlations (r_{bs} ; Becker, 1986; Hunter & Schmidt, 1990).

Meta-Analytic Procedure

The meta-analysis involved four steps. The first step was to obtain correlations between deliberate practice and performance, along with their associated sampling error variances. The second step was to search for extreme values. One effect size exceeded 1.0 ($r = 1.15$); this effect size was judged to be invalid and deleted it. There also were four outliers—effect sizes whose residuals had z scores of 3 or greater ($rs = .91, .90, .90$, and $.84$); these values were Winsorized to z scores equaling 2.99 ($rs=.83, .83, .84$, and $.83$, respectively). The third step was to estimate overall effects and heterogeneity in those effects using random-effects meta-analysis modeling, and then to test whether some of the heterogeneity is predictable from moderator variables using mixed-effects meta-analysis modeling. The final step was to perform publication bias analyses. The Comprehensive Meta Analysis software package (Borenstein, Hedges, Higgins, & Rothstein, 2005) was used to conduct the meta-analyses and publication bias analyses.

¹⁵We reversed the sign of the correlation where appropriate before analyzing the data. For instance, negative correlations between deliberate practice and race times in sport indicate that more deliberate practice is associated with lower (faster) race times (i.e., more deliberate practice is associated with better performance).

6.4 Results

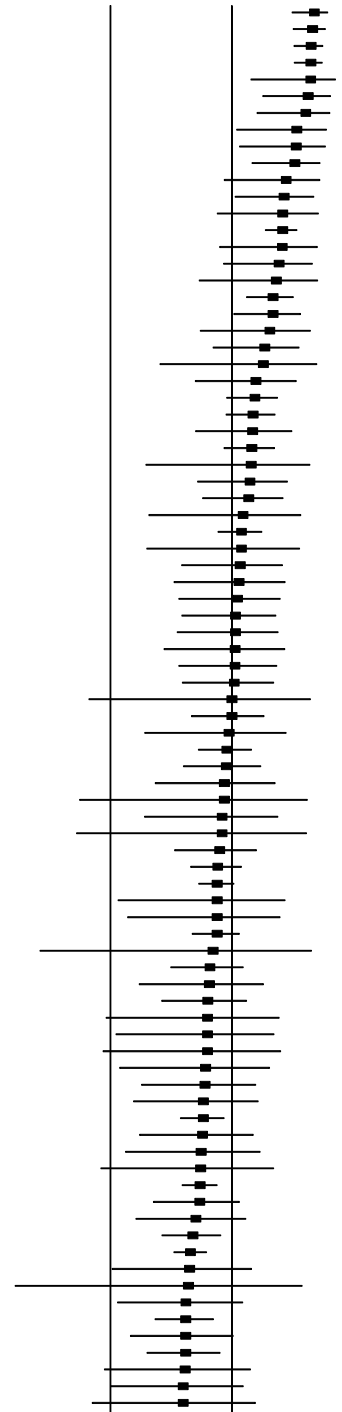
Figure 6.2 shows that nearly all correlations between deliberate practice and performance were positive: high levels of deliberate practice were associated with high levels of performance. Of the small number of negative correlations (10 of 157), only 2 ($< 1.5\%$ of all correlations) were statistically significant ($p < .05$).

The meta-analytic average correlation between deliberate practice and performance was .35, 95% CI [.30, .39], indicating that deliberate practice explained 12% of the variance in performance, 95% CI [9%, 15%], leaving 88% of the variance unexplained. However, as indicated by the I^2 statistic, which indicates the percentage of the between-study variability in effect sizes that is due to heterogeneity rather than random error, there was a high degree of heterogeneity in the effect sizes: $I^2 = 84.90$. The source of this heterogeneity through the moderator analyses was investigated and is reported next.

Study name

Correlation and 95% CI

Ward et al. (2007) - S6 *
 Ericsson, Krampe, & Tesch-Römer (1993) - S2 *
 Helsen, Starkes, & Hodges (1998) - S1 *
 Bilalic, McLeod, & Gobet (2007) - M1 *
 Ward et al. (2007) - S4
 Ward et al. (2007) - S5
 Harris (2008) - M1
 Da Matta (2004)
 Harris (2008) - M3
 Bilalic, McLeod, & Gobet (2007) - M2
 Weissensteiner et al. (2008) - S3
 Ward et al. (2007) - S1
 Harris (2008) - M2
 Zimmerman & Kitsantas (2005)
 Ward et al. (2007) - S3
 Ericsson, Krampe, & Tesch-Römer (1993) - S1
 Ward et al. (2007) - S7
 Hallam (1998)
 Meinz & Hambrick (2010)
 Ward et al. (2007) - S2
 Weissensteiner et al. (2008) - S1
 Guo (2006)
 Young et al. (2008)
 Elferink-Gemser et al. (2011) - S2
 Soboda et al. (1996)
 Tuffiash, Roring, & Ericsson (2007)
 Tuffiash (2002)
 Hodges et al. (2004) - S4
 Bruce, Farrow, & Raynor (2013)
 Gobet & Campitelli (2007) - M1
 Ward et al. (2007) - S8
 Charness et al. (2005) - S1
 Ruthsatz et al. (2008) - S3
 Massui et al. (2012) - S3 - M4
 Duffy, Baluch, & Ericsson (2004)
 Massui et al. (2012) - S3 - M5
 Helsen, Starkes, & Hodges (1998) - S2
 Schultetus & Charness (1997)
 Ford & Williams (2012)
 Baker, Deakin, & Cote (2005)
 Hallam (2001)
 Kopiez et al. (2011) - M4
 McPherson (2005) - M4
 Woody (2003)
 Charness et al. (2005) - S2
 McPherson (2005) - M3
 Hodges et al. (2004) - S1
 Kopiez et al. (2011) - M1
 Jabusch, Yong, & Altenmüller (2007)
 Kopiez et al. (2011) - M3
 De Bruin, Rikers, & Schmidt (2007)
 Urena (2004)
 Diseth et al. (2010)
 Hodges et al. (2004) - S3
 Hodges, Augaitis, & Crocker (in revision) - S1
 Svanum & Bigotti (2006)
 Lehmann & Ericsson (1996) - M2
 Meinz (2000)
 Weissensteiner et al. (2008) - S2
 McPherson (2005) - M2
 Hodges et al. (2004) - S2
 Hodges, Augaitis, & Crocker (in revision) - S3
 Jabusch et al. (2009)
 Hendry (2012) - S1 - M4
 Hutchinson, Sachs-Ericsson, & Ericsson (2013)
 Massui et al. (2012) - S2 - M4
 Ryabov (2012)
 Gobet & Campitelli (2007) - M2
 Massui et al. (2012) - S1 - M3
 Cathey (2010)
 Brunborg et al. (2010)
 McKinney & Davis (2003) - S1
 Maynard, Meinz, & Hambrick (2014)
 Ruthsatz et al. (2008) - S1
 Howard (2012)
 Massui et al. (2012) - S1 - M5
 Lehmann & Ericsson (1996) - M1
 Massui et al. (2012) - S3 - M3
 Duckworth et al. (2011)
 Ruthsatz et al. (2008) - S2
 van Gelder, Bissett, & Cumming (2004)
 Berry, Abernethy, & Cote (2008)
 Hendry (2012) - S2 - M2
 Hodges, Augaitis, & Crocker (in revision) - S4



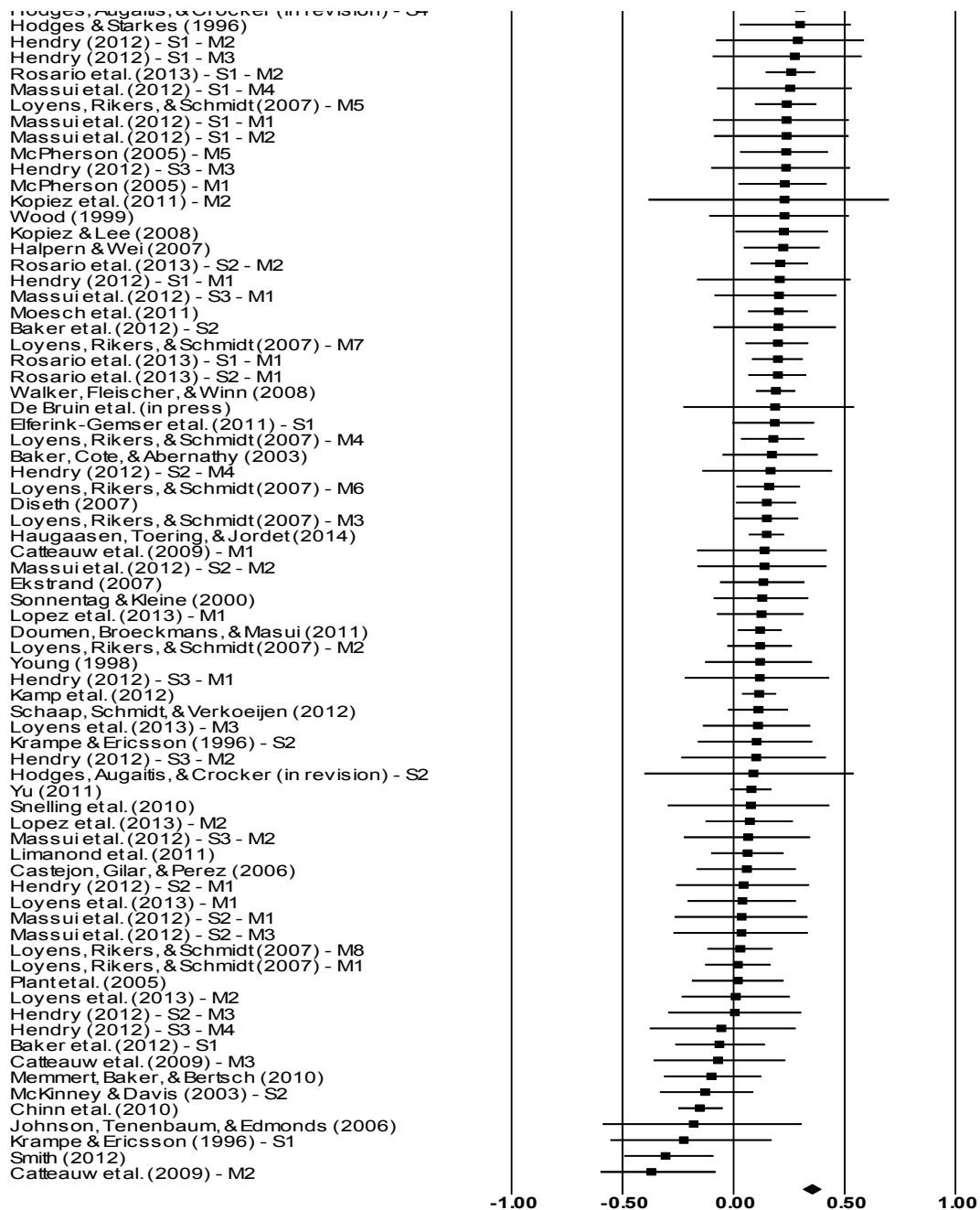


Figure 6.2. Correlations Between Deliberate Practice and Performance. Correlations (represented by squares) and 95% confidence intervals (represented by lines) of effects entered into the meta-analysis. The diamond on the bottom row represents the meta-analytically weighted mean correlation. S1 = independent sample #1 of the document, S2 = independent sample #2 of the document, etc. M1 = performance measure #1 of the sample, M2 = performance measure #2 of the sample, etc. Multiple measures were adjusted for dependency; see SOM-R for method. * = adjusted (Winsorized) outlier.

Moderator Analyses

Theoretical moderators. Domain was a statistically significant moderator, $Q(4)=49.09, p<.001$.

Percentage of variance in performance explained by deliberate practice was 26% for games ($\bar{r} = .51, p < .001$), 21% for music ($\bar{r} = .46, p < .001$), 18% for sports ($\bar{r} = .42, p < .001$), 4% for education ($\bar{r} = .21, p < .001$), and less than 1% for professions ($\bar{r} = .05, p = .62$) (see Figure 6.3).

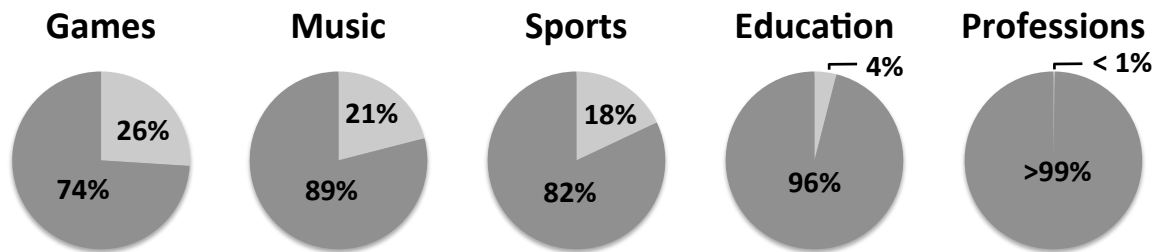


Figure 6.3. Percentage of Performance Variance Explained vs. Unexplained by Deliberate Practice Within Domains. Each pie chart depicts the percentage of variance in performance explained by deliberate practice (light gray area) vs. unexplained by deliberate practice (dark gray area). Percentage of variance explained is equal to $\bar{r}^2 * 100$.

Predictability of the task environment was also a statistically significant moderator, $Q(1) = 20.49, B = .14, T^2 = .05, p < .001$. As hypothesized, percentage of variance in performance explained by deliberate practice was largest for activities high in predictability at 24% ($\bar{r} = .49$), intermediate for activities moderate in predictability at 12% ($\bar{r} = .35$), and smallest for activities low in predictability at 4% ($\bar{r} = .21$).

Methodological moderators. Method used to assess deliberate practice was a statistically significant moderator, $Q(2) = 16.19, p < .001$. Percentage of variance in performance explained by deliberate practice was 20% for studies that used a retrospective interview ($\bar{r} = .45$,

$p < .001$), 12% for studies that used a retrospective questionnaire ($\bar{r} = .34, p < .001$), and 5% for studies that used a log method ($\bar{r} = .22, p < .001$).¹⁶

Method used to assess performance was also a statistically significant moderator, $Q(3) = 14.41, p = .002$. Percentage of variance in performance explained by deliberate practice was 26% for studies that used group membership ($\bar{r} = .51, p < .001$), 14% for studies that used laboratory tasks ($\bar{r} = .37, p < .001$), 9% for studies that used expert ratings ($\bar{r} = .30, p < .001$), and 8% for studies that used standardized objective scoring measures ($\bar{r} = .28, p < .001$).

Additional Meta-analytic Models

Three additional models were conducted. The first model excluded the 38 effect sizes for team sports, leaving 119 effect sizes (games = 11, music = 28, individual sports = 22, education = 51, professions = 7). This model was run because interpretation of correlations between deliberate practice and performance in team sports is complicated by the fact that an individual's performance is not independent of the team's performance (Hutchinson, Sachs-Ericsson, & Ericsson, 2013). Overall percentage of variance explained by deliberate practice was 11% in this model (games = 26%, music = 21%, sports = 19%, and education = 4%, all $ps < .001$; professions <1%, $p = .62$).

The second model included only the 59 effect sizes for *solitary* deliberate practice (games = 6, music = 9, sports = 14, education = 30, professions = 0). This model was tested to address the question of whether deliberate practice must be performed in isolation to be maximally effective (Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005; Ericsson et al., 1993). Overall percentage of variance explained by deliberate practice was 11% in this model (games =

¹⁶Whether deliberate practice hours were transformed for analyses was not a statistically significant moderator, $Q(1) = 1.77, p = .18$.

23%, music = 23%, sports = 22%, and education = 3%, all $ps < .001$), indicating that solitary deliberate practice is not a stronger predictor of performance than deliberate practice with others.

The third model only included the 53 effect sizes for solitary deliberate practice and excluded effect sizes for team sports (games = 6, music = 9, individual sports = 8, education = 30). Overall percentage of variance explained by deliberate practice was 10% in this model (games = 23%, music = 23%, sports = 28%, and education = 3%, all $ps < .001$).

Thus, results were similar across the additional analyses, and results of these analyses were consistent with the overall analysis in indicating that deliberate practice explains a considerable amount of the variance in performance, but leaves a large amount unexplained and potentially explainable by other factors.

Publication Bias Analyses

Publication bias analyses were conducted to investigate whether null or weak results have been systematically suppressed from publication in the deliberate practice literature, and to investigate whether there were effect sizes missing from our meta-analysis due to publication bias. The first step was to create a funnel plot depicting the relationship between study size and effect size. Visual inspection of the funnel plot indicated that it was approximately symmetrical, suggesting that smaller studies with weak effect sizes were not missing from our meta-analysis. A trim-and-fill analysis (Duval & Tweedie, 2000a, 2000b) confirmed this, indicating that zero effects prone to publication bias were missing.

6.5 Discussion

More than twenty years ago, Ericsson et al. (1993) argued that “individual differences in ultimate performance can largely be accounted for by differential amounts of past and current

levels of practice” (Ericsson et al., 1993, p. 392). Ericsson and Moxley (2011) reiterated this claim, stating that “[t]he concept of deliberate practice can account for the large individual differences between experts and novices” (p. 145). The results of this meta-analysis do not support these strong claims. Regardless of domain, deliberate practice leaves a large amount of variance in performance unexplained and potentially explainable by other factors. We conclude that amount of deliberate practice—while important as a predictor of individual differences in performance from both a statistical and a practical perspective—is not as important as Ericsson and colleagues have argued.

Moderator analyses revealed that the strength of the relationship between deliberate practice and performance varied by domain. In terms of percentage of variance in performance explained, the effect of deliberate practice was strong for games (26%), music (21%), and sports (18%), and much weaker for education (4%) and professions (<1% and not statistically significant). Why were the effect sizes for education and professions so much smaller? One possibility is that deliberate practice is less well defined in these domains. It could also be that, in some of the studies, participants differed in amount of pre-study expertise (i.e., amount of domain knowledge prior to taking an academic course or accepting a job), and thus in the amount of time they needed to engage in deliberate practice to achieve a given level of performance.

Moderator analyses further revealed that the effect of deliberate practice on performance tended to be larger for activities that are highly predictable (e.g., running) than for activities that are less predictable (e.g., handling an aviation emergency), as we hypothesized. Furthermore, the effect of deliberate practice on performance was stronger for studies that used retrospective methods to elicit estimates of deliberate practice than for those that used a log method. In fact,

for studies that used the log method, which presumably yields more valid estimates than retrospective methods, deliberate practice accounted for only 5% of the variance in performance. This finding suggests that the use of what Ericsson (2014) termed a “high-fidelity” (p. 13) approach to assessing deliberate practice, an example of which might be video monitoring, may reveal that the relationship between deliberate practice and performance is weaker than the results of this meta-analysis indicate. Finally, the relationship between deliberate practice was weaker for studies that used a standardized objective measure of performance (e.g., chess rating) than for studies that used group membership as the measure of performance.

Individual effect sizes were not corrected for the attenuating effect of measurement error (i.e., measurement unreliability), because very few studies in the meta-analysis reported a reliability estimate for both deliberate practice and performance. However, measures of deliberate practice and performance typically demonstrate “acceptable” or better reliability ($\geq .70$). For example, Tuffiash, Roring, and Ericsson (2007) stated that “self-report practice estimates repeatedly from experts in sports and music have reported test-retest reliabilities at or above .80” (p. 129), and Hambrick et al. (2014) found reliability of .91 for chess rating. Furthermore, using the standard formula for correcting a correlation for measurement unreliability¹⁷, the percentage of variance in performance explained by deliberate practice is smaller than the percentage of variance unexplained across a wide range of reliability assumptions. For example, if it is assumed that reliability of both deliberate practice and performance is .80, the mean overall correlation between deliberate practice and performance is $\bar{r} = .43$ after correction for unreliability, indicating that deliberate practice accounts for 19% of the reliable variance, leaving 81% potentially explainable by other factors; corresponding values

¹⁷ $\hat{r} = r_{xy} / (r_{xx} r_{yy})^{1/2}$, where r_{xx} and r_{yy} are reliability coefficients for x and y, respectively (Schmidt & Hunter, 1996).

(percentage of variance explained) by domain are 41% for games, 33% for music, 28% for sports, 7% for education, and <1% for professions.

What explains variance in performance that deliberate practice leaves unexplained? There are likely many factors. One may be the age at which one starts serious involvement in a domain. Ericsson et al. (1993) argued that any performance advantage associated with starting age simply reflects the fact that a person who starts at a young age has more time to accumulate deliberate practice than a person who starts at a later age. However, inconsistent with this account, Gobet and Campitelli (2007) and Howard (2012) found that starting age negatively predicted chess rating even after statistically controlling for deliberate practice. This evidence suggests that there may be an optimal developmental period for acquiring complex skills, as there seems to be for acquiring language (Lenneberg, 1967).

Research suggests that general intelligence and more specific abilities may also explain some of the variance in performance that deliberate practice leaves unexplained. General intelligence (Hunt, 2011; Jensen, 1998)—which is highly stable and substantially heritable (Plomin, DeFries, McClearn, & McGuffin, 2008)—positively predicts performance in a wide range of domains, including music (Shuter, 1968), chess (Grabner, Stern, & Neubauer, 2007), academics (e.g., Binet & Simon, 1911; Brody, 1997; Laidra, Pullmann, & Allik, 2006), and virtually any occupation (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004). Working memory capacity—the ability to maintain information in the focus of attention (Engle, 2002)—is an example of a specific ability that may predict performance differences. Mainz and Hambrick (2010) found that working memory capacity positively predicted pianists' performance in a sight-reading task, above and beyond deliberate practice. There was no significant interaction between deliberate practice and working memory capacity, indicating that working memory

capacity was as important a predictor of performance for beginning pianists as it was for pianists who had engaged in thousands of hours of deliberate practice.

Conclusion

Ericsson and colleagues' (1993) deliberate practice view has significantly increased scientific understanding of expert performance, but their claim that individual differences in performance is largely accounted for by individual differences in amount of deliberate practice is not supported by the available empirical evidence. An important goal for future research on expert performance is to draw upon existing theories of individual differences (e.g., Ackerman, 1987; Gagné, 2009; Schmidt, 2014; Simonton, 2013) to identify basic abilities and other individual-difference factors that explain variance in performance, and to estimate their importance as predictor variables relative to deliberate practice. Another important goal is to continue to investigate how and when task and situational factors such as task predictability moderate the impact of deliberate practice and other individual-difference factors on performance. Research aimed at addressing these goals will shed new light on the underpinnings of expert performance.

Chapter 7

Conclusion and future directions

7.1 Chapter summary

This chapter summarizes empirical findings, contributions and implications, and future directions and conclusions.

- Chapter 2: Expertise in a complex bilingual processing task is predicted by domain-general cognitive abilities and not domain-specific experience.
- Chapter 3: Domain-cognitive abilities are enhanced from experience with a complex bilingual processing task.
- Chapter 4: Both domain-specific experience and general cognitive abilities predict early skill acquisition in a complex bilingual processing task.
- Chapter 5: Predictability of the task environment moderates the relationship between domain-specific experience and performance and the relationship between domain-general cognitive abilities and performance.
- Chapter 6: A meta-analysis revealed that domain-specific experience—deliberate practice—accounts for about 12% of performance variance overall and that this amount is moderated by domain and predictability of the task environment.

7.2 Introduction

The goal of the present research program is to gain a fundamental understanding of the mechanisms and moderators of skill acquisition, the underlying cognitive abilities, and the interrelations among experiential factors, cognitive abilities and performance variance. This research integrates three distinct areas of investigation: the study of abilities as predictors of performance (i.e., prototypical individual differences research), the study of practice and training as predictors of performance (i.e., prototypical skill acquisition and expertise research), and the study of changes to general cognitive abilities from specific types of experience (i.e., bilingual advantage research).

In order to integrate the three areas of investigation, a complex bilingual processing task—simultaneous interpreting—was chosen as the primary performance task for three of the five studies, two that investigated predictors of performance and one that investigated enhancements to cognitive abilities. The two other studies included a laboratory study investigating a potential moderator of the explanatory power of both domain-specific experience and cognitive abilities on performance, and a meta-analysis investigating the explanatory power of domain-specific cognitive abilities on performance variance and the role of potential moderators on this relationship.

7.3 Overview of empirical findings

Chapter 2 reported a field study in which measures of experiential factors, cognitive abilities, and affective traits were administered to expert and non-expert simultaneous ASL-English simultaneous interpreters. There were no differences between the experts and the non-experts on any demographic or experiential factor variables, including years of professional

interpreting experience. Discriminant function analysis revealed that a combination of general cognitive abilities and an affective trait—mental flexibility, task switching speed, psychomotor speed, perceptual speed, and risk sensitivity—reliably differentiated experts from non-experts. These results suggest that cognitive abilities and not experience are important for explaining variance in high levels of simultaneous interpreting performance. However, the design of the study limited the causal inferences that could be made about whether the expert interpreters were born with the cognitive profile or whether cognitive abilities had become enhanced among some individuals.

The goal of the study reported in Chapter 3 was to investigate when and how cognitive abilities become enhanced from bilingual experience. The primary hypothesis for the source of the bilingual advantage—the finding that bilinguals tend to outperform monolinguals on certain measures of cognitive abilities—is that experience engaging general cognitive in order to manage two languages enhances those cognitive abilities. Supporting this hypothesis is evidence that ASL-English bilinguals do not demonstrate such an advantage, presumably because the distinct language modalities reduce conflict and control demands (Emmorey et al., 2008). However, inter-language conflict alone does not account for other findings in which bilinguals do not outperform monolinguals, such as among college students (e.g., Paap & Greenberg, 2013). The study presented in Chapter 3 was designed based on the hypothesis that that the mechanism responsible for the bilingual advantage is the interplay between (a) the magnitude of bilingual management demands and (b) the amount of experience managing those demands. That is, that bilinguals with experience engaging in highly demanding bilingual activity, such as simultaneous interpreting, will demonstrate cognitive enhancements, regardless of the inter-language conflict. In order to test this hypothesis, adult ASL-English bilinguals with high bilingual management

demands were recruited and cognitive control and working memory capacity were examined longitudinally. After gaining experience managing high bilingual management demands, participants outperformed themselves from 2 years earlier on cognitive abilities associated with managing the bilingual demands of simultaneous interpreting. These results suggest that cognitive control outcomes for bilinguals vary as a function of the mechanisms recruited during bilingual management and the amount of experience managing the bilingual demands.

Chapter 4 presented a study investigating the roles of domain-specific experience—training—and general cognitive abilities on early skill acquisition of ASL-English simultaneous interpreting over a two-year period. Change in cognitive abilities followed the same patterns revealed in Chapter 3. Training had a significant effect the early stages of skill acquisition: the more training accumulated, the better the simultaneous interpreting performance. Also, general cognitive abilities predicted simultaneous interpreting performance. This relationship was positive and linear, generally following the same pattern as the relationships revealed in Chapter 2. There was no positive relationship between change in cognitive abilities and simultaneous interpreting performance, suggesting that the higher one's cognitive abilities, the better one's simultaneous interpreting performance, regardless of whether the cognitive abilities were initially high and remained so or whether they became heightened with experience.

The first three studies all investigate ASL-English simultaneous interpreting performance and experience in the field. The studies presented in Chapter 5 presents a controlled environment in which to study skill acquisition more generally. The experiments are the first of their kind to investigate the role of predictability of the task environment—which takes into account the rate of change of the task environment and the range of potential stimuli the performer is likely to encounter—on performance. Participants completed five blocks in which the two aspects of

predictability were parametrically manipulated with interleaved deliberate practice. In the first experiment, participants were randomly assigned to one of four groups in which they either completed the two conditions where rate of change was low—static at the time of performer response—or high—completely in flux—during performer response. Half the participants completed a condition in which the range of probable stimuli is highly constrained—pattern movements—followed by a condition in which the range of probable stimuli has low constraints—random movements. The other half completed these conditions in the opposite order. In the second experiment, participants were randomly assigned to one of two groups in which they completed two conditions where the range of probable stimuli was highly constrained—pattern movements—or had low constraints—random movements. All participants completed the condition in which the rate of change was low—static at the time of performer response—and then the condition in which the rate of change was high—completely in flux at the time of performer response. The results supported the hypothesis that when the task environment is highly predictable, that is, the stimuli are static at the time of performer response, and the range of probable stimuli is highly constrained, practice best accounted for performance variance. Performance on conditions that were semi-predictable correlated to performance on cognitive ability measures. The results of these experiments suggest that predictability of the task environment moderates the relationships between practice and performance variance and between general cognitive abilities and performance variance.

Chapter 6 presented a meta-analysis investigating the relationship between practice and performance variance across all domains in which deliberate practice has been studied, including sports, music, games, education, and professions. Predictability of the task environment was also tested as a moderator variable. Deliberate practice explained a much smaller proportion of

performance variance than proponents of deliberate practice claim, only 12% overall. This amount varied by domain—highest in games, lowest in professions and by predictability of the task environment—highest for highly predictable tasks, lowest for less predictable tasks.

Taken together, the empirical findings suggest that 1) domain-specific experience is important during initial skill acquisition, regardless of the task's predictability, 2) domain-specific experience explains less variance in performance after initial skill acquisition in tasks with low predictability than in tasks with high predictability, 3) general cognitive abilities explain less variance in performance in tasks with high predictability than in tasks with low predictability, and 4) domain-specific experience in highly demanding tasks affects domain-general cognitive abilities.

7.4 Contributions and Implications

Are experts “born” or “made?” Proponents of the deliberate practice view suggest that experts are made, that only experiential variables are of any import, while most other psychologists assume that both experiential and stable individual differences factors are important without a way to distinguish relative importance. The presented studies and discussion within this document suggest that the deliberate practice view is too extreme to apply to most tasks, and that task demands moderate the relative importance of assumed nature-based and assumed nurture-based variables in explaining individual differences in performance. In other words, the appropriate question to ask is *not* which is correct, nature or nurture, but when and why does the relative contribution of each vary?

The major focus of this document is examining the effect of a particular task characteristic—predictability of the task environment—as a moderator of the relative importance

of nature-based and nurture-based determinants of performance. The current research took into account the types of tasks in which researchers of individual differences typically investigate performance—tasks that demonstrate the importance of general cognitive abilities over practice—and the types of tasks in which researchers of skill acquisition and expertise typically investigate performance—tasks that demonstrate the importance of practice over general cognitive abilities—in order to examine what differentiates these types of tasks, and how that may interact with determinants of performance.

The results of the literature search and present studies suggest that experiential factors may be more important when the task environment is more predictable than when it is less predictable. That is, when all potential stimuli to which a performer may respond can be learned, when the problem space is narrow enough that the performer can anticipate the upcoming stimuli, when a regimen of activities designed to improve performance in response to those stimuli can be engaged, and when the stimuli are consistent while the performer is executing her response, then amount of task experience, practice, and training will strongly determine skill acquisition rates, expertise, and performance variance; individual differences in cognitive abilities will be less important than experience, practice, and training.

In contrast, individual differences in cognitive abilities may be more important when the task environment is less predictable than when it is more predictable. That is, when all potential stimuli to which a performer may respond are innumerable and therefore cannot be learned or practiced, when the problem space is not narrow enough for the performer to anticipate upcoming stimuli, and when the stimuli are in flux while the performer is executing her response, then individual differences in updating the contents of working memory, problem solving, and reasoning will strongly determine skill acquisition rates, expertise, and performance

variance; amount of task experience, practice, and training will be less important than individual differences in cognitive abilities.

This work contributes to our scientific understanding of human performance and has implications for cognitive psychology, industrial/organizational psychology, developmental psychology, and training and professional domains such as interpreting, education, sports, and music. This research contributes to scientific investigations of cognitive psychology by attempting to understand how we interact with the information presented in our environment. Specifically, this research enhances our understanding of which cognitive processes become engaged depending on the features of the task environment, and therefore which cognitive abilities are predictive of performance under those conditions. This research also contributes to our understanding of when general cognitive abilities can become adaptive to the demands of the task, that is, when cognitive abilities that are generally presumed to be stable become enhanced through continuous activation demanded by the task.

The results of these investigations in turn can inform investigations in industrial/organizational psychology, as well as inform individuals making decisions about their own training or career selection and those who have the power to make such decisions about others' training or career (e.g., hiring parties, admissions administrators, athletic scouts, coaches) based on predictions of future potential or likely performance. This and future research can be passed on to non-scientists and applied to real-world scenarios, allowing individuals to make more informed decisions.

These findings also challenge the deliberate practice view. The deliberate practice view holds that the task must be repeatable—stimuli drawn from a limited pool (Ericsson et al., 1993)—which is inline with the view that deliberate practice explains less variance in

performance when the task is not highly predictable. However, the deliberate practice view also holds that general cognitive abilities are not predictive of performance (Ericsson et al., 1993), which is unsupported by the present research as well as most of the literature within differential psychology.

7.5 Future directions and Conclusion

This body of work represents the first steps of a research program designed to understand complex relationships among experiential factors, general cognitive abilities, performance, and predictability of the task environment. As such, there is a multitude of future directions. The remaining section will describe current and future research projects whose findings will further elucidate the relationships among the aforementioned variables.

Investigating predictors of simultaneous interpreting performance has specific applications. Little is known about the source of the very high failure rate among simultaneous interpreting students discussed in Chapter 2. Research investigating predictors of simultaneous interpreting performance has the potential to produce aptitude and assessment measures as well as enhanced curricula designs. Aptitude and assessment measures benefit the prospective student by providing guidance in their career. Information gained from these measures allow the student to make better decisions about education and career choices before investing thousands of dollars and years engaging in a training program that may not yield a successful career. Aptitude and assessment measures also benefit the training programs by providing potential selection criteria. Information gained from these measures allows the administrators to decide the level of competitiveness of their admissions. Improved curricula designs can work to develop cognitive ability enhancements and tailor training programs to individuals' cognitive profiles.

Another area of future research is to examine the generalizability of the Bilingual Management Demands hypothesis. The support for this hypothesis was limited to bimodal bilingual simultaneous interpreters. While this population provides an ideal testing ground, given that bimodal bilingual non-interpreters—that is, bimodal bilinguals with low bilingual management demands—do not demonstrate a bilingual advantage, this finding may be special to bimodal bilinguals. Current and future research is targeted at examining the effects of bilingual management demands on cognitive abilities among unimodal bilinguals with differing bilingual management demands. Specifically, the cognitive abilities of beginning and advanced spoken-language simultaneous interpreter students, beginning and advanced spoken-language translation students, spoken-language college bilinguals without special bilingual training, and spoken-language second-language students at a range of proficiency levels are all currently being measured. The results of these examinations will provide insight on the role of differing bilingual management demands and experience with these demands on general cognitive abilities among individuals with a variety of bilingual profiles.

More broadly, in addition to the general need for more longitudinal studies investigating skill acquisition and expertise across domains, one future direction is to investigate the bidirectional causal relationship between deliberate practice and performance. While deliberate practice accounted for approximately 12% of the variance in performance in the meta-analysis, this relationship is complicated by the aspect of human nature noted by Sternberg in 1996: we tend to pursue endeavors in which we feel we are performing well and stop pursuing endeavors in which we feel we are failing. In this way, amount of accumulated deliberate practice is likely to reflect, to some degree, ability or talent. In order to test this causal direction of the relationship, that is, the effect of performance on practice rather than practice on performance, a

longitudinal experiment can be designed in which perception of one's performance is manipulated. For this study, a task that incites motivation to perform well and includes a mechanism for providing not-necessarily-true feedback to the performer is all that is necessary. The amount of practice engaged by the performer would be measured. Likely, those told they are performing well (but not at ceiling) relative to others would engage in more practice over time than participants who were told that they are performing below average.

Finally, more work is needed to further clarify and investigate the novel construct introduced in this body of work—predictability of the task environment. One clarification that is necessary is deciphering how predictability of the task environment varies for response-based tasks versus completely self-generated tasks. For completely self-generated tasks (e.g., running, writing a novel) the task environment is internal to the performer and adjustments to the definition are necessary. While these differences were largely clear as indicated by inter-rater reliability in the meta-analysis, more objective assessment measures should be developed in order to examine tasks in the future within this framework.

In addition to an assessment measure of predictability of the task environment, an assessment of the complexity of the task should also be developed. This is a difficult task. As Neil Johnson (2009) pointed out "Even among scientists, there is no unique definition of Complexity. Instead, the scientific notion of Complexity -- and hence of a Complex System -- has traditionally been conveyed using particular examples of real-world systems which scientists believe to be complex" (p. 3). Despite this difficulty, both differences and an interaction between predictability and complexity are likely to emerge. That is, that predictability of the task environment is distinct from complexity of the task, but that complexity might moderate the role of predictability. If one assumes complexity of the task to be the degree of elaborateness among

interconnected components, this may be thought of as the number of task aspects that must be understood in order to perform the task. For example, most people would agree that chess is more complex than checkers. However, according to the Predictability of the Task Environment hypothesis, both chess and checkers are highly predictable—they are both static at the time of performer response and the range of probable immediate future inputs is numerable—and thus practice is likely to be a strong predictor of performance variance. Complexity will increase the amount of practice necessary to reach a given level of performance. In other words, more practice is necessary to become a competent chess player than a competent checkers player. Presumably, this relationship will hold for less predictable tasks. This is, for less predictable tasks, general cognitive abilities will explain more performance variance than amount of experience (after initial skill acquisition). A complex task with low predictability will increase the level of general cognitive abilities necessary to reach a given level of performance. In other words, an individual must have stronger cognitive abilities to competently perform a complex task with low predictability than a simple task with low predictability. Future work should be aimed at testing these hypotheses.

Complex and interesting relationships exist among experiential factors, cognitive abilities, and performance variance. Examinations of interactions among the individual's accumulated experience, cognitive abilities, and features of the task environment are necessary in order to understand any one of these variables, their inter-relationships, and the underpinnings of human performance. From this work, is the potential to understand how cognitive abilities can be altered and improved, to better predict future performance of individuals, and to develop successful training regimens allowing achievement where there otherwise might not have been.

References

- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, 102(1), 3-27.
- Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, 117(3), 288-318.
- Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, 1, 270-304.
- Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: John Wiley and Sons.
- Baker, J., Bagats, S., Büsch, D., Strauss, B. & Schorer, J. (2012). Training differences and selection in a talent identification system. *Talent Development & Excellence*, 4(1), 23–32.
- Baker, J., Côté, J., & Abernathy, B. (2003). Learning from the experts: Practice activities of expert decision-makers in sport. *Research Quarterly for Exercise and Sport*, 74, 342-347.
- Baker, J., Deakin, J. & Côté, J. (2005). On the utility of deliberate practice: Predicting performance in ultra-endurance triathletes from training indices. *International Journal of Sport Psychology*, 36, 225-240.
- Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, 83, 413–422.

- Bayliss, D. M., Jarrold, C., Gunn, D. M. & Baddeley, A. D. (2003). The complexity of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General* 132, 71–92.
- Becker, G. (1986). Correcting the point-biserial correlation for attenuation owing to unequal sample size. *The Journal of Experimental Education*, 55(1), 5-8.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 39, 15–22.
- Berry, J., Abernathy, B., & Cote, J. (2008). The contribution of structured activity and deliberate play to the development of expert perceptual and decision-making skill. *Journal of Sport & Exercise Psychology*, 30, 685-708.
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, 70, 636–644.
- Bialystok, E. (2009). Claiming evidence from non-evidence: A reply to Morton and Harper. *Developmental Science*, 12, 499–501.
- Bialystok, E., & Craik, F.I.M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science*, 19(1), 19-23.
- Bilalić, M., & McLeod, P. (2007). Does chess need intelligence?—A study with young chess players. *Intelligence*, 35(5), 457-470.
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Englewood, NJ: Biostat.
- Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, 52(10), 1046.
- Bruce, L., Farrow, D., & Raynor, A. (2013). Performance milestones in the development of expertise: Are they critical? *Journal of Applied Sport Psychology*, 25, 281-297.

- Brunborg, G. S., Pallesen, S., Diseth, A., & Larsen, S. (2010). Preoccupation with failure affects number of study hours—Not academic achievement. *Scandinavian Journal of Educational Research*, 54(2), 125-132.
- Burke, H. R. (1972). Raven's Progressive Matrices: Validity, reliability, and norms. *Journal of Psychology: Interdisciplinary and Applied* 82 (2), 253–257.
- Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology* 42 (1), 116–131.
- Calabria, M., Hernández, M., Branzi, F. M., & Costa, A. (2011). Qualitative differences between bilingual language control and executive control: Evidence from task-switching. *Frontiers in Psychology*, 2, 399.
- Carver, C. S. & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology* 67, 319–333.
- Castejón, J. L., Gilar, R. & Pérez, A. M. (2006). Complex learning: The role of knowledge, intelligence, motivation and learning strategies. *Psicothema*, 18(4), 679-685.
- Cathey, R. M. (2010). Retrospective practice histories of expert and novice baseball pitchers. Unpublished dissertation. University of South Carolina.
- Catteeuw, P., Helsen, W., Gilis, B., & Wagemans, J. (2009). Decision-making skills, role specificity, and deliberate practice in association football refereeing. *Journal of Sports Sciences*, 27(11), 1125-1136.
- Cattell, R. B., Cattell, A. K. S., & Institute for Personality and Ability Testing. (1960). *Measuring intelligence with the culture fair tests*. Champaign, IL: Institute for Personality and Ability Testing.

- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 641.
- Charness, N., Krampe, R. Th., & Mayr, U. (1996) The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.) *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games*. Mahwah, NJ: Erlbaum, pp. 51-80.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, 19, 151-165.
- Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. *Cognitive skills and their acquisition*, 141-189.
- Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *The psychology of learning and motivation*, 16, 1-58.
- Chi, M. T. H., R. Glaser, & M. J. Farr (Eds.). (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Chinn, D., Sheard, J., Carbone, A., & Laakso, M-J. (2010). Study habits of CS 1 students: What do they do outside the classroom? *Proceedings of the 12th Australasian Computing Education Conference (ACE 2010)*: Brisbane, Australia.
- Christoffels, I. K., & de Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 326–348). New York, NY: Oxford University Press.

- Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language* 54 (3), 324–345.
- Christoffels, I. K., de Groot, A. M. B. & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting. *Bilingualism: Language and Cognition* 6 (3), 201–211.
- Cokely, D. (1992). *Interpretation: A sociolinguistic model*. Burtonsville, MD: Linstok Press.
- Coleman, M.; and Liao, T. L. (1975); *A computer readability formula designed for machine scoring*, Journal of Applied Psychology, Vol. 60, pp. 283–284.
- Colonomos, B. (1997). *Pedagogical model of the interpreting process*. College Park, MD: Bilingual Mediation Center.
- Colonomos, B. (2008). *Revised pedagogical model of the interpreting process*. College Park, MD: Bilingual Mediation Center.
- Colvin, G. (2008). *Talent is overrated: What really separates world-class performers from everybody else*. New York: Penguin Group.
- Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30(2), 163-183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769-786.
- Cowan, N. (1995). *Attention and memory: an integrated framework*. Oxford: Oxford University Press.

- Cowan, N. (2000). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting* 5 (2), 117–146.
- Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill. *Ergonomics*, 2(2), 153-166.
- Da Matta, G. B. (2004). The influence of deliberate practice and social support systems on the development of expert and intermediate women volleyball players in Brazil. Unpublished dissertation. University of South Carolina.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of verbal learning and verbal behavior*, 19(4), 450-466.
- Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... & Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular psychiatry*, 16(10), 996-1005.
- de Bruin, A. B. H., Kok, E. M., Leppink, J. M., & Camp, G. (2014). Practice, intelligence, and enjoyment in novice chess players: A prospective study at the earliest stage of a chess career. *Intelligence*, published online ahead of print.
- de Bruin, A. B. H., Rikers, R. M. J. P., Schmidt, H. G. (2007). The influence of achievement motivation and chess-specific motivation on deliberate practice. *Journal of Sport & Exercise Psychology*, 29, 561-583.
- de Groot, A. (1978). *Thought and choice and chess*. The Hague: Mouton. (Original work published 1946).
- Dean, R. K. & Pollard, Jr., R. Q. (2005). Consumers and service effectiveness in interpreting work: A practice profession perspective. In M. Marschark, R. Peterson & E. A. Winston

- (Eds.), *Sign language interpreting and interpreter education: Directions for research and practice*. New York: Oxford University Press, 259–282.
- Deary, I., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.
- Diseth, Å. (2007). Approaches to learning, course experience and examination grade among undergraduate psychology students: testing of mediator effects and construct validity. *Studies in Higher Education*, 32(3), 373–388.
- Diseth, Å., Pallesen, S., Brunborg, G. S., & Larsen, S. (2010). Academic achievement among first semester undergraduate psychology students: the role of course experience, effort, motives and learning strategies. *Journal of Higher Education*, 59, 335–352.
- Doumen, S., Broeckmans, J., & Masui, C. (2011). Study time and academic performance: A conditional relation? Paper presented at the 16th ELSIN conference in Antwerp, June 30, 2011.
- Duckworth, A. L., Kirby, T. A., Tsukayama, E., Berstein, H., & Ericsson, K. A. (2011). Deliberate practice spells success: Why grittier competitors triumph at the national spelling bee. *Social Psychological and Personality Science*, 2(2), 174-181.
- Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sport Psychology*, 35, 232-245.
- Duval, S. J., & Tweedie, R. L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463.

- Duval, S. J., & Tweedie, R. L. (2000b). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89-98.
- Ekstrand, J. (2007). Measuring and explaining economics students learning efficiency. Unpublished document.
- Elferink-Gemser, M. T., Starkes, J. L., Medic, N., Lemmink, K. A.P.M., & Visscher, C. (2011). What discriminates elite and sub-elite youth field hockey players? *Annals of Research in Sport and Physical Activity*, 50-68.
- Emmorey, K., Borinstein, H. B., Thompson, R., & Gollan, T. H. (2008a). Bimodal bilingualism. *Bilingualism*, 11, 43–61.
- Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008b). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science*, 19, 1201–1206.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology: General*, 128, 309 – 331.
- Ericsson, K. A. (2007). Deliberate practice and the modifiability of the body and mind: Toward a science of the structure and acquisition of expert and elite performance. *International Journal of Sport Psychology*, 38(1), 4-34.
- Ericsson, K. A. (2013). *Psychotherapy and the science of human excellence*. Keynote presentation at the 2013 Psychotherapy Networker Symposium, Washington, DC.

- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211-245.
- Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. *Annual Review of Psychology*, 47, 273-305.
- Ericsson, K. A., & Moxley, J. H. (2011). The expert performance approach and deliberate practice: Some potential implications for studying creative performance in organizations. In M. D. Mumford (Ed.), *The handbook of organizational creativity* (pp. 141-167). Academic Press.
- Eysenck, M. W. (1979). Anxiety, learning and memory: A reconceptualization. *Journal of Research in Personality* 13, 363–385.
- Eysenck, M. W. (1985). Anxiety and cognitive task performance. *Personality and Individual Differences* 6 (5), 579–586.
- Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological science*, 18(10), 850-855.
- Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions*, 6(5), 1–12.
- Ford, P. R., Ward, P., Hodges, N. J., & Williams, A. M. (2009). The role of deliberate practice and play in career progression in sport: the early engagement hypothesis. *High Ability Studies*, 20(1), 65-75.

- Ford, P. R. & Williams, A. M. (2011). The developmental activities engaged in by elite youth soccer players who progressed to professional status compared to those who did not. *Psychology of Sport Exercise*, 1-4.
- Frauenfelder, U. H. & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting* 2, 55–89.
- Gagné, F. (2009). Talent development as seen through the differentiated model of giftedness and talent. In T. Balchin, B. Hymer, & Matthews, D. (Eds.) *The Routledge international companion to gifted education*, 32-41.
- Galton, F. (1869). Hereditary genius. London: Macmillan.
- Gardner, H. (1995). "Expert performance: Its structure and acquisition": Comment. *American Psychologist*, 50(9), 802-803.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Brislin (Ed.), *Translation: Applications and research*. New York: Gardner Press, 165–207.
- Gerver, D., Longley, P. E., Long, J. & Lambert, S. (1989). Selection tests for trainee conference interpreters. *Meta* 34 (4), 724–735.
- Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 196–214). Thousand Oaks, CA: Sage.
- Gladwell, M. (2008). *Outliers: The Story of Success*. New York: Little, Brown and Company.

- Gobet, F. & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology*, 43(1), 159-172.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24, 79–132.
- Grabner, R.H., Stern, E., & Neubauer, A.C. (2007). Individual differences in chess expertise: A psychometric investigation. *Acta Psychologica*, 124, 398–420.
- Grant, D. A. & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38(4), 404–411.
- Gray, J. A. (1982). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system*. New York: Oxford University Press.
- Gray, J. A. (1987). *The psychology of fear and stress*. (2nd ed.) Cambridge: Cambridge University Press.
- Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion*, 4(3), 269–288.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism*, 1, 67–81.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534–537.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism*, 1, 67–81.
- Grimm, L. G. & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.

- Gunning, Robert (1952). *The Technique of Clear Writing*. McGraw-Hill. pp. 36–37.
- Guo, A. (2006). Competition preparation and deliberate practice: A study of the 2005 National Spelling Bee finalists. Unpublished dissertation.
- Hallam, S. (1998). The predictors of achievement and dropout in instrumental tuition. *Psychology of Music*, 26, 116-132.
- Hallam, S. (2001). The development of metacognition in musicians: Implications for education. *British Journal of Music Education*, 18(1), 27-39.
- Halpern, D. F., & Wai, J. (2007). The world of competitive Scrabble: Novice and expert differences in visuospatial and verbal abilities. *Journal of Experimental Psychology: Applied*, 13(2), 79-94.
- Hambrick, D. Z. & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology* 44, 339–387.
- Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence*.
- Harris, K. R. (2008). Deliberate practice, mental representations, and skilled performance in bowling. Unpublished dissertation. Florida State University.
- Haugaasen, M., Toering, T., Jordet, G. (2014). From childhood to senior professional football: A multi-level approach to elite youth football player's engagement in football specific activities. *Psychology of Sport & Exercise*, 15.
- Helsen, W. F., Starkes, J. L., Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport & Exercise Psychology*, 20, 12-34.

- Hendry, D. T. (2012). The role of developmental activities on self determined motivation, passion and skill in youth soccer players. Unpublished dissertation. The University of British Columbia.
- Hodges, N. J., Augaitis, L., & Crocker, P. R. E. (in revision). Sport commitment and deliberate practice among male and female triathletes.
- Hodges, N. J., Kerr, T., Starkes, J. L., & Weir, P. L. (2004). Predicting performance times from deliberate practice hours for triathletes and swimmers: What, when, and where is practice important? *Journal of Experimental Psychology: Applied*, 10(4), 219-237.
- Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature of expertise: A sport-specific test of Ericsson, Krampe, and Tesh-Römer's (1993) theory of deliberate practice. *International Journal of Sport Psychology*, 27(4), 400-424.
- Hoffman, R. R. (Ed.). (1992). *The Psychology of Expertise: Cognitive Research and Empirical AI*. New York: Springer-Verlag.
- Howard, R. W. (2012). Longitudinal effects of different types of practice on the development of chess expertise. *Applied Cognitive Psychology*, 26, 359-369.
- Hunt, E. (2011). *Human intelligence*. Cambridge, UK: Cambridge University Press.
- Hunter, J. E., & Schmidt, F. L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, 75(3), 334.
- Hutchinson, C. U., Sachs-Ericsson, N. J., & Ericsson, K. A. (2013). Generalizable aspects of the development of expertise in ballet across countries and cultures: A perspective from the expert performance approach. *High Ability Studies*, 2013.

- Jabusch, H-C., Alpers, H., Kopiez, R., Vauth, H., & Altenmuller, E. (2009). The influence of practice on the development of motor skills in pianists: A longitudinal study in a selected motor task. *Human Movement Science*, 28, 74–84.
- Jabush, H-C., Yong, R., & Altenmuller, E. (2007). *Biographical predictors of music-related motor skills in children pianists*. Proc. International Symposium on Performance Science, In Williamon, Aaron & Coimbra, Daniela (Eds.). Utrecht, The Netherlands: The European Association of Conservatoires (AEC).
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829-6833.
- Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Johnson, M. B., Tenenbaum, G., & Edmonds, W. A. (2006). Adaptation to physically and emotionally demanding conditions: The role of deliberate practice. *High Ability Studies*, 17(1), 117-136.
- Johnson, N. F. (2009). *Simply complexity: A clear guide to complexity theory*. Oxford: Oneworld.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99, 122–149.
- Kamp, R. J., Dolmans, D. H., van Berkel, H. J., & Schmidt, H. G. (2012). The relationship between students' small group activities, time spent on self-study, and achievement. *Higher Education*, 64(3), 385-397.

- Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive *g* and academic achievement *g* one and the same *g*? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, 40, 123–138.
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L., & Chissom, B.S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease formula) for Navy Enlisted Personnel. Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95, 163–182.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363–394.
- Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14(7), 317-324.
- Kopiez, R., Jabusch, H-C., Galley, N., Homann, J-C., Lehmann, & Altenmuller, E. (2011). No disadvantage for left-handed musicians: The relationship between handedness, perceived constraints and performance-related skills in string players and pianists. *Psychology of Music*, 40(3), 357-384.
- Kopiez, R. & Lee, J. I. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, 10(1), 41-62.
- Köpke, B. & Nespoulous, J. (2006). Working memory performance in expert and novice interpreters. *Interpreting* 8 (1), 1–23.
- Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, 128, 416–430.

- Krampe, R. Th. & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General* 125(4), 331-359.
- Lambert, S. (1992). Shadowing. *Meta*, 37, 263–273.
- Law, M. P., Cote, J., Ericsson, K. A. (2007). Characteristics of expert development in rhythmic gymnastics: A retrospective study. *International Journal of Sport and Exercise Psychology*, 5(1), 82-103.
- Lehmann, A. C. & Ericsson, K. A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, 15, 1-29.
- Lemay, Be'dard, Rouleau, & Gilbert Tremblay (2004). Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, 18(2), 1-19.
- Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: John Wiley and Sons.
- Leon, M. R. & Revelle, W. (1985). The effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology* 49, 1302–1315.
- Levitt, S. D. & Dubner, S. J. (2009). *SuperFreakonomics: Global cooling, patriotic prostitutes, and why suicide bombers should buy life insurance*. New York: HarperCollins.
- Limanond, T., Jomnonkwao, S., Watthanaklang, D., Ratanavaraha, V., & Siridhara, S. (2011). How vehicle ownership affect time utilization on study, leisure, social activities, and academic performance of university students? A case study of engineering freshmen in a rural university in Thailand. *Transport Policy*, 18, 719-726.

- Lopes, L. (1987). Between hope and fear: The psychology of risk. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. San Diego: Academic Press, 255–296.
- Lopez, E. J., Nandagopal, K., Shavelson, R. J., Szu, E. & Penn, J. (2013). Self-regulated learning study strategies and academic performance in undergraduate organic chemistry: An investigation examining ethnically diverse students. *Journal of Research in Science Teaching*, 50, 660-676.
- López Gómez, M. J., Bajo Molina, T., Padilla Benítez, P. & Santiago de Torres, J. (2007). Predicting proficiency in signed language interpreting: A preliminary study. *Interpreting* 9 (1), 71–93.
- Lovett, M., Reder, L., & Lebiere, C. (1999). An ACT-R perspective on working memory. In: A. Miyake, & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control*. (pp. 135 – 182). New York: Cambridge University Press
- Loyens, S. M. M., Gijbels, D., Coertjens, L., & Cote, D. J. (2013). Students' approaches to learning in problem-based learning: Taking into account professional behavior in the tutorial groups, self-study time, and different assessment aspects. *Studies in Educational Evaluation*, 39, 23-32.
- Loyens, S. M. M., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The impact of students' conceptions of constructivist assumptions on academic achievement and drop-out. *Studies in Higher Education*, 32(5), 581–602.
- Macnamara, B. (2009). Foundational cognitive aptitudes of sign language interpreters. *Journal of Interpretation* 2008–2009, 9–32.

- Macnamara, B. N. & Conway, A. R. A. (in prep). Skill acquisition in a complex bilingual processing skill.
- Macnamara, B. N. & Conway, A. R. A. (2014). Novel evidence in support of the bilingual advantage: Influence of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review*, 520-525.
- Macnamara, B. N. & Conway, A. R. A. C. (2014, March). *Cognitive Enhancement from Sign Language Interpreting: The Interpreter Advantage*. Presented at the International Symposium on Signed Language Interpretation and Translation Research, Gallaudet University, Washington, D.C.
- Macnamara, B. N., & Conway, A. R. A. (2013, May). *Behind the Bilingual Advantage: Cognitive Enhancements Depend on Magnitude of Bilingual Demands*. Poster presented at the 25th Association for Psychological Science Annual Convention, Washington, DC.
- Macnamara, B. N., & Conway, A. R. A. (2012, November). *Beyond Bilingual: The Interpreter Advantage*. Poster presented at the annual meeting of the Psychonomic Society, Minneapolis, MN.
- Macnamara, B. N., Drach, R., Chow, M., & Conway, A. R. A. (in prep). The role of predictability of the task environment in skill acquisition.
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (in press). Deliberate Practice and Performance in Music, Games, Sports, Education, and Professions: A Meta-analysis. *Psychological Science*.
- Macnamara, B. N., Hambrick, D. Z., & Oswald, F. L. (2013, November). *Deliberate practice and performance in music, games, sports, professions, and education: A meta-analysis*. Poster presented at the annual meeting of the Psychonomic Society, Toronto, ON.

- Macnamara, B. N., Moore, A., & Conway, A. R. A. (2009, May). *Domain-general cognitive abilities and interpreter skill prediction*. Paper presented at the Aptitude for Interpreting: Towards Reliable Admission Testing symposium, Antwerp, Belgium.
- Macnamara, B. N., Moore, A. B., Kegl, J. A., & Conway, A. R. A. (2011). Domain-general cognitive abilities and interpreter skill prediction. *Interpreting*, 13(1), 121-142.
- MacWhinney, B. (1997). Simultaneous interpretation and the Competition Model. In J. H. Danks, G. M. Shreve, S. B. Fountain & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting*. Thousand Oaks, CA: Sage, 215–233.
- Marcus, G. (2012). *Guitar zero: The science of becoming musical at any age*. New York: Penguin.
- Masui, C., Broeckmans, J., Doumen, S., Groenen, A. & Molenberghs, G. (2012). Do diligent students perform better? Complex relations between student and course characteristics, study time, and academic performance in higher education. *Studies in Higher Education*, 2012, 1–23.
- Maynard, D., Hambrick, D. Z., & Meinz, E. J. (2014). Practice vs. play as predictors of individual differences in bowling skill. Unpublished data.
- McKinney, E. H. & Davis, K. J. (2003). Effects of deliberate practice on crisis decision performance. *Human Factors*, 45(3), 436-444.
- McLaughlin, G. H. (1969). SMOG grading—a new readability formula. *Journal of Reading*, 12(8), 639–646.
- McPherson, G. E. (2005). From child to musician: skill development during the beginning stages of learning an instrument. *Psychology of Music*, 33(5), 5-35.

- Meinz, E. J. (2000). Experience-based attenuation of age-related differences in music cognition tasks. *Psychology and Aging, 15*(2), 297-312.
- Meinz, E. J. & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science, 21*(7), 914-919.
- Memmert, D., Baker, J. & Bertsch, C. (2010). Play and practice in the development of sport-specific creativity in team ball sports. *High Ability Studies, 21*(1), 3–18.
- Milner, B. (1962). Les troubles de la memoire accompagnant des lesionshippocampiques bilaterales. In *Physiologie de l'hippocampe*. Paris: C. N. R.S., [English translation (1965) in P. M. Milner & S. Glickman (Eds.) *Cognitive processes and the brain*, Princeton: Van Nostrand.]
- Moesch, K., Elbe, A.-M., Hauge, M.-L. T., & Wikman, J. M. (2011). Late specialization: The key to success in centimeters, grams, or seconds (cgs) sports. *Scandinavian Journal of Medicine & Science in Sports, 1-9*.
- Monsell, S. (2003). Task switching. *TRENDS in Cognitive Sciences 7* (3), 134–140.
- Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review, 18*(1), 46-60.
- Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science, 10*, 719–726.
- Moser, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication*. New York/London: Plenum Press, 353–368.

NAD-RID National Interpreter Certification Candidate Handbook (2013). Alexandria, VA:

Registry of Interpreters for the Deaf.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment*, 21, 244–260.

Oberauer, K. (2004). The measurement of working memory capacity. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications, 393–407.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66, 232–258.

Padilla, P., Bajo, M. T., Cañas, J. J. & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tammola (Ed.), *Topics in interpreting research*. Turku: University of Turku, Centre for Translation and Interpreting, 61–71.

Plant, E. A., Ericsson, K. A., Hill, L., Asberg, K. (2005). Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. *Contemporary Educational Psychology*, 30, 96–116.

Plomin, R., Shakeshaft, N. G., McMillan, A., & Trzaskowski, M. (2013). Nature, nurture, and expertise. *Intelligence*.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2008). *Behavioral Genetics* (Vol. 5). New York: Worth.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of International Neuropsychological Society*, 17, 682–691.

- Pyers, J. E., & Emmorey, K. (2008). The face of bimodal bilingualism. *Psychological Science*, 19, 531.
- Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis.
- Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92.
- Rosário, P. & Núñez, J. C., & Valle, A., González-Pienda, J., & Lourenço, A. (2012). Grade level, study time, and grade retention and their effects on motivation, self-regulated learning strategies, and mathematics achievement: a structural equation model. *European Journal of Psychology of Education*, 1-21.
- Rudser, S. F., & Strong, M. (1986). An examination of some personal characteristics & abilities of sign language interpreters. *Sign Language Studies* 53, 315–331.
- Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. *Intelligence*, 36, 330-338.
- Ryabov, I. (2012). The effect of time online on grades in online sociology courses. *Journal of Online Learning and Teaching*, 8(1), 1-12.
- Salthouse, T. A. & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27(5), 763–776.
- Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M. et al. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology*, 14(1), 102–111.
- Salthouse, T. A., & Tucker-Drob, E. M. (2008). Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*, 22, 800–811.

- Schaap, L., Schmidt, H. G., & Verkoeijen, P. P. J. L. (2012). Assessing knowledge growth in a psychology curriculum: which students improve most? *Assessment & Evaluation in Higher Education*, 37(7), 875–887.
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440–452.
- Schmidt, F. L. (2014). A general theoretical integrative model of individual differences in interests, abilities, personality traits, and academic and occupational achievement: A commentary on four recent articles. *Perspectives on Psychological Science*, 9, 211–218.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schneider, S. L., & Lopes, L. L. (1986). Reflection in preferences under risk: Who and when may suggest why. *Journal of Experimental Psychology: Human Perception and Performance* 12, 535–548.
- Schultetus, S. & Charness, N. (1997). Fencing data. Unpublished data. Cited in Deakin, J. M., & Copley, S. (2003). *A search for deliberate practice: Expert performance in sports: Advances in research on sport expertise*. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research on sport expertise* (115–136). Champaign, IL: Human Kinetics.
- Seal, B. C. (2004). Psychological testing of sign language interpreters. *Journal of Deaf Studies and Deaf Education*, 9(1), 39–52.

- Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreters' Newsletter* 12, 37–49.
- Shuter, R. (1968). *The Psychology of Musical Ability*. Methuen, London, pp. 113-176.
- Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, 61, 394-403.
- Simonton, D. K. (2013). Creative performance, expertise acquisition, individual differences, and developmental antecedents: An integrative research agenda. *Intelligence*.
- Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of practice in the development of performing musicians. *British Journal of Psychology*, 87, 287-309.
- Smith, A. (1992). Retrospective practice histories of Division I and Division II female basketball players in the Carolinas. Unpublished dissertation. University of South Carolina.
- Smith, E. A. & Senter, R. J. (1967). *Automated readability index*. AMRL-TR-6620. Dayton, OH: Wright-Patterson Air Force Base.
- Snelling, P. C., Lipscomb, M., Lockyer, L., Yates, S., & Young, P. (2010). Time spent studying on a pre-registration nursing programme module: An exploratory study and implications for regulation. *Nurse Education Today*, 30, 713-719.
- Snoddy, G. S. (1926). Learning and stability. *Journal of Applied Psychology*, 10, 1-36.
- Sonnentag, S. & Kleine, B. M. (2000). Deliberate practice at work: A study with insurance agents. *Journal of Occupational and Organizational Psychology*, 73, 87-102.
- Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between bilingual language switching and executive functions? Introducing a within-group analysis approach. *Frontiers in Psychology*, 2, 183.

- Starkey, P. (1992). The early development of numerical reasoning. *Cognition* 43 (2), 93–126.
- Sternberg, R. J. (1982). Reasoning, problem-solving, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press, 225–307.
- Staszewski, J. J. (1988b). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71-128). Hillsdale, NJ: Erlbaum.
- Sternberg, R. J. (1996). The costs of expertise. In K. A. Ericsson (Ed.) *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 347–354). Mahwah, NJ: Erlbaum.
- Svanum, S. & Bigatti, S. M. (2006). The influences of course effort and outside activities on grades in a college course. *Journal of College Student Development*, 47(5), 564-576.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics*. 5th ed. Boston: Allyn and Bacon.
- Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology*, 56 (4), 369–379.
- Tuffiash (2002). Predicting individual differences in piano sight-reading skill: Practice, performance, and instruction. Unpublished dissertation.
- Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: Implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied*, 13(3), 124-134.
- Unsworth, N., Heitz, R. P., Schrock, J. C. & Engle, R.W. (2005). An automated version of the operation span task. *Behavior Research Methods* 37, 498–505

- Urena, C. A. (2004). Skill acquisition in ballet dancers: The relationship between deliberate practice and expertise. Unpublished dissertation. The Florida State University.
- van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology*, 58(2), 142-152.
- Walker, C., Fleischer, S. & Winn, S. (2008). A path analysis of first-year social science students' engagement with their degree and Level 1 academic outcome. *ELiSS*, 1(2), 1-19.
- Ward, P., Hodges, N. J., Starkes, J. L., & Williams, M. A. (2007). The road to excellence: Deliberate practice and the development of expertise. *High Ability Studies*, 18(2), 119-153.
- Waters, G. S. & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers* 35 (4), 550–564.
- Watkins, M. W., Lei, P-W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, 35, 59–68.
- Watson, J. B. (1930). *Behaviorism*. Chicago, IL: The University of Chicago Press.
- Wechsler, D. (1997). *Wechsler adult intelligence scale—WAIS III (3rd ed.)*. San Antonio, TX: Harcourt Assessment.
- Weissensteiner, J., Abernethy, B., Farrow, D., Müller, S. (2008). The development of anticipation: A cross-sectional examination of the practice experiences contributing to skill in cricket batting. *Journal Of Sport And Exercise Psychology*, 30(6), 663-684.
- Wicherts, J. M., Dolana, C. V., Hessena, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I. & Spanc, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509–537.

- Williams, B. A., & Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with Raven scores, but not with measures of cognitive processing. *Intelligence*, 34, 177-191.
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin* 76, 92–104.
- Wood, L. J. (1999). Practice, ability, and expertise in computer programming. Unpublished dissertation.
- Woody, R. H. (2003). Explaining expressive performance: Component cognitive skills in an aural modeling task. *Journal of Research in Music Education*, 51(1), 51-63.
- Young, B. W. (1998). Deliberate practice and the acquisition of expert performance in Canadian middle-distance running. Unpublished dissertation.
- Young, B. W., Medic, N., Weir, P. L., Starks, J. L. (2008). Explaining performance in elite middle-aged runners: Contributions from age and from ongoing and past training factors. *Journal of Sport and Exercise Psychology*, 30, 737-754.
- Yu, D. D. (2011). How much do study habits, skills, and attitudes affect student performance in introductory college accounting courses? *New Horizons in Education*, 59(3), 1-15.
- Yudes, C., Macizo, P., & Bajo, T. (2011). The influence of expertise in simultaneous interpreting on non-verbal executive processes. *Frontiers in Psychology*, 2, 309.
- Yudes, C., Macizo, P., & Bajo, T. (2012). Coordinating comprehension and production in simultaneous interpreters: Evidence from the articulatory suppression effect. *Bilingualism*, 15, 329–33.
- Zimmerman, B. J., & Kitsantas, A. (2005). Homework practices and academic achievement: The mediating role of self-efficacy and perceived responsibility beliefs. *Contemporary Educational Psychology*, 30, 397–417.