ECOLOGICAL COGNITIVE ASSESSMENT:
A NEW FRONTIER FOR THE CONCEPTUALISATION AND MEASUREMENT OF COGNITIVE ABILITY

Arabella Charlotte Vaughan
BSc. (Adv) (Hons. I and the University Medal)

School of Psychology
The University of Sydney
Sydney, New South Wales
AUSTRALIA

A thesis submitted to fulfil requirements for the degree of
Doctor of Philosophy

The research reported in this thesis was supported by the award of a Research Training
Program scholarship to the PhD Candidate

28 September 2024

STATEMENT OF ORIGINALITY AND APPROVAL

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

The data obtained through the research presented in this thesis was collected with approval by the Human Research Ethics Committee at the University of Sydney (Protocol 2021/463).

Arabella Charlotte Vaughan
PhD Candidate

AUTHORSHIP ATTRIBUTION STATEMENT

Chapter II of this thesis is partially published as: [Vaughan, A. C., & Birney, D. P. (2023). Within-Individual Variation in Cognitive Performance Is Not Noise: Why and How Cognitive Assessments Should Examine Within-Person Performance. *Journal of Intelligence, 11(6), 11060110*]. I conceptualised and wrote the drafts of the manuscript.

Some data from Chapters III – VI was partially presented as: [Vaughan, A. C., & Birney, D. P. (2022). Within-Person Variation in Cognitive Ability is Not Noise: Using the Experience Sampling Method (ESM) to Assess Cognitive Ability. Paper presented at the 20th conference of the Australasian Congress on Personality and Individual Differences, Churchill, Australia.] I designed and conducted the study, analysed the data, constructed the presentation, and wrote the manuscript version of the paper.

Some data from Chapters III – VI was partially presented as: [Vaughan, A. C., & Birney, D. P. (2023). Within-Individual Variation in Cognitive Performance is Not Noise: A Case for Examining Within-Person Variation on Cognitive Assessments. Paper presented at the 44th Annual Meeting of the Cognitive Science Society, Sydney, Australia.] I designed and conducted the study, analysed the data, constructed the poster, and wrote the manuscript version of the paper.

Chapters III – VI is partially published as: [Vaughan, A. C., & Birney, D. P. (2024). Ecological cognitive assessment has incremental validity for predicting academic performance over and above single occasion cognitive assessments. *Intelligence, 107, 101869*]. I conceptualised and wrote the drafts of the manuscript.

Some data from Chapter VII was partially presented as: [Vaughan, A. C., & Birney, D. P. (2024). Short-term within-person variation in cognitive performance is an adaptive process. Paper presented at the 21st European Conference on Personality, Berlin, Germany.] I designed and conducted the study, analysed the data, wrote and presented the conference presentation, and wrote the manuscript version of the paper.

Arabella Charlotte Vaughan, 28th September 2024.

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Damian P. Birney, 28th September 2024.

Acknowledgements

Firstly, I would like to express my deepest thanks to my supervisor Professor Damian Birney for his invaluable mentorship and guidance over the last five years of working together. My PhD experience has been enriching, challenging, and life-changing, and I would not have been able to succeed without your unwavering support. Thank you!

I would also like to thank my extended supervisory and lab team for their advice and feedback throughout my journey. Many thanks to Associate Professor Micah Goldwater for providing his support and expertise on early iterations of my thesis. I am also thankful to Dr Kit Double for his research methods and statistics advice. Lastly, thank you to Professor Carolyn MacCann and Dr Rebecca Pinkus for broadening my perspective on many of the issues covered in this thesis through their feedback and expertise.

I am also deeply grateful to Dr Sarah Walker for her friendship and encouragement throughout my PhD journey. Without you, my PhD experience would not have been the same. You’re an incredible researcher and friend, and it’s been a privilege to go on this journey with you.

When I started my PhD, something I didn’t anticipate was the sheer number of wonderful friendships I would forge along the way. I’ve been lucky to have found myself a community of the most incredible young female researchers, who all support and encourage each other to succeed. Massive thanks to Lilly, Yueting, Dora, Hester, Hannah, and Hayley for their friendship, collegiality, and enthusiasm for brunch. You’re all amazing women who are going to do amazing things!

This PhD would not have been possible without the support of my family. My deepest thanks go to my parents, Julia and Chris, and my brother, Sebastian. Thank you for all your love and support, and for encouraging me to take a leap into the unknown and go back to university to do a PhD. Mum and Dad, a special thank you for instilling in me a curiosity for learning and thirst for knowledge, and for supporting me over the last five years of my research journey with your love, encouragement, and food!

Finally, I’d like to thank my husband Tom for being the most wonderful partner, best friend, and occasional editor/sounding board/conference talk practice audience. Thank you for your advice, your encouragement, your pep talks, and for being my biggest fan.

Table of Contents

Abstract

Predominant theories of cognitive ability and methods of cognitive assessment assume that cognitive abilities differ between individuals but remain consistent within individuals, at least over short periods of time (e.g., weeks and months). As a result, cognitive ability is generally assessed at a single point in time in a static context. However, this approach fails to account for the empirical evidence that shows our performance on cognitive assessments varies over several weeks and months. In this thesis, I investigate whether this short-term within-person variation in cognitive performance is systematic and meaningful, or just noise (as predominant theories and methods would suggest). To do so, I use ecological assessment methods and analysis processes that allow a more comprehensive understanding of if, how, and why cognitive performance varies over time within the individual. Across three studies, I explore how the data generated from EMA can be parameterised to capture between-person differences in ecological within-individual variation in cognitive performance across time and contexts. I show that many of these parameters incrementally predict university performance over and above performance on a traditional single session cognitive assessment and that this relationship appears to emerge because short-term within-person variation in cognitive performance is a substantive construct above and beyond cognitive ability as we traditionally conceptualise and assess it. Moreover, I demonstrate that this short-term within-person variation is an adaptive process that facilitates resilience to changes in internal states, which enables individuals to perform well under varied situational cues. The findings of this thesis suggests that short-term within-person variation in cognitive performance is systematic and meaningful, and should be incorporated into our theories of cognitive ability and methods of cognitive assessment.

ECOLOGICAL COGNITIVE ASSESSMENT:
A NEW FRONTIER FOR THE CONCEPTUALISATION AND MEASUREMENT OF COGNITIVE ABILITY

I. INTRODUCTION

The human experience is richly dynamic, characterised by constant change and complexity. Each day, we encounter a variety of situations, people, and tasks that require us to adapt and respond with flexibility. Intuitively, most people would recognise that their thoughts, feelings, and behaviours are heavily influenced by these environmental changes and that their emotional and cognitive responses are variable as a result. Despite this intuitive understanding, psychological theories and methods have historically tended to conceptualise cognitive and individual differences as static traits that do not vary much within an individual over time, at least in the short term.

In recent decades, however, there has been a shift towards using ecological assessment to better understand how and why individuals vary in their responses to dynamic environmental changes. Ecological assessment involves studying individuals in their natural environments, capturing data on how they react and adapt to real-world situations (Csikszentmihalyi & Larson, 2014). This approach has provided valuable insights, particularly in personality research, where it has demonstrated that personality traits, once considered static unchanging traits, are better understood as a density distribution of states more or less likely to be expressed depending on the context an individual is in (Fleeson, 2001, 2007; Fleeson & Jayawickreme, 2015). In contrast, cognitive ability research has been slower to adopt ecological assessment methods and has historically preferred single occasion, context invariant assessments.

Cognitive assessment is used extensively in education, industry, and clinical settings. In New South Wales alone, over 18,500 children sat the selective school test in 2024 (Carroll, 2024) while 400,000 children sat the NAPLAN tests in the same year (Education, 2024). In industry, cognitive ability tests remain one of the best predictors of workplace performance (Schmidt et al., 2016) and estimates suggest they are used by around 40% of Australian companies for graduate recruitment (Bradley et al., 2021). In practice, these cognitive assessments are typically conducted as single-occasion, point-in-time evaluations. During these assessments, individuals are tested and a sum score of correct responses is derived from their performance. These sum scores are then used to make comparisons between individuals, aiming to understand their relative cognitive abilities. For example, in educational settings, such scores might be used to place students in gifted and talented programs or to identify those who need additional support. In industrial settings, these scores could influence hiring decisions and professional development opportunities.

The current approach to cognitive assessment is based on the widely accepted notion that cognitive abilities differ between individuals but remain consistent within individuals, at least over short periods of time (e.g., weeks and months). Therefore, it is believed that comparisons of single-occasion sum scores of performance will provide a stable and reliable indication of an individual's cognitive ability at present and in the near future. For instance, if a student scores high on a cognitive test, it is assumed that this score reliably reflects their cognitive capacity, which will remain consistent in subsequent evaluations. However, this approach fails to account for the large body of empirical evidence that shows that our performance on cognitive ability assessments does vary over several weeks and months (Allaire & Marsiske, 2005; Bunce et al., 2004; Li et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009; Salthouse, 2007, 2012; Salthouse & Berish, 2005; Salthouse et al., 2006; Schmiedek et al., 2013). By not considering this variation, the current method of cognitive assessment may provide an incomplete and potentially misleading picture of an individual's cognitive abilities.

The implications of neglecting ecological within-person variation in cognitive assessments are significant. If our theories and methods are incorrect due to the exclusion of within-person variability, we may have an incomplete conceptualisation of cognitive ability. This could lead to assessment methods that do not provide a holistic perspective on an individual’s cognitive strengths and weaknesses. As a result, we may fail to identify the right people for the right opportunities, such as job selection, educational program selection, and learning support.

Using ecological assessment methods and analysis processes could allow us to more comprehensively understand if, how, and why cognitive performance varies over time within the individual. In turn, this would allows us to build more dynamic and holistic theories of cognitive ability, as well as develop cognitive assessments that truly reflect an individual’s capabilities, leading to more informed and equitable decisions in education, industry, and clinical settings.

1. Structure and Aims

Research Question 1: Why is understanding within-individual variability in cognitive performance important for our theories of cognitive ability?

My first research question is addressed in Chapter 2, which presents the case for examining within-individual variation in cognitive performance. In this chapter, I outline the large body of empirical evidence that shows that, contrary to predominant theories and methods, there is substantial within-individual variability in cognitive performance. I argue that cognitive ability as it is traditionally conceived of and operationalised is a point in time assessment that does not account for ecological variation in the application of cognitive ability across time and contexts, and is therefore fundamentally asymmetrical with the criterion outcomes it is commonly used to predict. These outcomes, chiefly academic and workplace performance, demand adaptability, flexibility, and ongoing engagement with cognitive challenges. I highlight the utility of a within-individual assessment and analysis approach in related fields, particularly the personality literature, where ecological assessment techniques and analysis methods have been used to understand within-individual variation in personality states over time and contexts. Finally, I outline how these ecological assessment techniques and analysis methods could be adapted for cognitive assessment.

Research Question 2: How can we assess and analyse within-individual variability in cognitive performance?

My second research question is addressed in Chapter 3, which explores how the data generated from ecological assessment techniques can be used to understand ecological within-individual variability in cognitive performance. I explore and propose three classes of quantitative parameters that leverage short-term repeated measures data generated from an EMA approach to model the extent, magnitude, and importance of ecological within-individual variation in cognitive performance. I refer to these parameters collectively as "cognitive dynamics" to highlight they are indicators of performance derived across different repeated measures and settings more so than is typically the case.

Research Question 3: Does this assessment and analysis approach lead to better prediction of academic performance than cognitive ability as traditionally conceptualised and assessed?

My third research question is addressed in Chapters 4, 5, and 6. In Chapter 4, I apply the cognitive dynamics identified in Chapter 3 in an EMA study using a standard reasoning task administered repeatedly in idiosyncratic situations over two weeks. I compare the predictive utility of the cognitive dynamics with a measure of cognitive ability as traditionally conceptualised and assessed: a single occasion, context-invariant assessment where performance is measured as a sum score of correct responses. In Chapter 5, I apply the cognitive dynamics in a second EMA study using a microworld task requiring dynamic decision-making administered repeatedly over three weeks, also in idiosyncratic situations. Across Chapters 4 and 5 I also briefly discuss engagement patterns in both empirical studies to ensure that the EMA is feasible and acceptable to participants. In Chapter 6, I synthesise the findings of the two empirical studies. I discuss that many of these cognitive dynamics incrementally predict university performance over and above performance on a traditional single session cognitive assessment, and propose that this is because these parameters capture a substantive construct beyond cognitive ability as it is traditionally assessed.

Research Question 4: Why might this assessment and analysis approach lead to better prediction of academic performance than cognitive ability as traditionally conceptualised and assessed?

My final research question is addressed in Chapter 7. In this chapter, I begin building an explanatory process account for why ecological within-individual variability in cognitive performance predicts academic performance than cognitive ability as traditionally conceptualised and assessed. To do this, I use an EMA approach and administer a measure of situational features at each EMA measurement occasion. I then derive within-person situation contingencies (i.e., situation-performance relationships) and show that these predict university performance over and above cognitive ability as traditionally conceptualised and personality and motivational traits associated with cognitive engagement. In doing so, we demonstrate that short-term within-person variation in cognitive performance is an adaptive process that facilitates resilience to changes in internal states, which enables individuals to perform well under varied situational cues. We suggest that cognitive ability is a meaningful predictor of university performance because it can be used in this flexible and adaptive manner, and it is this flexibility and adaptivity that determines success in cognitively demanding real world scenarios rather than cognitive ability as it is traditionally conceptualised and operationalised.

2. Summary of Findings

Ecological cognitive assessment can improve the theoretical and predictive utility of cognitive assessment by better aligning assessment and analysis approaches with corresponding criterion outcomes. Relatively simple ‘cognitive dynamics’ parameters derived from EMA data consistently predicted academic performance above and beyond traditional single session, context-invariant cognitive assessments. This predictive relationship appeared to emerge because EMA-derived cognitive dynamics accounted for something broader than, or different to, what traditional assessments measure. Moreover, we found that individuals whose performance varied relatively less in response to changes in situational cues were better adapted to succeed academically. We propose that these findings are due to the increased symmetry between predictors and criterion when using ecological cognitive assessment. Criterion outcomes, such as university or workplace performance, are an accumulation of performance instances across time and contexts. When cognitive ability is also assessed as an accumulation of performance instances across time and contexts (i.e., it is assessed using ecological cognitive assessment), its predictive utility is greater because it better reflects the way in which we actually use our cognitive ability in the real world.

II. THE CASE FOR EXAMINING WITHIN-INDIVIDUAL VARIATION IN COGNITIVE PERFORMANCE

1. Introduction

Our world is increasingly complex and dynamic. As a society, we face ever more challenging problems which have a multitude of ethical considerations and resource constraints. In our attempts to meet these challenges as individuals, we are often barraged with information from multiple sources of varying reliability. Yet, despite these progressively more complex and dynamic cognitive demands placed on many facets of our lives, our view of human cognitive abilities remains decidedly static. Attempts have been made to broaden our understanding of intelligence, for example through dynamic testing and complex problem-solving movements, but it is apparent that progress has not been promising enough to substantially shift the status quo. We argue that what is missing in intelligence research is an understanding of the role of within-individual variability in cognitive performance (Birney & Beckmann, 2022).

Variability in cognitive performance has long been recognised as important for understanding the full range of human cognitive abilities (Fiske & Rice, 1955). For instance, Spearman (1927) postulated the existence of an oscillation factor that influences the efficiency and accuracy of cognitive performance. Yet it remains common practice to construct intelligence tests in ways that ensure stability of within-individual differences in the construct they purport to measure. That is, intelligence is assumed to differ between individuals but remain constant within individuals. From this perspective, within-individual variability is an indicator of poor reliability and low validity that should be eliminated during test construction (Birney et al., 2019; Cronbach, 1957). Psychometric tests often include practice items, sometimes with feedback, to provide experience with the style of items to be used. This serves to minimises errors in initial items due to construct-extraneous factors that a lack of familiarity with task requirements might introduce. In doing so, this approach further acts to eliminate within-individual differences in cognitive performance and confounds their analysis. Moreover, standard practice ignores a substantial body of research that shows people’s performance on basic cognitive tasks varies even within a short time period (Allaire & Marsiske, 2005; Bunce et al., 2004; Li et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009; Salthouse, 2007, 2012; Salthouse & Berish, 2005; Salthouse et al., 2006; Schmiedek et al., 2013).

Our exposition of the nature and implication of within-individual variation in cognitive ability is presented in three parts. In Part 1, we build a case for why understanding within-individual variability in cognitive performance is important for enhancing our theories of human intelligence and outline a proposed approach for doing so. To date, within-individual variability in cognitive performance has predominantly been conceptualised as noise (Birney et al., 2019; Voelkle et al., 2014) or an indicator of cognitive problems (Bunce et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009). We propose that within-individual variability in cognitive performance can be construed as a meaningful property of an individual's intelligence, and that this is necessary for the development of a process theory of cognitive ability that can provide an explanatory account of why individuals with similar cognitive test results might perform differently in the real world. We argue that the predominantly between-individual approach to the measurement of cognitive ability obscures important within-individual variation in cognitive performance that may have significant implications for the use of cognitive tests (Birney & Beckmann, 2022).

In Part 2, we outline methods for examining within-individual variability in cognitive performance. In doing so, we explicate our contention that established paradigms for understanding the magnitude and implications of within-individual variability used in related fields can be adapted to understand within-individual variability in cognitive performance. Specifically, in personality research, within-individual variability has been studied using short-term repeated-measures paradigms, such as the Experience Sampling Method (ESM), which have enriched the field's understanding of within-person variation as a property of personality that should be assessed independently (Beckmann et al., 2020; Fleeson, 2001, 2007).

In presenting this case, we hope to encourage researchers to look beyond the traditional approach to assessing cognitive ability. We argue that embracing a new approach has the potential to enhance both our theoretical understanding of the dynamic nature of human intelligence and our capacity to optimise practical applications of cognitive tests. At the very least, we argue that the extent to which within-individual variation has systematic between-individual differences should be considered more fully.

2. Part 1

In the following sections, we present the case for examining within-individual variability in cognitive performance. First, we define within-individual variability in a broad framework of performance variability. Second, we outline the empirical evidence that suggests within-individual variability is a meaningful property of cognitive ability, rather than noise. Finally, we describe why understanding within-individual variability is essential for the interpretation and use of cognitive test results in practical contexts.

2.1. What is Within-Individual Variability?

We begin by situating within-individual variability in a broader framework for studying cognitive performance variability. Hultsch et al. (2008) distinguish between three related forms of cognitive performance variability, illustrated in Figure 2.1. Firstly, between-individual differences are the differences between people when measured on a single construct on a single occasion, for example, differences between people’s performance on a general mental ability test. This is the predominant form of variability compared in assessment and selection scenarios. Secondly, within-individual differences are the differences within a person measured on multiple constructs on a single occasion, for example, differences between an individual’s performance on different domains of a general mental ability test, such as Gf, Gc, Gv, etc. Finally, intra-individual variability is the differences within a person measured on a single construct across multiple occasions, for example, differences in an individual’s performance on a Gf test across time.

Nesselroade (1991) distinguishes between two temporal periods of intra-individual variability. Firstly, within-individual change is a form of intra-individual variability whereby an individual’s performance can change slowly and (sometimes) permanently over extended temporal periods, such as years or decades. In Figure 2.1, within-individual change would be demonstrated if the points on the x-axis represented performance over a long temporal period (e.g., years). Secondly, within-individual variability represents fluctuations in an individual’s performance over shorter temporal periods, such as days or weeks. Understanding this variability is closely related to the ESM approach that we will describe shortly and would be represented in Figure 2.1 if the points on the x-axis represented shorter intervals (e.g., days). Of the two, within-individual change has received far more empirical attention through a plethora of longitudinal and cross-sectional research investigating cognitive changes throughout the lifespan, while the attention devoted to shorter-term within-individual variability has been comparatively sparse.

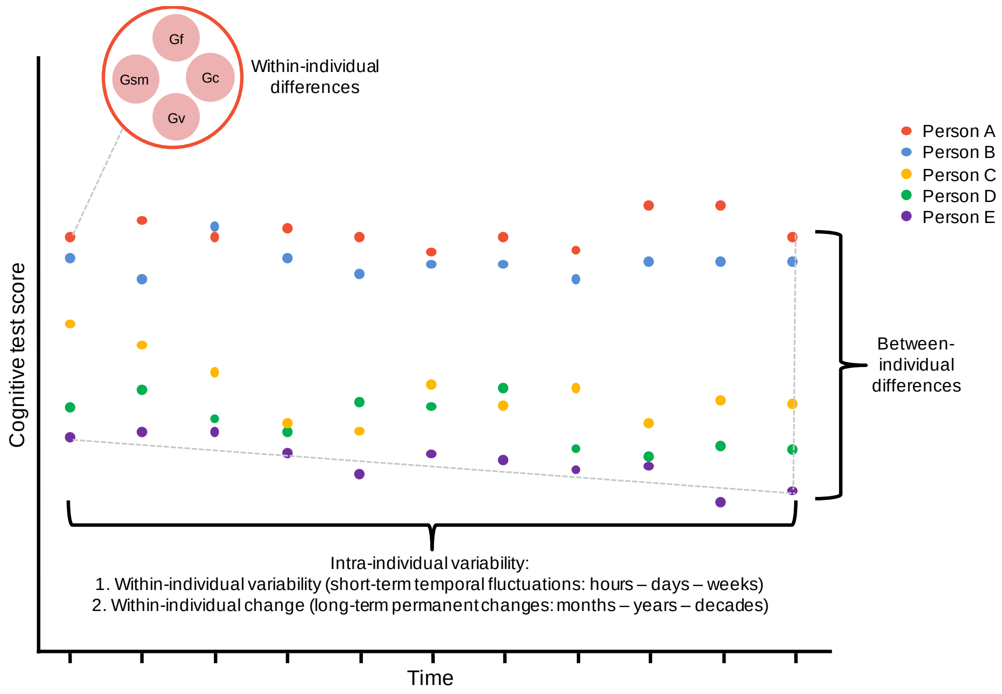


Figure 2.1. Types of cognitive performance variability outlined in Chapter 2, Section 2.1.

Within-individual variability further encompasses both random and structured variability components. Fiske and Rice (1955) distinguish between Type 1 variability, which is spontaneous or random variability, and Type 2 variability, which is reactive or structured variability. When considering within-person variation in cognitive performance, it is structured variability that we are interested in studying as doing so may help tells us how and why people systematically vary in their typical intellect-related behaviour. This structured variability can occur both across and within measurement occasions (as detailed in Chapter 2, Section 3.2.1). Cognitive performance variation is likely to be structured in regard to a variety of factors; in this article we focus on changes in situation, task, and person-related factors (as detailed in later sections in this chapter).

2.2. What is the Evidence for Within-Individual Variability?

We now turn to the question of whether there is evidence for within-individual variability in cognitive performance. Empirical evidence suggests that people demonstrate both within-individual change and within-individual variability on cognitive tests. Firstly, within-individual change is evidenced when both the scores and factor structure of cognitive abilities are temporally unstable (McArdle et al., 2002; Tucker-Drob, 2009). Different cognitive abilities show consistent age-related changes; fluid intelligence (Gf) and speed-related abilities show initial increases and then pronounced declines throughout the lifespan, whereas crystallised intelligence (Gc) and auditory/visual processing plateau and decline at a far shallower rate (McArdle et al., 2002). Furthermore, the first factor extracted from large cognitive test batteries, the *g* factor, accounts for more or less variance in cognitive performance depending on both age and baseline cognitive ability (Tucker-Drob, 2009). It too (i.e., *g*) is also considered to undergo qualitative change developmentally over time (Demetriou et al., 2023).

There is also substantial empirical evidence that suggests within-individual variability is an important aspect of cognitive performance, although most of the extant research has focused on basic cognitive processes such as reaction time (see e.g., Bunce et al., 2004; Hultsch et al., 2008; Li et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009; Salthouse & Berish, 2005). Salthouse and colleagues have demonstrated that within-individual variability in cognitive performance is large on various tasks. For reaction time measures it is roughly of the same proportion as between-person variability (see Salthouse & Berish, 2005 who conduct an ESM-style study similar to what we will do in Part 3). On more complex cognitive tasks (e.g., vocabulary, inductive reasoning, spatial visualisation, episodic memory, and perceptual speed) the median ratio of within-person variability to between-person variability was 0.54 (ranging from 0.28 to 0.78) (Salthouse, 2007). Allaire and Marsiske (2005) found that across a 60-day testing period people displayed substantial within-individual variability in performance and that variability in one domain was generally unrelated to variability in another domain. Moreover, higher performance variability was associated with higher mean performance in a given domain (r = .34 for perceptual speed and r = .43 for working memory). Similarly, Salthouse et al. (2006) found substantial within-individual variability in cognitive test performance that varied in magnitude between individuals across a two to 10 week period (the median ratio of within to between-person variation was 0.46). In a follow-up study, Salthouse (2012) reported that individuals exhibited between-session variability equivalent to one or more decades of cross-sectional age-related performance differences (with a mean between-session variability of 8.8 years of age-related performance differences), suggesting people vary substantially in their performance in relatively short temporal frames.

In sum, the extant literature suggests that despite the best efforts of cognitive test developers to weed out within-individual sources of variation in test development, people do vary in their performance on cognitive assessments across relatively short time periods. That is, there is indeed evidence for systematic within-individual variability in cognitive performance. Further, this variation cannot be attributed to cognitive decline or maladaptive cognitive functioning because it has been replicated in younger populations without clinically significant cognitive functioning impairments (e.g., in Salthouse, 2007; Salthouse et al., 2006). However, the magnitude and nature of this within-individual variability has not been well integrated into our theories of human intelligence. In fact, outside of motivation as an explanatory account, little is known of the sources of within-individual variability in higher-level cognitive abilities.

2.3. Why is Within-Individual Variability Important?

Without integrating an understanding of within-individual variability into our theories of human intelligence, there is and will continue to be a disconnect between the predominant theory and methodology for operationalising and measuring cognitive ability and the actual nature of cognitive abilities in the real world. There are a number of reasons for this. First, consider a typical cognitive assessment situation. A major national bank is recruiting for their graduate program. As a desirable workplace, they have thousands of applicants to assess and rank, most of whom have impressive academic records. The bank invites several hundred candidates to complete an online cognitive assessment. On the basis of these test results, as well as interviews and reference checks, the bank selects the highest performing students for its graduate program. Yet, several months into the graduate program, it is apparent that the actual workplace performance of the graduates varies substantially – both between and within the individual graduates. All came in with stellar records and all achieved well on the cognitive assessment, so why?

At present, theoretical and practical research into intelligence predominantly relies on such between-individual comparisons of single-occasion scores, or rather it is often framed in that way (that is, to compare individuals and to select one person over another). Between-individual comparisons of test scores obtained at a single point in time in a controlled environment cannot well-capture those aspects of cognitive ability that vary within individuals. The test situation facilitates maximum concentration and cognitive engagement, either through the test being taken in a testing centre or by the test taker selecting an optimal physical location to take the test. The content and design of the task also demands a narrow form of cognition; cognitive assessments are generally static tasks comprised of a series of independent items, with no elements that change dynamically or based on previous test taker responses[1]. Mostly, only multiple-choice response options with a binary correct/incorrect outcome are available. Further, motivation is usually at its peak because something valued is at stake (e.g., a job or a spot on an educational/training program), so the situational contexts are often similarly narrow.

This approach is extremely different to the actual cognitively challenging situations we encounter on a day-to-day basis. Moreover, it ignores the empirical evidence that individuals can exhibit very different patterns of within-individual variability in cognitive performance across relatively short time periods. Successful real-world cognitive performance requires dealing with dynamic problems, which necessitates the capacity to be sensitive to changes in the problem as well as to changes in one's goals over time, flexibility to adapt to these changes, and the capacity to integrate new information into existing schemas. Even with recognition that individual differences in motivation and engagement will fluctuate, dynamic changes in our environment demand a cognitive capacity unlikely to be captured sufficiently in static intelligence tests which have been designed to be sensitive to between-person differences. While some have proposed different constructs, such as practical intelligence (e.g., Sternberg et al., 2000), to deal with contextual variability, our focus here remains on the set of classic cognitive abilities. However, we frame these classic cognitive abilities from a within-person perspective described as short-term temporal fluctuations in Figure 2.1, rather than as only between-individual differences.

Understanding real-world cognitive performance therefore requires an understanding of structured within-individual variability in response to changes, in other words, we need to embrace heterogeneity (Bryan et al., 2021). This is not to the exclusion of between-individual comparisons. Rather, it is intended to complement and enhance the utility of between-individual comparisons by providing a within-person process account of between-person differences in cognitive ability, much in the same way as nuanced personality *states* have augmented our understanding of personality *traits*. As an example, consider two individuals who achieve similar results on a cognitive assessment when a place on a competitive graduate program is at stake, as represented by the red X in Figure 2.2. Despite their similar peak score, the individuals exhibit different patterns of within-individual variability in other situations as represented by different distribution curves. This between-person variability in within-person variability (distributions) may be structured in regard to situation, task, or

---

[1] By change based on past responses, we do not mean adaptive psychometric testing, rather we mean a more qualitative change in item type.

personal factors (e.g., one individual may learn skills and concepts faster than the other, or be more sensitive to changes in the problem space and thus quicker to incorporate these into their problem-solving strategy).

While Figure 2.2 is purely illustrative, the intention is to draw attention to the fact that single-occasion tests taken in conditions designed to minimise within-person variability do not provide a holistic understanding of how or why someone is likely to dynamically vary in their performance across days, weeks, months, or even several years. If we are selecting for a role where consistent high performance is important and mistakes are costly, such as a surgeon or pilot, most people would prefer one whose performance is of consistently high quality and less prone to variation, rather than one whose performance varies substantially. It is for such reasons we suggest that understanding within-individual cognitive performance variability is crucial to holistically understand the cognitive abilities underlying human intelligence, design appropriate cognitive tests, and to interpret results appropriately to make an optimal decision.

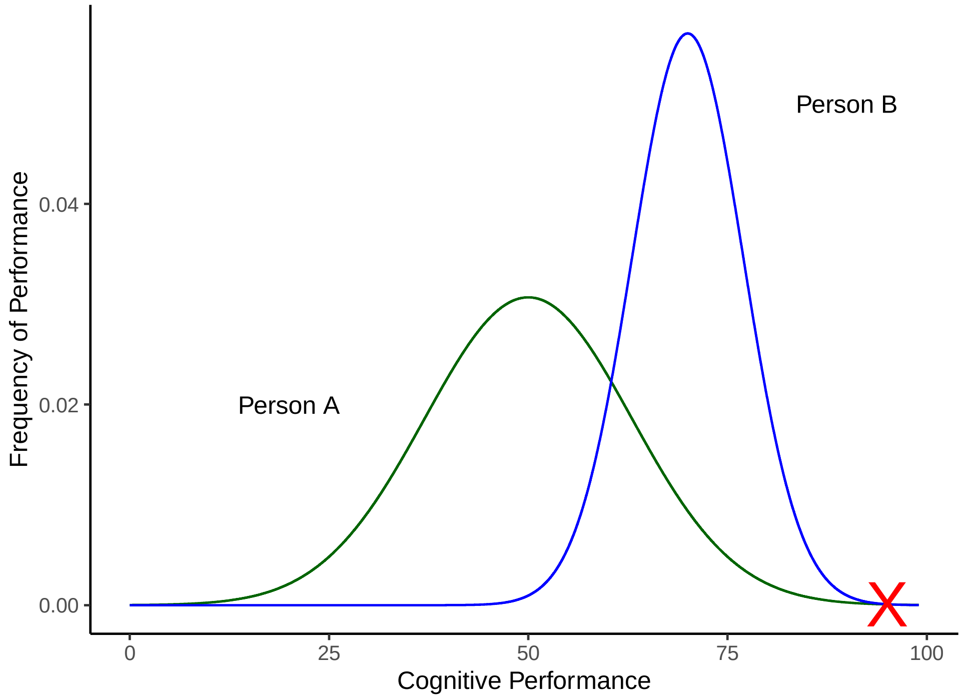


Figure 2.2. Individuals with the same maximal performance and different typical performance.

3. Part 2: Methods for Examining Within-Individual Variability in Cognitive Performance

In Part 2 we review how related fields have incorporated an understanding of within-individual variability into theories and measures, particularly Experience Sampling Method (ESM), and discuss implications for cognitive assessment (Chapter 2, Section 3.1). We then outline the considerations for researchers when adapting the ESM paradigm to capture short-term within-individual variability in assessed cognitive ability. We focus on theoretical (Chapter 2, Section 3.2), methodological (Chapter 2, Section 3.3), and analytical considerations (Chapter 2, Section 3.4). In doing so, we compare existing methods for examining within-individual variability in personality with our proposed approach for examining the same in cognitive ability.

3.1. How do Related Fields Capture Within-Individual Variability?

Within-individual variability has been well integrated into theories and methods for assessing a range of psychological constructs, notably affect, emotion, and personality (Beckmann et al., 2020; Conner et al., 2009; Ilies & Judge, 2002; Trull et al., 2015). Importantly, all of these fields have gleaned an understanding of the role of within-individual variability by pivoting away from the predominant use of single-occasion between-individual assessments and comparisons. In particular, within-individual repeated-measures research designs have allowed researchers to uncover how and why individuals vary in their experiences across a typical week or month (Beckmann et al., 2020; Conner et al., 2009; Ilies & Judge, 2002; Trull et al., 2015). Generally, this research leverages the Experience Sampling Method (ESM)[2]. The ESM takes many short measures a week, fortnight, or month, and is based on a simple premise – taking a sample of people's experiences over a short time period allows us to more deeply understand the variability in their thoughts, feelings, behaviours, and experiences as they unfold in their everyday lives (Conner et al., 2009). The spreading of measurement occasions across several weeks allows researchers to sample fluctuating experiences across a range of contexts and to understand how variability is related to various outcomes (Conner et al., 2009; Fisher & To, 2012).

In personality research, the ESM has allowed researchers to study within-individual variation in personality, which has resoundingly challenged many of the central tenets of nomothetic personality theory. Importantly, these findings have complemented between-

---

[2] ESM is also known as ecological momentary assessment (EMA) and ambulatory assessment.

individual comparisons by providing a richer process account of how and why individuals differ in their personality expression over time. This idiographic approach has notably been explored by Fleeson and colleagues (Fleeson, 2001, 2007; Fleeson & Jayawickreme, 2015), who demonstrated distinctive within-individual personality profiles using the Big Five personality traits. Across multiple studies, Fleeson asked participants to complete a measure of the Big Five several times a day for several weeks in a row. Fleeson found that people exhibited relatively low consistency in self-reported personality states. Rather, there was more variability within individuals than between individuals. As a result of these findings, he proposed that personality is better conceptualised as a density distribution of states that is unique to an individual, with each state more or less likely to be expressed depending on situational cues (e.g., if the situation is a novel social environment, then personality state is more/less extraverted). More recent research has sought to understand the factors that systematically influence within-person variability in state personality expression and has uncovered a range of situation and person factors implicated in the dynamic nature of personality (Beckmann et al., 2020; Beckmann et al., 2021; Wood et al., 2019).

We propose that the ESM and accompanying within-individual analysis techniques be used to integrate an understanding of within-individual variability in cognitive *performance* into our theories of and methods for assessing cognitive *ability* (and intelligence generally). It is necessary to obtain multiple measures of individuals' cognitive ability over short time periods under structured conditions so as to better understand how and why they vary in their cognitive performance. We do not propose such a paradigm be used to the exclusion of traditional, well-validated methods for assessment and analysis. Instead, we believe that the findings in personality research are indicative of potential to also enhance our theoretical understanding of the extent and magnitude of within-individual cognitive performance variability, which can be integrated into existing theories of intelligence and methods for cognitive assessment.

3.2. Theoretical Considerations

The theoretical rationale for studying within-person variability in cognitive ability will inform the way in which short-term repeated measures paradigms are designed. In this section we compare how within-individual variability in personality is conceptualised relative to how within-individual variability in cognition is conceptualised, as well as how personality and cognitive states and traits are similar and different. We then consider the implications of this for specifying the structure of within-individual variation and task choice.

3.2.1. Conceptualisation of Variability

In Chapter 2 Section 2.1, we situated within-individual variability in a framework of cognitive performance. We outlined two forms of within-individual variability described by Fiske and Rice (1955)—Type 1 (spontaneous or random variability) and Type 2 (structured variability)—and emphasised that it is the latter that we are interested in understanding in this line of research. We now elaborate on why structured variability is of interest when studying personality and cognitive ability.

For personality, variability is construed as a systematic and structured response to situational cues (Fleeson, 2001, 2007). Personality state expression is primarily determined by an individual’s distinct pattern of if-then contingencies, e.g., if the social context is familiar then act more extraverted but if the social context is unfamiliar then act less extraverted (Fleeson, 2001, 2007). Empirical evidence would suggest that the same is likely to be true of cognitive performance; within-individual variability may be a systematic response to changes in situation, task, and person factors. Cognitive performance can be construed as having two layers of structured variability, between-occasion variability and within-occasion variability.

Between-occasion variability entails structured differences in response to situation, person, and task factors that change across measurement occasions. The role of the situation in which a task is completed has been explored extensively in the theory of maximal and typical performance. In high stakes situations with high perceived importance, people are expected to aspire for maximal performance (Goff & Ackerman, 1992; von Stumm et al., 2011). Variability is primarily dictated by one’s cognitive ability because non-cognitive factors such as motivation and engagement are mostly constant across individuals, thus performance is at its peak. As the stakes and perceived importance of the situation decrease, people display more typical performance (Goff & Ackerman, 1992; von Stumm et al., 2011). Typical performance is dictated by variation in both cognitive ability and non-cognitive factors, and between-occasion differences in situation valence are likely to interact with task demands. More discretionary effort and motivation is required to voluntarily engage greater cognitive resources in lower-stakes activities because the expectancy of a valued outcome is relatively lower (Wigfield, 1994). Furthermore, person factors including motivational and conative dispositions such as Conscientiousness, Openness/Intellect, Need for Cognition, and Goal Orientation have been shown to incrementally predict typical performance beyond cognitive ability (Schmidt & Hunter, 1998; Schmidt et al., 2016; Steinmayr et al., 2011;

Strobel et al., 2019). Finally, differences in task demands and design are likely to contribute to between-occasion variability as the tasks presented to us change across measurement occasions.

Within-occasion variability primarily entails structured differences in task factors that change while one attempts the task. Changes in the impact of situation and person factors are a potential additional source of within-occasion performance variability. For example, in both cognitive tests and real-world cognitive challenges there may be changes in the content and design of the task and the framing of the situation while the task is underway that affects ongoing task performance. Within-occasion variability may also be impacted by some of the aforementioned person factors, which affect engagement with cognitive challenges and responses to changes in the task or situation. Analysing within-occasion variability is important because it reveals information about people's response to change, their strategy use, cognitive exploration, and flexibility (Schmiedek et al., 2010). These critical aspects of cognitive ability cannot be revealed by between-occasion analysis alone. Variability in cognitive performance represents something both similar and distinct from variability observed in personality research which necessitates the analysis of different layers of structured within-individual variability.

3.2.2. Conceptualisation of States and Traits

Theoretically there are also similarities and differences in the conceptualisation of states and traits for personality and cognition that are important to consider when adapting the ESM paradigm to assess within-individual variation in cognitive performance. In personality research, personality traits are assumed to reflect an average level of behaviour across time (Fleeson & Jayawickreme, 2015). This is quantified based on an individual's self-reflection at a single time point on their usual behaviour or, in the case of repeated-measures studies, based on the average of assessments across time. Conversely in cognitive ability research, the ability traits are generally thought to reflect the maximum level of behaviour, which is quantified as a score (usually) observed in a high stakes situation at a single time point (von Stumm et al., 2011). Consequently, trait personality and "trait" cognitive ability are theoretically different; trait personality reflects typical behaviour whereas "trait" cognitive ability reflects maximum behaviour.

Both personality and cognitive states share similarities in their conceptualisations. The expression of personality and cognitive states is based on a conscious or subconscious judgment of the environment and a conscious or subconscious decision to allocate resources to navigating the situation (Fleeson, 2007; Wigfield, 1994). When these states are measured on multiple occasions in personality research, a distribution of within-person variation in states is observed (Fleeson & Jayawickreme, 2015). This distribution shows both the trait and its nature of within-person variation (e.g., range, shape, skew) and can be used to understand the factors that influence the expression of a particular state (Fleeson & Jayawickreme, 2015). Moreover, the degree of adaptability in personality or cognitive state expression is likely to be linked to different outcomes (Ram & Gerstorf, 2009), such as in the workplace or education. In other words, both the average trait level and the shape of one’s personality and cognitive performance distribution are likely to be reflected in successful life outcomes, albeit in different ways.

However, the conceptualisations of personality and cognitive states differ in an important way. Personality states are captured in a relatively simple momentary reflection on one’s current feelings. This is because personality states are assumed to unfold organically and subconsciously; we do not have to exert substantial effort to answer personality survey items because our personality state is generally not something we consciously influence. On the other hand, our theoretical conceptualisation of successful cognitive performance in demanding environments is of one’s capacity to consciously engage and respond to discrete cognitive challenges, and to changes in these challenge once we are engaging with it (e.g., at school, university, or the workplace) (Birney & Beckmann, 2022). To respond to these cognitive challenges, one must effortfully disengage from other tasks, reflect on the challenge presented, and deliberately and effortfully allocate resources to complete it. Faithfully capturing a cognitive state, whether using an ESM paradigm or not, requires longer and more engagement-demanding measures than capturing personality states, and this has implications for the type of task used.

3.2.3. Structure of Variation

The theoretical conceptualisation of variability, states, and traits has implications for the likely structure of cognitive ability variation and this has ramifications for the adaptation of ESM paradigms to cognitive assessment. At its core, expression of momentary personality and cognitive ability states involves the allocation of our finite cognitive, behavioural, and emotional resources based on some appraisal of the situation. This allocation of resources represents a structured and systematic response to situation, task, and person factors at the time the measurement is taken. In personality research, situation factors have increasingly been shown to systematically influence within-individual variation, for instance by the degree of familiarity of those we interact with, whether it is a social or workplace context, and the particular reason for the social interaction (Beckmann et al., 2020; Rauthmann et al., 2016; Rauthmann et al., 2015; Sherman et al., 2015).

However, as we have outlined in Chapter 2 Section 3.2.1, cognitive performance is structured between tasks and within tasks. Moreover, as described in Chapter 2 Section 3.2.2, the complex nature of cognitive task performance means that it is likely to be interactively structured in respect to situation, person, and task factors. Therefore, when developing an ESM paradigm to assess structured within-individual variability in cognitive performance, we will likely need to capture more sources of structured variability than in personality research. As a starting point for further research, it is instructive to briefly outline some situation, person, and task factors that may systematically influence cognitive task performance.

Situation factors that influence performance may include the stakes of the situation in which a task is completed (Goff & Ackerman, 1992; von Stumm et al., 2011). In personality and affective dynamics research, measures have been established to evaluate situation-contingent responding. One popular measure is the DIAMONDS scale (Rauthmann et al., 2014; Rauthmann et al., 2016; Rauthmann et al., 2015), which captures eight situational features that may influence behaviour: Duty, Intellect, Adversity, Mating, Positivity, Negativity, Deception, and Sociality. The empirical evidence for the interaction between DIAMONDS and personality is limited (Rauthmann et al., 2016; Sherman et al., 2015), however, this may be due to the similarity of content captured in DIAMONDS and personality measures, e.g., Sociality bears strong similarities to Agreeableness and Extraversion, Intellect bears strong similarities to Openness/Intellect. For the study of cognitive ability, the DIAMONDS or something similar may be useful to understand situational influences. As there is little, if any, construct overlap between these situation characteristics with cognitive ability, DIAMONDS ratings may explain unaccounted for variance in cognitive states. Equally however, the factors may simply be too dissimilar to those that influence cognitive performance. In either case, there is research to be conducted to better outline the situational features that systematically influence within-individual variation in cognitive performance. In fact, we begin this work in Chapter 7 of this thesis.

Task factors include task complexity and design[3], the cognitive load of the task (including the presence of intrinsic, extraneous, and germane cognitive load) (Beckmann,

---

[3] Task design is related, in part, to the particular ability being assessed. For instance, Gf, Gc, and Working-memory should each entail task designs that capture their respective within-person processes.

2010), and the introduction of new features or information into the problem space (Goff & Ackerman, 1992). In addition to the specific cognitive ability being assessed, *person factors* include aforementioned individual differences traits like Conscientiousness, Openness/Intellect, Need for Cognition, and Goal Orientation, which predict typical cognitive performance (Schmidt & Hunter, 1998; Schmidt et al., 2016; Steinmayr et al., 2011; Strobel et al., 2019). Other more transient person factors include skill acquisition, neuromodulatory processes, stress, and motivation (Schmiedek et al., 2010).

Importantly, these situation, task, and person factors are likely to interactively determine patterns of within-individual variability in cognitive performance. For example, individuals may respond differently to changes in the stakes of the situation depending on their motivation, which may in turn be influenced by the complexity of the task presented to them. A key strength of using the ESM to understand within-individual variability is that by holding certain factors constant, we can elicit how people respond to changes in other factors over repeated administrations. To elaborate on the previous example, we could hold the stakes of the situation constant (e.g., high stakes), manipulate task complexity, and ask participants to provide indications of their motivation and engagement over multiple occasions across time. Thus, what we would capture and isolate would be both motivational and task performance changes in response to changes in task complexity under a specific situation. Consequently, a critical aspect of using ESM paradigms to capture within-individual variability in cognitive ability will be the manipulation of one or more situation, task, or person factors while holding other factors constant so as to elucidate the unique and interactive contribution of various factors to performance differences.

3.2.4. Task Choice

If our aim is to understand structured within-individual variation in cognitive performance, we must employ tasks that enable observation and analysis of variability in the processes people typically use to solve a cognitive challenge. Typical every-day cognitive tasks are dynamic and complex, and require motivation and engagement to learn, apply, and master skills and concepts across different activities and contexts (Goff & Ackerman, 1992; von Stumm et al., 2011). A static task, such as a standard reasoning measure employing only one type of test item (e.g., all figural matrices) is unlikely to replicate a typical environment because it does not require dynamic processes and this type of engagement diversity. Moreover, single-item or very short cognitive tests do not allow for the extraction of metrics that are important for understanding typical ability processes such as learning trajectories, cognitive exploration, and complex problem-solving (Birney et al., 2019). How rapidly people learn and apply new concepts, and their cognitive exploration styles has theoretically stronger links to typical cognitive abilities than performance on a binary score on a single item.

Consistent with Birney and Beckmann (2022), it is for this reason that we propose that dynamic cognitive tasks are by definition the most appropriate tasks to elicit meaningful within-individual variability. Specifically, these tasks allow analysis of different levels of time-structured variability – between-occasion variation (i.e., performance trajectories across different measurement occasions) and within-occasion variation (i.e., performance trajectories within measurement occasions, across different trials), both of which were outline in Chapter 2 Section 3.2.1. These trajectories allow researchers to understand factors that influence learning and insight in typical cognitive performance environments, and may be modelled in, for instance, complex problem-solving tasks and microworlds. Other interesting tasks features that may be susceptible to manipulation (and therefore investigation) include insight problems (for understanding dynamic aspects of creative cognition) and strategy tasks (for understanding strategy selection and application in cognitively challenging scenarios).

3.3. Methodological Considerations

There are also methodological considerations when designing an ESM style cognitive assessment. Here we focus on several key method and design issues: the appropriate duration and spacing of repeated measures assessments, and considerations for participant recruitment and retention.

3.3.1. Duration and Spacing of Measurements

To summarise Chapter 2 Section 3.2, within-individual variability in cognitive performance is likely to be structured between and within measurement occasions, and influenced by situation, task, and person factors. It is consequently a complex phenomenon and significant consideration should be given to task choice, as well as the choice of situation, task, and person variables to manipulate for investigation and assessment. When analysing within-person variation it is also critical to align the duration and spacing of measurements with a theoretical conception of the phenomena being measured (Ram & Gerstorf, 2009). Within-individual variability in personality state expression is a phenomenon that usually unfolds below the threshold of conscious awareness and thus likely requires participants to merely reflect or report on their current state. Consequently, personality states have been relatively well measured with one or only a few items per trait that collectively take one or two minutes to complete (Rauthmann et al., 2016; Rauthmann et al., 2015; Sherman et al., 2015). Accordingly, personality measures are relatively unobtrusive, can be completed in a minute or two, and do not require a substantial shift in focus from whatever the participant was engaged with prior to the prompt to respond (Fisher & To, 2012). As a result, it is feasible to have several measurement occasions per day (Church et al., 2013; Fleeson, 2007; Jones et al., 2017). This allows researchers to capture transient changes in personality state expression triggered by fluctuations in the external environment experienced during a typical day, and then investigate the structure of this variation (Rauthmann et al., 2016; Rauthmann et al., 2015; Sherman et al., 2015).

Conversely, within-individual variability in cognitive ability states may be more enduring within a day and thus the factors linked to structured variability may sometimes persist for longer, for example, due to cognitive fatigue or motivational factors, such as workplace engagement and perceived task importance. Moreover, as outlined in Chapter 2, Section 3.2.4, regardless of whether a static or dynamic task is chosen, ideally it will need to be of sufficient length to capture between- and within-occasion variation. Consequently, ESM cognitive studies are likely to present participants with longer tasks than is standard in personality studies. In addition, cognitive tasks require a significant shift in attention and motivational resources away from whatever the participant was previously doing. This will be more taxing and arguably necessitate greater participant commitment and engagement than for personality studies. Accordingly, participants may need to be given longer time windows to complete a cognitive task after receiving an ESM prompt, e.g., several hours rather than an hour or less as in personality research.

When studying cognitive ability variation, we may also only be interested in understanding variation of experience across a constrained time period. For example, if our goal is to assess ability variation as an indicator of future workplace performance, we might focus measurement occasions during usual working hours, rather than throughout a whole day as is commonplace in personality variation studies. Taken together, this suggests that cognitive studies with repeated measures will need to be limited to fewer but longer measurement occasion/response windows; that is, rather than five 1-minute tasks per day we might administer one 5-minute task per day.

3.3.2. Participant Recruitment and Retention

As we have outlined, short-term repeated measures cognitive studies are likely to be substantially more demanding than short-term repeated measures personality studies. As a result, strategies to boost participant recruitment and retention should be considered. Participant remuneration may need to be altered, for example, paying above standard rates. Researchers may also want to consider awarding bonuses to those who complete a certain number of tasks to incentivise high completion rates. In student samples where remuneration is in the form of course credit, there may be a benefit to setting high completion rate thresholds so that there are more data points per participant which increases measurement reliability and statistical power.

Another method for incentivising participant retention is task design. Gamification of the task environment via engaging semantics, animations, and interface design may assist to boost participant engagement. This could be done by embedding tasks within a relevant context and using interactive elements (task feedback). Such manipulations are entirely consistent with a goal to use more dynamic rather than static tasks (Birney & Beckmann, 2022).

3.4. Analytical Considerations

Once theoretical and methodological considerations have been factored into assessment design, there are several analytical considerations that will inform the data cleaning and analysis approach. Here, we consider how data pre-processing and cleaning could be approached. We also present considerations for extracting ability parameters from data, whether ability is construed as performance scores or response time, and introduce the implications of these parameters for reliability and boundary effects within the data.

3.4.1. Data Pre-processing

Because cognitive tasks take longer to complete and are more intrusive on participants' time, researchers may need to be more circumspect when pre-processing repeated measures cognitive data. In repeated measures personality studies it is common to exclude tasks completed more than an hour after it is sent to participants (Church et al., 2013; Fisher & To, 2012; Jones et al., 2017; Rauthmann et al., 2016; Rauthmann et al., 2015; Sherman et al., 2015). As outlined above, this approach may be unfeasible in cognitive ESM studies. Furthermore, exclusion criteria for purportedly non-serious responding may need consideration. Excluding based on extreme performance scores or response times may lead to an unrepresentative sample of data variability, particularly if the cognitive task is dynamic.

Dynamic tasks present participants with the opportunity to explore a novel system and, as a result, extreme scores may reflect exploration for some, and non-serious responding for others. Including extreme scores in data analysis may tell us something important about people's typical cognitive ability variability. The fact that the distribution of a person's performance variation includes extreme exploration strategies and/or non-serious attempts may be an indicator of a substantive cognitive strategy when presented with challenges. In turn, this may be important for understanding applications of cognitive abilities in educational and occupational contexts. In sum, we suggest that researchers develop their data cleaning approach in consideration of faithfully replicating the phenomena of interest rather than simply adhering to conventional exclusion criteria.

3.4.2. Extracting Parameters From Data

What parameters to extract from cognitive ESM data is an open question and likely to require further theoretical consideration and experimentation. Personality state items are restricted in their response options. Generally, participants are asked to indicate how well an item describes their current state. The variability in these ratings across time can then be analysed using a range of metrics. Common metrics include individual standard deviation (iSD), variance, absolute range, interquartile range, and mean squared successive differences (MSSD) (Ram & Gerstorf, 2009; Wendt et al., 2020). These metrics all provide an indication of how much an individual varies across time. More recently, multilevel modelling methods have been employed to understand how personality states vary contingent on other factors over time (see e.g., Beckmann et al., 2020; Beckmann et al., 2021; Wood et al., 2019). Regardless of the modelling method used, there is an emerging consensus that one's current personality state can be captured by a single, point-in-time score. How we select an outcome variable from a cognitive ability task intended to capture within-individual variability is not as clear.

To explicate the challenge, consider performance on a typical dynamic inventory-management style microworld task (e.g., Birney et al., 2018), as represented in Figure 2.3. Here, the participant is asked to maintain a system state of 100 and given 15 trials to do so. There are numerous possibilities for parameterising the state of the system, some are illustrated in Figure 2.3. As for personality data, we could take a simple metric such as the mean or median state, the variance, range, or interquartile range of states, the end point (final trial state), or the cost (cumulative state penalty of all decisions), shown in Figure 2.3 points 1-6 and 10. These metrics tell us something about someone's mean level of performance and variability, as well as parameterising the amount of exploration of the system. However, these metrics likely miss fundamental structural aspects of within-individual variability in cognitive ability, such as variation in effectiveness and efficiency of within-occasion learning. Understanding these aspects require focusing on the analysis of performance trajectories and change points (Figure 2.3, points 7-9). When we ask whether and how such within-individual parameters differ between individuals, we open the potential for more holistic models of cognitive ability and cognitive performance differences under more typical everyday conditions.

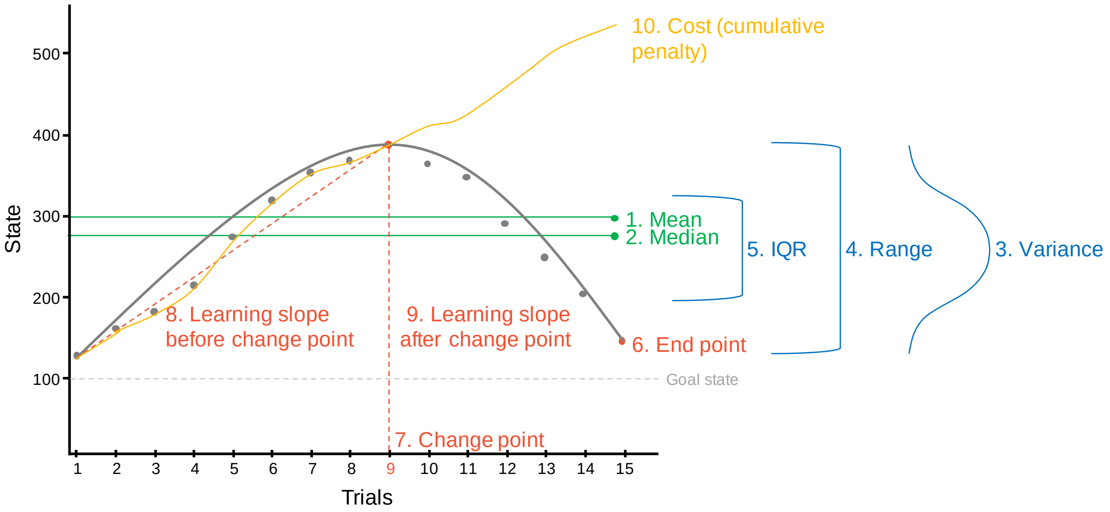


Figure 2.3. Example of performance metrics extracted from a dynamic microworlds task.

3.4.3. Variability in Performance vs. Response Time

The choice of dependent variable in cognitive ESM studies is a further consideration. The preceding discussion assumes accuracy/error scores are the dependent variable; however, another candidate could be response time, or some function of it. Understanding within-individual variation in personality does not necessitate the analysis of response time, as they are not likely to tell us much about the dynamic nature of participants’ personality variation beyond, say, identifying non-serious responding. However, response time is often important for understanding cognitive performance even on tasks where the focus is on accuracy. A critical aspect of typical cognitive performance is motivation and engagement; it is not enough to just finish the task, the participant must actively engage with it, particularly if it is a dynamic one (Goff & Ackerman, 1992; von Stumm et al., 2011). Analysing structured variability in response time data alone and in conjunction with accuracy data may provide insights into strategy use, cognitive exploration, and flexibility, as well as task engagement.

3.4.4. Reliability and Boundary Effects

Well-established methods exist for gauging the reliability of personality and cognitive measures. They rely on standardised administration as well as statistical assumptions, such as local independence (Birney et al., 2022). However, applications of standard reliability coefficients are challenging when applied to data from dynamic cognitive tasks and repeated measures which entail within- and between-person variation in structural features, such as administration under different situations and when differential feedback is provided.

A potential solution lies in the fact that repeated measures data lends itself to multilevel modelling (MLM) and the decomposition of variance into structured effects both between- and within-individuals. MLM allows for the estimation of random effects, which are model-implied individual parameter estimates within and across levels of the data. These have promise for use as adjusted predictors of other outcomes of interest. Liu et al. (2021) examined the reliability of MLM random effects using both simulated and empirical data and multilevel regression and structural equation models. They concluded that Empirical Bayes estimated random effects produce biased regression coefficients but that a wide variety of factors influence their reliability: larger variance in Level 1 predictors, larger variance in Level 2 random effects, and a higher number of observations per person each increase reliability, whereas larger Level 1 residual variance decreases reliability. To caveat these findings, they reported that individual-specific differences in these factors further influence the reliability of random effects, e.g., random effects of individuals with a low number of observations are likely to be biased towards zero. Accordingly, they caution researchers to understand why data is missing before using random effects as predictors, e.g., individuals may consistently miss response occasions for a reason that is relevant to the variables of interest, thus not providing an accurate picture of within-person factors.

Depending on the parameters extracted from the data, ESM studies may also yield boundary effects and heteroscedasticity. This is particularly likely if the dependent variable under investigation is an indicator of within-person variability. People whose performance is consistently at the higher or lower end of a trait dimension will by definition have less space for their states to vary in the scale-bounded direction. For example, those both low and high in Conscientiousness or Abstract Reasoning capacity tend to vary less. Methods for adjusting for such boundary effects exist (Mestdagh et al., 2018) and have been successfully applied in personality data (Beckmann et al., 2020), and could be leveraged when analysing within-person cognitive data also.

4. Conclusions

While there is empirical evidence for within-individual variability in cognitive task performance, our theories of and methods for understanding cognitive ability have not accounted for this. Our typical cognitive performance in education and work is inherently variable; over time, we perform differently in response to changes in the situation and tasks we are presented with, and differences in person factors such as personality motivational traits, and task learning and strategy use. However, the study of intelligence continues to rely on between-person comparisons of single-occasion test scores that do not consider the possibility or implications of within-individual variability in *cognitive ability* impacting *cognitive performance*. Furthermore, there have been few attempts made to understand the extent, magnitude, and structure of this variability. Without understanding this, our theories of cognitive ability will remain incomplete.

In this chapter, we presented a case for why we should study within-individual variability in cognitive ability. In doing so, we highlighted the utility of a within-individual approach in related fields, particularly the personality literature. In personality research, studying within-individual variability in personality expression has fundamentally altered our understanding of personality as a construct by facilitating a within-person process account of structured personality variation over time. These findings have been unearthed by pivoting away from between-individual single-occasion administration and analysis methods towards within-individual methods. We believe that there are lessons to be learned from this approach for those who research cognitive ability. Specifically, we advocate for the use of the ESM to generate rich, within-person data. However, we cannot simply copy and paste the methods used in personality as there are important distinctions to be made between the theory, methods, and analytical approaches used.

Within-individual variability in cognitive performance is multifaceted and complex, possessing both between-occasion and within-occasion variability components and likely to be influenced by a broad range of situation, task, and person factors. Conversely, within-individual variability in personality represents relatively transient momentary fluctuations in personality states in response to situational changes that can be captured with single or few item assessments. Consequently, significant consideration must be given to the design of cognitive ESM studies and the analysis of resulting data. We suggest that dynamic tasks will be required to capture between- and within-occasion performance variation. Moreover, researchers should seek to experimentally manipulate various situation, task, and person factors to elucidate their unique and interactive contributions to between-person differences in within-individual variability. As a result, cognitive ESM studies are likely to be longer and more demanding than their corresponding personality assessments, which has implications for the spacing of assessments and participant recruitment and retention strategies. Moreover, researchers will need to carefully consider the phenomena of interest (i.e., within-individual cognitive performance variability) when deciding what performance parameters to extract from the data and when making data cleaning decisions.

The following chapters will focus on the feasibility of this style of assessment, how we can extract parameters from this data, and whether these parameters have incremental predictive utility over and above between-person performance indices for assessing success in educational environments. In presenting these arguments, we hope that researchers will be prompted to consider how integrating a within-person perspective into cognitive ability could benefit our theories and methods for assessing cognitive ability in the 21st century.

III. MEASURING ECOLOGICAL COGNITIVE PERFORMANCE

1. Introduction

Over the following four chapters, we implement the ideas proposed in Chapter 2 across two empirical studies. Our aim across these chapters is to understand whether a more authentic assessment and analysis approach that models time and context as substantive sources of differences, rather than treating it as noise, leads to better prediction of common criterion measures than single occasion composites. Across these chapters, our criterion measure of interest is university performance.

In this chapter, we explore and propose three classes of quantitative parameters that leverage short-term repeated measures data generated from an EMA approach to model the extent, magnitude, and importance of ecological within-individual variation in cognitive performance. We refer to these select parameters collectively as “cognitive dynamics” to highlight they are indicators of performance derived across different repeated measures and settings more so than is typically the case. In Chapter 4, we apply these cognitive dynamics in an EMA study (Study 1) using a standard reasoning task administered repeatedly in idiosyncratic situations over two weeks. We compare the predictive utility of the cognitive dynamics with a measure of cognitive ability as traditionally conceptualised and assessed: a single occasion, context-invariant assessment where performance is measured as a sum score of correct responses. In Chapter 5, we apply the cognitive dynamics in a second EMA study (Study 2) using a microworld task requiring dynamic decision-making administered repeatedly over three weeks, also in idiosyncratic situations. As for the first EMA study, we compare the predictive utility of the cognitive dynamics with cognitive ability as traditionally conceptualised and assessed. We believe understanding ecological cognitive performance on both types of tasks is important as they reflect different operationalisations of cognitive ability. Moreover, the two study tasks were chosen as they generate different types of data; reasoning tasks generate binary response data and microworld tasks generate continuous response data. We apply our hypothesised cognitive dynamics parameters to each. Any assessment method can only be useful if it is feasible and acceptable to those who use it, therefore across Chapters 4 and 5, we also briefly discuss engagement patterns in both empirical studies. Finally, in Chapter 6 we synthesise and discuss the findings of Study 1 and Study 2.

2. Background

2.1 What is ecological cognitive performance?

As explicated at length in Chapter 2, cognitive ability tests are traditionally administered during a single testing session, from which mean or total test scores are extracted and interpreted. In the case of assessment and selection, these single session performances are often taken under high stakes and thus scores will reflect maximal performance because the expectancy of a valued outcome and the concordant motivation to exert maximum effort are both high (Goff & Ackerman, 1992; von Stumm et al., 2011). These *maximal performance measures* are typically then used to rank the likelihood of future performance. In doing so, its assumed that maximal performance scores from single testing sessions are valid indicators of criterion outcomes which instead reflect *ecological cognitive performance* across varied situations where maximal performance is often impossible to sustain (Wittmann & Süß, 1999, referred to such situations as a violation of Brunswik symmetry). If we want to understand ecological cognitive performance, cognitive ability assessments must be designed to capture this broad, dynamic, and longitudinal cognitive capacity that varies over time and context rather than relying on single occasion, context-invariant assessments.

2.1.1 A note on the predictive utility of cognitive ability

Despite this theoretical and measurement asymmetry, we do acknowledge that historically, single occasion cognitive ability assessments have predictive utility of important outcome criterion variables – we are not disputing this in and of itself. For instance, Schmidt et al. (2016) report that roughly 42% of variability in job performance can be accounted for by single occasion assessments of general cognitive ability, while Zaboski et al. (2018) report that 54% of variability in school performance can be explained by single occasion assessments of general cognitive ability. However, recent research has called into question the magnitude and robustness of these relationships (Richardson & Norgate, 2015; Sackett, Berry, et al., 2023; Sackett, Demeke, et al., 2023; Sackett et al., 2022).

Re-examinations of many prominent meta-analyses regarding the predictive utility of cognitive ability has demonstrated a range of statistical issues, including heterogeneity of ability measures, small sample sizes, and assumptions about normality and randomness that are unverifiable due to unavailability of the original data (Richardson & Norgate, 2015; Sackett, Berry, et al., 2023; Sackett, Demeke, et al., 2023; Sackett et al., 2022). Furthermore, Richardson and Norgate (2015) noted that there is substantial circularity in the predictors and criterion; many of the underlying cognitive ability tests used in these meta-analyses were designed to capture capacities that are acquired in formal education (e.g., general crystallized (Gc) abilities), therefore it is self-evident that they also predict academic success and workplace success in roles reliant on accumulated knowledge. In addition, causality has been disputed with research demonstrating that the relationship between cognitive ability and job performance is fully mediated by job knowledge (Palumbo et al., 2005, as cited in Richardson & Norgate, 2015).

Finally, the inclusion of very old studies in many meta-analyses may obscure the true relationship between single occasion assessments of general cognitive ability and criterion outcomes because the nature of education and work has changed substantially over time (Sackett, Demeke, et al., 2023; Sackett et al., 2022). Sackett et al. (2022) report that when accounting for the aforementioned statistical issues, cognitive ability has a mean corrected correlation of .31 for predicting job performance in 20th century data. However, when using only 21st century data and accounting for statistical issues, Sackett, Demeke, et al. (2023) report that cognitive ability has a mean corrected correlation of only .22 and relatively large residual standard deviation of 0.11. Collectively, this research suggests that previously higher meta-analytic estimates of the relationship between cognitive ability and educational and/or job performance should be considered theoretical maximum correlations only. This also implies that single occasion measures of general cognitive ability may not be as powerful a predictor as it was once (almost) universally accepted in the field to be and that there is room for improvement in our cognitive assessment practices.

2.2 Prior research on ecological cognitive performance

Capturing ecological cognitive performance has a rich and storied history in psychological research that, for much of the past century, has fallen by the wayside. Danziger (1990) recounts that much psychological research from the late 19th and early 20th centuries focused on a small number of subjects whose behaviour, thoughts, and performance were studied and reported individually. This ecological, within-individual perspective was most of interest because, at the time, many considered that the pursuit of psychological knowledge was about understanding what was happening within the mind of the individual. However, such research paradigms rightly drew criticisms owing to the statistical complexities of analysing small sample sizes and generalising findings to wider populations. As a result, the study of the individual fell out favour and was replaced by survey-style paradigms that allowed researchers to capture the behaviour, thoughts, and performance of large groups of individuals in a more robust and generalisable way. This mass testing paradigm laid the foundations for much of modern intelligence testing, which assumes that abilities differ between individuals but not within individuals and consequently relies on between individual comparisons of single occasion cognitive ability test scores for assessment and selection (Ackerman, 1996). This paradigm has continued to dominate intelligence testing for over a century.

Despite this prevalent between-individual perspective, prior research has demonstrated that performance on a range of cognitive tasks varies within individuals (Allaire & Marsiske, 2005; Bunce et al., 2004; Li et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009; Salthouse, 2007, 2012; Salthouse & Berish, 2005; Salthouse et al., 2006; Schmiedek et al., 2013). However, to date, the extant research has not comprehensively examined whether this variation is more than just measurement error and is in fact systematic and meaningful for our theories of cognitive ability and methods of cognitive assessment. This is the question that we hope to begin answering here.

2.3 How can we assess ecological cognitive performance?

Capturing ecological cognitive performance requires an assessment method that generates rich, within-person data. In Chapter 2, we proposed the use of the EMA paradigm to assess ecological cognitive performance. Such an assessment method, we argued, is more symmetrical with criterion outcomes like educational and workplace performance because the context in which the assessments are taken is symmetrical with the context in which educational and workplace performance unfolds – dynamic, variable, and often unpredictable.

3. Measuring Ecological Cognitive Performance

We now turn to the question of how data generated from the EMA can be used to understand ecological cognitive performance. Statistical parameters that capture ecological cognitive performance, or *cognitive dynamics* in our framing, have not yet been well identified. Related fields such as personality and affect have derived some relatively simple parameters to measure central tendency and distributions of performance, a detailed overview of which can be found in Wendt et al. (2020) for those interested. In affect research, Wendt et al. (2020) refer to these as ‘indicators of affective dynamics’. We consider many of these parameters here. However, there are other parameters that capture performance variability that lie outside of this literature. Therefore, we also propose a range of parameters informed by our theoretical concept of ecological within-person variation in cognitive performance (a detailed exposition of this perspective can be found in Vaughan & Birney, 2023). In this section we propose three classes of statistical parameters that capture cognitive dynamics: 1) central tendency, 2) dispersion, and 3) contingencies. We outline these three classes in more detail and summarise our proposed parameters in Table 3.1. Again, we refer to these collectively as ‘cognitive dynamics’ to highlight that they are indicators of performance derived in more diverse ways than typical single occasion assessments.

3.1 Central tendency parameters

Central tendency parameters can be used to describe ecological within-person variation when the underlying data is repeated measures data (e.g., data generated from EMA). This is because these parameters will capture the typical level of one’s performance over time and varied contexts. Two such parameters are the mean and the median. While both parameters are likely to be correlated with traditional single-point-in-time measures of performance, they inherently capture something different about an individual: central points around which the ecological application of an individual’s cognitive ability varies over a short period of time and across dynamic assessment contexts, as opposed to total scores taken at a single point in time in an unchanging context. This variation in the application of cognitive ability may be attributable to differences in various factors that influence cognitive performance (e.g., different situations, varied task complexity, motivation, and cognitive engagement). Comparisons of mean EMA performance are likely to be more suitable when data is symmetrical distributed whereas comparisons of median EMA performance are likely to be more suitable when it is not.

3.2 Dispersion parameters

Dispersion parameters can also be used to describe ecological within-person variation because they can capture characteristics of an individual’s cognitive performance distribution over time. How individuals vary in the magnitude or shape of their performance distribution over repeated EMA measures is meaningful to consider in and of itself. However, the variation captured by these parameters may also be meaningful in the context of an individual’s central tendency parameters; two individuals could theoretically have the same dispersion parameters and different points of central tendency, or the same points of central tendency and different dispersion parameter scores.

The first dispersion parameter we propose is within-person standard deviation (SD), which provides information about how scattered an individual’s performance is around their own mean over the repeated measures. Using within-person SD in isolation allows the identification of between-individual differences in the magnitude of ecological cognitive performance variation, while using within-person SD in conjunction with central tendency parameters allows the identification of between-individual differences in the location and magnitude of ecological cognitive performance variation. These between-individual differences may have implications for predicting performance in cognitively challenging environments such as school, university, or the workplace, where consistency of performance across time and contexts is important.

Where the underlying data is time-structured and autocorrelated, ecological cognitive performance variation could also be captured by the within-person mean square successive difference (MSSD), calculated as:

$$MSSD = \frac{\sum (X_{i+1} - X_i)^2}{(N-1)}$$

Which can be contrasted with the regular standard deviation:

$$SD = \frac{\sum (X_i - \bar{X})^2}{(N-1)}$$

This is because the within-person MSSD demonstrates the degree of consistency in an individual’s performance from trial-to-trial in circumstances where future trial responses are dependent on previous trial responses. A low within-person MSSD would indicate relative consistency in an individual’s response pattern, whereas a high within-person MSSD would indicate inconsistency in an individual’s response pattern. Such differences can prima facie capture the basis of adaptability and flexibility of an individual’s strategy when used alone or combined with central tendency parameters. An individual with good[4] within-person central tendency and low within-person MSSD may generally perform well and be adaptable and flexible in their strategy use, whereas an individual with good within-person central tendency and high within-person MSSD may generally perform well but do so despite a more maladaptive and inflexible strategy. These between individual differences may be predictive of performance in other cognitively challenging scenarios where adaptability and flexibility

---

[4] Where good is defined as some optimal level on the performance scale.

in response to dynamic and variable situations are important, for example, in certain occupations or educational scenarios demanding rapid problem-solving and the integration of multiple competing perspectives.

Ecological cognitive performance variation can also be characterised by within-person skewness, which represents asymmetry in performance around its central tendency across measurement occasions. Skewness of performance may be a useful parameter for comparing between-person differences in ecological within-person variability because it provides additional information about the spread of people's performance variability over time and contexts that is not captured by their within-person SD or within-person MSSD. Specifically, individuals with within-person skewness statistics close to zero have relatively symmetrical performance variation. If symmetrical variation around the mean is systematic, then it may be indicative of inflexibility in strategy use depending on the nature of the outcome variable. Equally, it could indicate that an individual has predominantly performed around the same level, give or take some noise. Conversely, individuals with non-zero within-person skewness have performance variation that is tilted in one direction versus another. Such tilting is suggestive of systematic responding (in one direction vs the other). Consideration of the individual's within-person central tendency in conjunction with their skewness can provide some additional information to distinguish the functionality of the skewed response distribution.

The final dispersion parameters we propose are within-person best and worst performance, which are meaningful between-person comparators because they provide information about relative levels of maximum and minimum performance across people. People with similar points of central tendency and/or similarly shaped distributions may have different maximal and minimal performance levels. Comparing individuals best and worst performance scores may have practical utility in typical cognitively challenging situations such as university or the workplace, where our performance will vary over time and contexts, and relative levels of peak and lowest performance may be important for success. For example, in our Figure 2.1 scenario, individuals with relatively better worst performance may be preferred for situations in which lapses of performance are more costly.

3.3 Contingency parameters

While the dispersion parameters outlined in the previous section are indicators of interest, the variation they capture is just variation without further information to contextualise whether it is adaptive or not. Ideally, one needs to know whether this variation is contingent in some way on other situational, task, or person-level characteristics to better understand if it is psychologically meaningful. This information can be captured by our third and final set of parameters, contingencies. Following on from work done in personality (e.g., Beckmann et al., 2021; Minbashian et al., 2010), Birney and Beckmann (2022) define adaptive contingencies as within-person trajectories of performance that are conditional on some task, situation, and/or person feature.

Task features may include item difficulty, item presentation order, and within- or between-occasion manipulations, and person-related task manipulations (e.g., task design, availability of information, task complexity) (Beckmann, 2010; Goff & Ackerman, 1992). *Situation features* may include the stakes/importance of the situation (Goff & Ackerman, 1992; von Stumm et al., 2011), environmental comfort such as room temperature (Zhang et al., 2019) and lighting (Keis et al., 2014; Ko et al., 2020), and the presence of distractions (Graydon & Eysenck, 1988). Finally, *person features* may include traits with links to typical cognitive performance such as Conscientiousness, Openness/Intellect, Need for Cognition, and Goal Orientation (Schmidt & Hunter, 1998; Steinmayr et al., 2011; Strobel et al., 2019), as well as more momentary features such as state personality, affect, neuromodulatory processes, stress, and motivation (Beal et al., 2005; Schmiedek et al., 2010), and metacognitive states (Birney et al., 2017). Contingencies reflect ecological within-person variation in cognitive performance in response to such factors. For example, the cognitive performance of an individual with a near zero contingency is not especially affected by changes in the factor but an individual with a stronger positive or negative contingency is affected more substantially. The implications of these different slopes may be of varied importance for predicting an individual’s success in a specific educational program or occupation, where adaptability and flexibility in response to task, situation, or person changes is important for performance.

In our studies we focus on two contingencies of performance: experience and cognitive difficulty. Such contingencies can be captured by multilevel models. Between-person differences in the direction and magnitude of within-task and between-task experience trajectories capture information about how individuals differ in their capacity to learn from experience within the short period of time in which a task is conducted and across the longer period of time of EMA testing (e.g., 2-3 weeks). This short and long-term learning capacity facilitates ecological cognitive performance in situations such as school, university, and the workplace in which we are constantly faced with new information and must adapt rapidly to solve problems, drawing on prior experiences. Between-person differences in the direction and magnitude of difficulty trajectories capture information about how individuals differ in their capacity to deal with increased cognitive challenge. For example, individuals who are less affected by changing levels of cognitive challenge may be better equipped to deal with the complex demands of the modern education system and many professional occupations.

Contingencies can also be captured by piecewise linear models that allow for changes in the slope of performance trajectories over time. Piecewise linear models demonstrate whether an individual’s performance changes at any point during the task and, if it does, provide an estimate of the location of the change point, and of the pre- and post-change point performance slopes for the individual. Such models may provide useful information about within-person performance on cognitive tasks where learning is expected, and trial-on-trial performance is autocorrelated. This is because the presence of a change point indicates some significant change in performance. The pre- and post-change point slopes demonstrate the direction and magnitude of the trajectories around the change point. These differences may provide valuable information about within-person differences in ecological cognitive performance because they likely reflect between-person differences in learning efficiency that are critical for educational and occupational success.

Table 3.1. Proposed cognitive dynamics to capture ecological cognitive performance.

Class	*Parameters*	*Description*	*Limitations/constraints*
Central tendency	Mean	Average of repeated measures performance scores	- More suitable for symmetrical data than asymmetrical data
	Median	Middle of repeated measures performance scores	- Suitable for both symmetrical and asymmetrical data
Dispersion	Standard deviation (SD)	Spread of repeated measures performance scores	- More suitable for symmetrical data than asymmetrical data - Requires a sufficient number of observations
	Mean squared successive differences (MSSD)	Variability in trial-on-trial performance scores (where scores are autocorrelated)	- Suitable for both symmetrical and asymmetrical data - Only applicable when scores are autocorrelated - Requires trial-on-trial performance data
	Skewness	Shape of the distribution of repeated measures performance scores	- Suitable for both symmetrical and asymmetrical data - Requires a sufficient number of observations
	Best performance	Best performance score out of all repeated measures performance scores	- Suitable for both symmetrical and asymmetrical data
	Worst performance	Worst performance score out of all repeated measures performance scores	- Suitable for both symmetrical and asymmetrical data
Contingencies	Experience trajectories (multilevel model)	Slope of performance trajectory across trials or blocks	- Suitable for both symmetrical and asymmetrical data - Requires trial-on-trial performance data
	Difficulty trajectories (multilevel model)	Slope of performance trajectory across difficulty levels	- Suitable for both symmetrical and asymmetrical data - Requires trial-on-trial performance data
	Experience trajectories (piecewise linear model)	Presence and location of a change in performance over time, and direction and magnitude of pre- and post-change point slopes	- Suitable for both symmetrical and asymmetrical data - Only applicable when within-task learning/performance change is expected - Requires trial-on-trial performance data

4. Implementing the Parameters

4.1 The present studies

We now turn to our empirical investigation of using the EMA to capture ecological cognitive performance. Across the next two chapters, we present the methods and discuss the results of two EMA studies. Our aims are to:

a) test our proposed set of EMA-derived cognitive dynamics for the first time
b) understand whether they have predictive utility beyond traditional single session cognitive ability measures, and
c) examine whether the predictive utility is greater for any proposed class of cognitive dynamics.

In both studies, we generate our proposed cognitive dynamics parameters and examine their practical utility by comparing how they predict university performance above and beyond performance on the single session cognitive ability task. That is, we construe *single session performance* as a measure of cognitive ability as typically conceived of and operationalised and *cognitive dynamics* parameters as measures of ecological cognitive performance. If the cognitive dynamics parameters do not predict academic performance beyond the single session task it suggests either the parameters do not capture anything distinct from cognitive ability as typically conceptualised and assessed or that our conceptualisation of ecological cognitive performance is not distinct from cognitive ability. However, if these parameters do predict academic performance beyond the single session task it suggests our parameters are capturing something distinct from cognitive ability as traditionally assessed that is important for real-world cognitive performance (i.e., ecological cognitive performance).

IV. STUDY 1: ECOLOGICAL COGNITIVE ASSESSMENT OF REASONING TASK PERFORMANCE

The purpose of this first empirical study is to derive and test cognitive dynamics parameters, proposed in Chapter 3, which are conceived to capture ecological cognitive performance. We hypothesise that ecological cognitive performance and associated cognitive dynamics parameters are more symmetrical conceptualisations and assessments of an important real-world criterion measure, academic performance, than traditional single session trait assessments. This is because short-term EMA assessment allows us to capture ecological cognitive performance variation across time and contexts over several weeks, which is an inherent aspect of academic performance. As a result, we expect that ecological cognitive performance and associated cognitive dynamics parameters will be better predictors of university performance than cognitive ability as traditionally conceptualised and assessed, as a single session assessment of a more static trait. In this study, we implement our cognitive dynamics parameters in a reasoning task consisting of abstract, numerical, and verbal reasoning items.

1. Methods

1.1. Participants

Participants were first- and second-year undergraduate psychology students ($N = 89$, $M_{age} = 20.89$ years, $SD_{age} = 4.45$ years, $N_{female} = 72$). In return for their participation they received course credit. When examining engagement with the EMA approach, we use the data from this full 98 person sample. However, when implementing the cognitive dynamics parameters, we excluded 21 participants who completed fewer than four EMA tasks or did not complete the single session task, leaving a final sample of 68 participants ($M_{age} = 20.38$ years, $SD_{age} = 2.84$ years, $N_{female} = 55$).

1.2. Materials

Reasoning Tasks. In this study we examine ecological cognitive performance on a reasoning task. The reasoning task is a standard fluid intelligence (Gf) task in which the participant makes a series of independent decisions based on an unchanging local[5] problem space to complete the task. The reasoning tasks were developed from the International

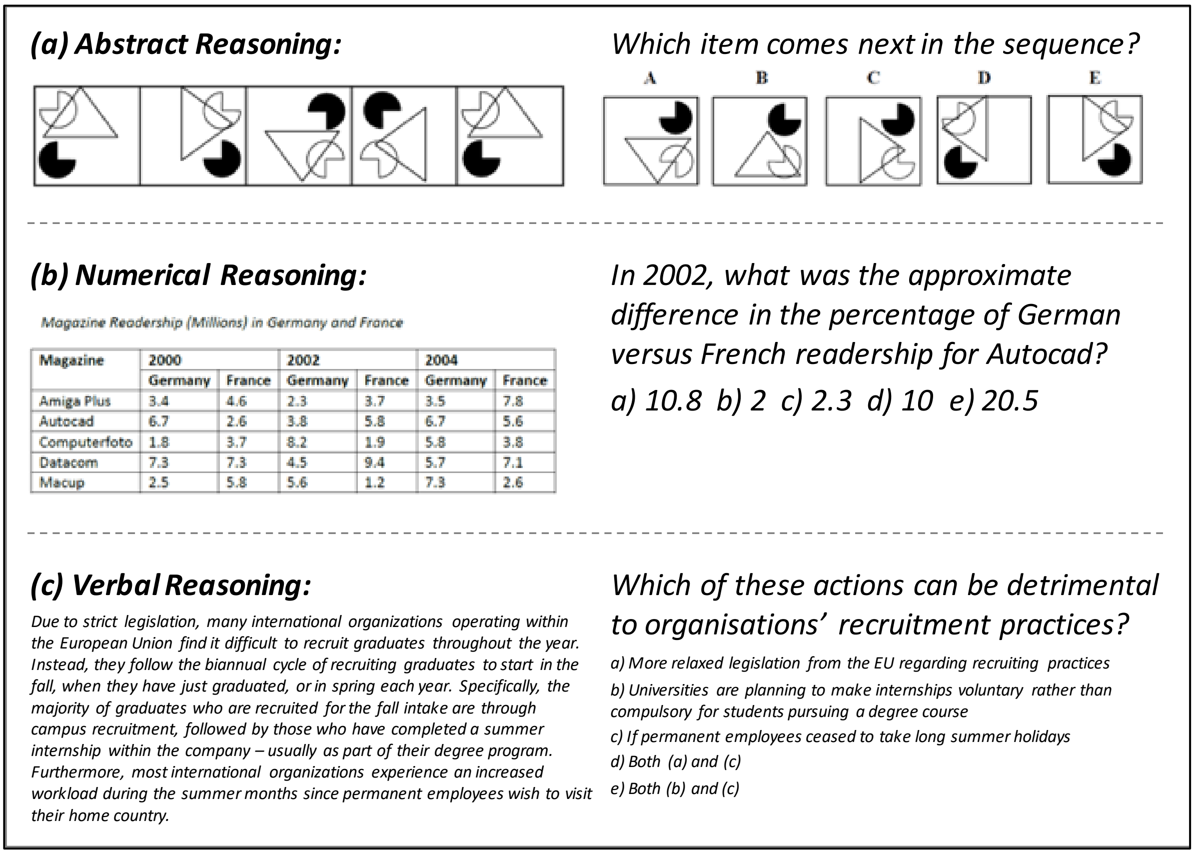
---

[5] Local in that it does not change while the task is being completed, though the broader situational context can vary from day to day.

Cognitive Ability Resource (ICAR) database (ICAR, 2014), which includes a variety of items assessing different cognitive capacities. Three sets of items were used: verbal reasoning, numerical reasoning, and abstract reasoning. *Verbal reasoning* items required the participant to read a short paragraph and answer a question based on the information it provided. *Numerical reasoning* items required the participant to read numerical information, usually in the form of a graph or table, and answer a question based on the information it provided. *Abstract reasoning* items required the participant to observe a series of five figures and determine which figure came next in the series. Sample items are contained in Figure 4.1.

As item difficulty data was not available for the selected items, for each reasoning domain items were rated as low, moderate, or hard difficulty by the first author and two research assistants. These ratings were then used to classify items such that 30 items each of low, moderate, and hard difficulty were extracted from the item pool (i.e., 90 items in total). Items and response options were reviewed by the first author and edited where required for clarity, brevity, and wording (in particular, some verbal reasoning items were updated to include more current dates than the original items). Furthermore, responses options were added where required to ensure each item had five options. The final set of items were then pseudo-randomly assigned to either the single session component or the EMA component such that there were 15 items from each reasoning domain in each component, with equal numbers of low, moderate, and hard items. The EMA items were then pseudo-randomised to EMA pulse tasks. Pulse tasks were short versions of the single session assessment intended to be completed via EMA and consisted of 9 items such that each pulse had 3 items from each reasoning domain, one of each difficulty level.

For the single session assessment, 45 items were presented. This consisted of 15 items each from the verbal, numerical, and abstract reasoning domains, presented as three domain-specific blocks of items. Block order and item order within block were randomised between participants. The single session assessment had good Cronbach's alpha reliability of $\alpha = .81$ For the EMA pulse component, pulse tasks were presented in the same randomised order to all participants, however, due to some participants missing a pulse the first time it was sent, not all completed the EMA in the order intended. The EMA pulse component had good Cronbach's alpha reliability of $\alpha = .82$. Reliability of between-person differences averaged over time was RkRn = .73 and reliability of within-person variation averaged over items was Rcn = .23 (mlr function from psych R package: Revelle, 2023).



(a) Abstract Reasoning:

Which item comes next in the sequence?

(b) Numerical Reasoning:

Magazine Readership (Millions) in Germany and France

Magazine	2000		2002		2004	
	Germany	France	Germany	France	Germany	France
Amiga Plus	3.4	4.6	2.3	3.7	3.5	7.8
Autocad	6.7	2.6	3.8	5.8	6.7	5.6
Computerfoto	1.8	3.7	8.2	1.9	5.8	3.8
Datacom	7.3	7.3	4.5	9.4	5.7	7.1
Macup	2.5	5.8	5.6	1.2	7.3	2.6

In 2002, what was the approximate difference in the percentage of German versus French readership for Autocad?

a) 10.8 b) 2 c) 2.3 d) 10 e) 20.5

(c) Verbal Reasoning:

Due to strict legislation, many international organizations operating within the European Union find it difficult to recruit graduates throughout the year. Instead, they follow the biannual cycle of recruiting graduates to start in the fall, when they have just graduated, or in spring each year. Specifically, the majority of graduates who are recruited for the fall intake are through campus recruitment, followed by those who have completed a summer internship within the company – usually as part of their degree program. Furthermore, most international organizations experience an increased workload during the summer months since permanent employees wish to visit their home country.

Which of these actions can be detrimental to organisations' recruitment practices?

a) More relaxed legislation from the EU regarding recruiting practices

b) Universities are planning to make internships voluntary rather than compulsory for students pursuing a degree course

c) If permanent employees ceased to take long summer holidays

d) Both (a) and (c)

e) Both (b) and (c)

Figure 4.1. Sample reasoning items (ICAR, 2014).

Note. Correct answers are E, A, and B.

Academic Performance. Academic performance was computed as the principal component extracted from a Principal Components Analysis of first year average mark and performance in two first year introductory psychology courses.

Reliability and Power. Low within-person reliability, relative to between-person reliability can reflect smaller number of observations within an occasion, as well as low within-occasion variance in the item responses. Although this can be offset to some extent by using multilevel analyses (Nezlek, 2017), along with the smaller than typical sample size for an individual differences study, low reliability in our measures also has implications for the statistical power. For regression analyses with three predictors and testing the incremental prediction of one of them with R^2 change between 0.08 to 0.20, a sample size ranging from 93 to 34 (respectively) is required to achieve power of 0.8 (Faul et al., 2007).

1.3. Procedure

On sign up, participants were randomly allocated to complete either the single session component or the EMA component first. There was roughly a two-week break between components (ranging from 10-21 days). For the single session component, participants completed the 45-item reasoning assessment, which took around 30 minutes. For the EMA component, participants were sent the five reasoning pulse tasks over a 2-week period. Each pulse was sent to participants' email addresses and once received participants were given 24 hours to complete it. One pulse was sent every day from Monday to Friday between 10am and 3pm during this 2-week period. Each pulse took around 5 minutes to complete. At the end of both components, participants were sent a 5-minute feedback survey to gauge their experiences of completing the study. The questions focused on the acceptability and practicality of the EMA paradigm for assessing cognitive ability, and captured any difficulties encountered their participation. The results of this survey are not the focus of this paper.

1.4. Analysis Approach

Data analysis was conducted using RStudio (2020) and JASP (2023). Figures were generated using ggplot2 (Wickham, 2016) and sjPlot (Lüdecke, 2023). Single session score was calculated as a sum score of correct responses for the 45-item single session assessment. Each pulse score is a sum score of correct responses for that 9-item pulse. The proposed parameters in Chapter 3 were computed as follows. The central tendency and distributional parameters were derived from the five pulse sum scores. For contingency parameters we used item-level scores as the basis for models run using the lme4 package in Rstudio (Bates et al., 2015). The basic multilevel model for extracting experience and difficulty contingencies was computed with observation-level performance (i.e., item score) as the dependent variable, item order and block order as independent variables representing experience, and one independent variable representing difficulty. We included both contingencies in the same multilevel model to control for the effects of one contingency on the other. Random intercepts and random slopes for all independent variables were also included. The trial slope is defined as the item-to-item order trajectory of the 9 items within each pulse (thus, there were 5 observations, one for each pulse, for each of item positions 1 to 9). To derive the difficulty trajectory, rather than the subjectively rated difficulty we used to equate pulses, we followed Birney et al. (2017) and empirically calibrated item difficulties using the Rasch measurement model. For each of the 5 pulses, we computed an individual Rasch model and extracted item difficulty using the ltm package in Rstudio (Rizopoulos, 2006). We then determined the performance-difficulty slope for the 45 items from all 5 pulses. This resulted in the below generic model:

Level 1:

$$Y_{ij} = \pi_0 + \pi_1 * trial + \pi_2 * block + \pi_3 * difficulty + e_{ij}$$

Level 2:

$$\pi_{0j} = \beta_{00} + r_{0j}$$

$$\pi_{1j} = \beta_{10} + r_{1j}$$

$$\pi_{2j} = \beta_{20} + r_{2j}$$

$$\pi_{3j} = \beta_{30} + r_{3j}$$

Where I = trial and j = person; trial = item-order; block = pulse-order; difficulty = Rasch estimate.

The random slopes were then extracted for each individual and used as estimates of within-person variation in within-task experience trajectories (π_{1j}), between-task experience trajectories (π_{2j}), and difficulty trajectories (π_{3j}) [6].

2. Results

2.1. Participant Engagement with the EMA Approach

We examined engagement across the full sample of participants, including those who did not complete the study as instructed (i.e., did not complete all five tasks) ($N = 89$, $M_{age} = 20.89$ years, $SD_{age} = 4.45$ years, $N_{female} = 72$). The reason for this is that we wanted to examine engagement across all participants to understand the broader feasibility and acceptability of the EMA approach. The mean number of tasks completed was 4.02 tasks (80.4%, SD = 1.77 tasks) with a median of 5 tasks. This is in line with other EMA studies which have a 65-90% task completion rate (Csikszentmihalyi & Larson, 2014; Fleeson, 2007; Rauthmann et al., 2016). The mean delay between receiving and completing the task was 6.24 hours (SD = 4.43 hours) and the median was 5.49 hours. There was no correlation between number of tasks completed and mean completion delay ($r = .024$, $p > .05$).

We also obtained feedback from participants to understand their experiences completing the study. Feedback data was missing for 27 participants. Feedback results are

---

[6] This multilevel model was computed using lme4 with the function: glmer(score ~ 1 + trial + block + difficulty + (1 + item + block + difficulty | subject)).

summarised in Table 4.2. Participants generally reported positive experiences with the study interface and technical experience, as well as the time given to complete each task. Feedback was slightly less positive for the time spent on each task, although the majority of participants agreed that the time spent on each task was reasonable.

Table 4.2. Proportion of participants who somewhat agreed, agreed, or strongly agreed with statements in Study 1 (N=62).

Statement	Study 1
"I was given enough time to complete each task"	.742
"The time I spent on each task was reasonable"	.565
"The interface of the task was easy to use"	.774
"I was able to complete this study without any technical difficulties"	.887

2.2. Preliminary Analysis

Summary statistics for reasoning task performance are reported in Table 4.3. Intercorrelations between the academic performance factor and reasoning tasks are contained in Table 4.4. Full within-person performance distribution plots can be found in the supplementary materials.

Table 4.3. Summary statistics for reasoning tasks.

Task	N	M(%)	SD(%)	Range(%)
Single Session	68	43.17	15.82	68.89
EMA Pulse 1	68	37.98	21.47	88.89
EMA Pulse 2	67	34.00	19.96	77.78
EMA Pulse 3	65	55.56	23.90	88.89
EMA Pulse 4	68	50.82	20.61	100
EMA Pulse 5	67	53.4	21.56	88.89

Table 4.4. Intercorrelations between academic and cognitive task performance.

	Academic Performance	Single Session Score	Pulse 1 Score	Pulse 2 Score	Pulse 3 Score	Pulse 4 Score
Single Session Score	.366					
Pulse 1 Score	.426	.565				
Pulse 2 Score	.473	.576	.536			
Pulse 3 Score	.451	.619	.444	.537		
Pulse 4 Score	.473	.491	.427	.506	.635	
Pulse 5 Score	.338	.595	.351	.408	.543	.467

2.3. Predictive Utility of Cognitive Dynamics

We then turned to whether the cognitive dynamics parameters would predict academic performance above and beyond cognitive ability as traditionally assessed (i.e., performance on a standard single session cognitive task). A range of parameters displayed numerically stronger correlations with academic performance than performance on the single session task, as shown in Table 4.5. All intercorrelations are contained in the supplementary materials.

Then we computed a series of OLS simple linear regression models with academic performance as the dependent variable, single session assessment score entered as an independent variable in the first block, and the cognitive dynamics parameter entered as an independent variable in the second block. The results of these regressions are in Table 4.6[7]. All variables that were more strongly correlated with academic performance than cognitive ability as traditionally assessed were also incremental predictors of academic performance over and above cognitive ability as traditionally assessed, suggesting that they predict an aspect of an important ecological cognitive performance criterion outcome (academic performance) that asymmetrical single session test scores do not.

To investigate whether any of these parameters were especially strong drivers of academic performance and to account for any multicollinearity of the cognitive dynamics parameters, we ran a further OLS multiple linear regression model. We included academic performance as the dependent variable, single session cognitive assessment score entered as an independent variable in the first block, and all nine cognitive dynamics parameters from

---

[7] We also conducted a series of regressions testing whether any of the distributional parameters predicted additional variation in academic performance beyond the central tendency parameters, however, none were significant incremental predictors.

Table 4.5 entered as independent variables in the second block. The nine cognitive dynamics parameters in the second block were entered using stepwise estimation to determine which parameter or parameters were the strongest predictors of academic performance. Median total pulse score was the only variable in the second block that was included in this stepwise model as a significant predictor of academic performance ($\beta = 0.603$, $R^2_{model} = .350$, $R^2\Delta = .182$, $p(F\Delta) = <.001$).

Table 4.5. Correlation between cognitive dynamics and academic performance.

	Academic Performance
Central Tendency Parameters	
Mean	**.566**
Median	**.582**
Dispersion Parameters	
SD	.060
Skewness	-.101
Best Score	**.496**
Worst Score	**.480**
Contingency Parameters	
Trial Slope	-.181
Block Slope	.296
Difficulty Slope	**-.500**

Table 4.6. Additional predictive utility of cognitive dynamics.

	R^2_{model}	$R^2\Delta$	$p(F\Delta)$
Central Tendency Parameters			
Mean	.326	**.192**	**<.001**
Median	.340	**.207**	**<.001**
Dispersion Parameters			
SD	.127	.003	.638
Skewness	.136	.002	.723
Best Score	.254	**.120**	**.002**
Worst Score	.234	**.100**	**.005**
Contingency Parameters			
Trial Slope	.185	.017	.266
Block Slope	.211	.043	.074
Difficulty Slope	.261	**.093**	**.007**

Note. Compared against a base model including single session score as a predictor, which had an $R^2 = .134$.

2.4 Mediation Analysis

A series of mediation analyses were conducted to investigate the mechanisms behind the predictive relationships uncovered in the previous analyses; this mediation analysis is depicted in Figure 4.2. For each of the cognitive dynamics parameter that was a significant incremental predictor in Table 4.6, the model tests whether the relationship between cognitive ability as traditionally conceptualised (single session performance) and academic performance is mediated by ecological cognitive performance (cognitive dynamics parameters).

The results of these models are reported in Tables 4.7 and 4.8, which shows that the cognitive dynamics parameter in all models fully mediated the relationship between single session cognitive assessment score and academic performance and that this mediation was significant for all models. This relationship was unidirectional, as single session cognitive assessment score did not mediate the relationship between any cognitive dynamics parameter and academic performance (Table 4.8). The standard interpretation of full mediation is a directional one (as indicated by the arrows), in which the following claim is consistent: that there is no direct relationship between traditional ability and academic performance, but rather this relationship emerges because traditional ability *causes* differences in cognitive dynamics, and it is dynamic performance that *cause* important for academic performance. While we feel such claims based on correlational data are fraught, the findings do suggests that cognitive ability as traditionally assessed may not be capturing as full account of the ability-performance relationship picture as it is typically thought to do.

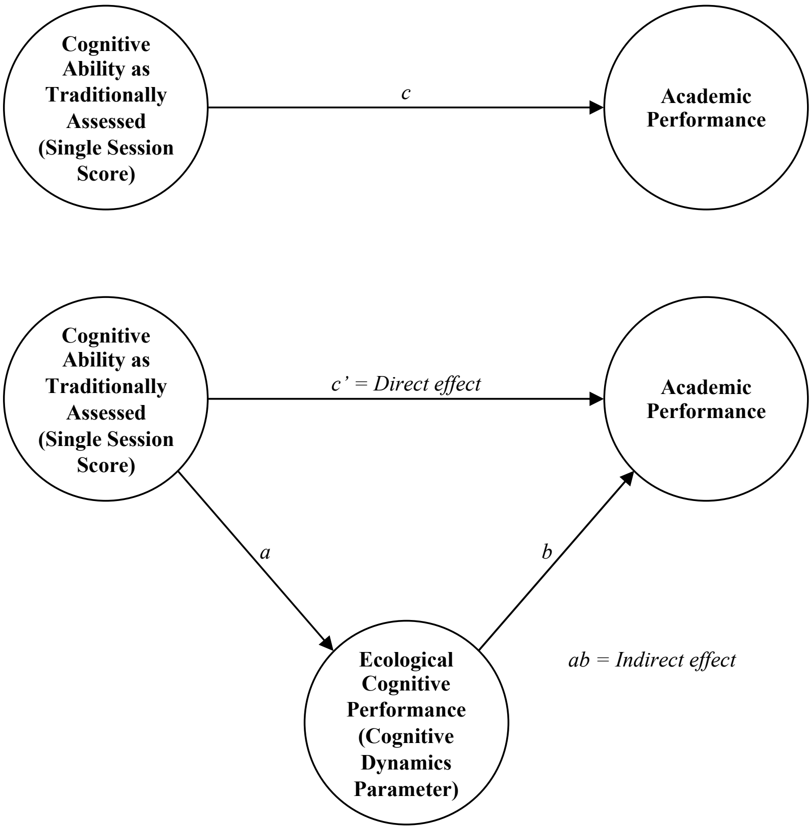


Figure 4.2. Representation of mediation models.

Table 4.7. Single session score predicting academic performance, mediated by cognitive dynamics parameters.

	Direct Effect		Indirect Effect		Significance	
Mediator	Estimate	p	Estimate	p	Sobel	p
Central Tendency Parameters						
Mean	-.014	.477	**.063**	**<.001**	8.51	**<.001**
Median	-.009	.628	**.058**	**<.001**	4.62	**<.001**
Dispersion Parameters						
Best Score	.015	.391	**.034**	**.004**	3.64	**<.001**
Worst Score	.011	.570	**.038**	**.006**	3.80	**<.001**
Contingency Parameters						
Difficulty Slope	.013	.499	**.036**	**.007**	3.95	**<.001**

Table 4.8. Cognitive dynamics parameters predicting academic performance, mediated by single session score.

	Direct Effect		Indirect Effect	
Mediator	Estimate	p	Estimate	p
Central Tendency Parameters				
Mean	**.430**	**<.001**	-.051	.478
Median	**.380**	**<.001**	-.028	.628
Dispersion Parameters				
Best Score	**.253**	**<.001**	.039	.396
Worst Score	**.276**	**.003**	.035	.571
Contingency Parameters				
Difficulty Slope	**-1.585**	**.005**	-.364	.331

3. Study 1 Discussion

We found evidence to support our argument that ecological cognitive performance is an important aspect of real-world cognitive performance not captured by traditional single occasion assessments of cognitive ability. Moreover, participants engaged with the EMA approach at similar rates to EMA studies in related fields and reported positive experiences with the task interface and study method.

It is important to note that the items are isomorphic in both assessments. The only difference is that the first version presented 45 items in a single session, whereas the second presented the items in 5 blocks of 9 items across 2 to 3 weeks, suggesting that our findings are not simply due to an increase in the number of assessment items. These parameters were derived from all three classes proposed in Chapter 3. Numerically speaking, parameters capturing central tendency of performance (mean and median) appeared to be the strongest predictors of academic performance, followed by best and worst scores, and difficulty contingency slope. These parameters fully and significantly mediated the relationship between cognitive ability and academic performance, and post-hoc mediation analyses suggested this relationship was unidirectional. In other words, these cognitive dynamics parameters fully accounted for variance in academic performance explained by cognitive ability as traditionally assessed but cognitive ability as traditionally assessed did not fully account for variance in academic performance explained by cognitive dynamics parameters. We suggest that what these parameters capture is information about the extent and magnitude of how individuals vary in the application of their cognitive ability in real-world cognitively demanding environments, such as university, which are dynamic and variable across time and contexts. In other words, cognitive dynamics represent ecological cognitive performance, which is more symmetrically aligned with this criterion outcome than a single occasion assessment conducted at a single point in time under a static context.

V. STUDY 2: ECOLOGICAL COGNITIVE ASSESSMENT OF MICROWORLD TASK PERFORMANCE

As for Study 1, the purpose of this second empirical study is to derive and test cognitive dynamics parameters, proposed in Chapter 3, which are conceived to capture ecological cognitive performance. Again, we hypothesise that cognitive dynamics parameters will better predict academic performance than a traditional single session assessment because cognitive dynamics capture variation in ecological cognitive performance across time and contexts over several weeks that is not captured in single occasion assessments. Unlike Study 1, in this study we implement our cognitive dynamics parameters on a dynamic microworlds task. The purpose of this contrast is to understand whether the results observed in Study 1 generalise to a more dynamic cognitive task that requires continuous engagement and in which there is not always one correct response.

1. Methods

1.1. Participants

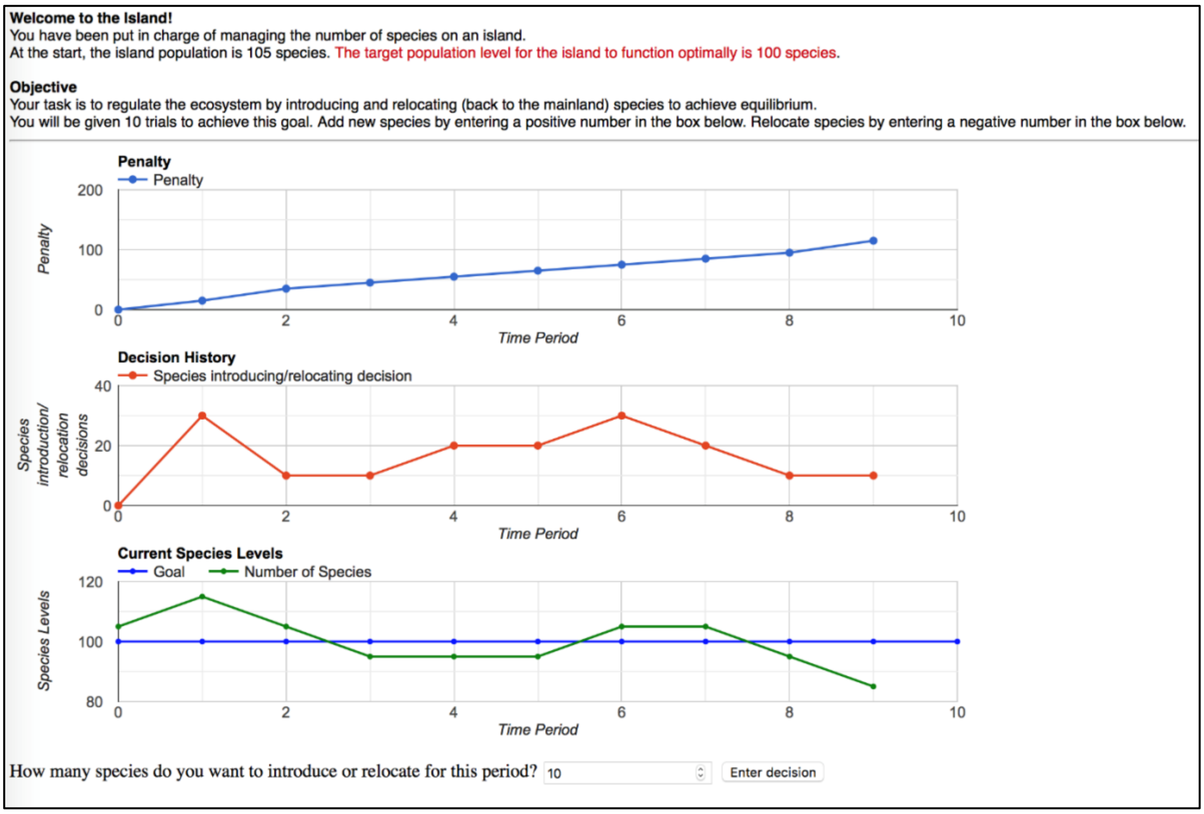
Participants were first- and second-year undergraduate psychology students ($N = 101$, $M_{age} = 21.13$ years, $SD_{age} = 5.46$ years, $N_{female} = 71$), who received course credit for participation. As for Study 1, when examining engagement with the EMA approach, we used the data from this full 101 participant sample but when implementing the cognitive dynamics parameters we excluded 19 participants who did not complete an acceptable minimum number of EMA pulse tasks (at least 8). A further two participants were excluded for having very high penalty scores on the microworld tasks, indicative of non-serious responding, leaving a final sample of 80 participants ($M_{age} = 21.13$ years, $SD_{age} = 5.49$, $n_{Female} = 57$).

1.2. Materials

Microworld Task. Microworld tasks require the manipulation of one or more inputs to achieve a target value in one or more outputs (Gonzalez, Thomas, et al., 2005; Gonzalez, Vanyukov, et al., 2005; Wood et al., 2009). The microworld task we used is a one-input, one-output task based on inventory management as presented by Birney et al. (2018), shown in Figure 5.1. In our shortened version, the "inventory" participants were put in charge of managing was an ecosystem of an island. On this island, the number of species decreases over time at some unstated rate. This loss is referred to as “outflow” and is independent of any input decision the participants make. The stated goal of the task is to keep the number of species on the island at a target level of 100. To do so, participants make a decision on each trial to introduce (add), relocate (subtract), or leave unchanged the number of species on the island.

There are two main rules governing the ecosystem: outflow, as just described, and delays in when a trial decision is actioned. *Outflow* could be of two types, either constant or random, and is described to participants as representing the number of species lost in a trial due to environmental influences. *Delays* were also of two types, either absent (no delay, trial decision is actioned immediately) or present (trial decision is actioned three trials later). Four variants of the task were created by crossing the outflow and delay manipulations as outlined in Table 5.1. In each session, participants completed 1 block of 15 trials of a single variant. Each variant was presented three times for a total of 12 pulse (EMA) blocks. The output variable is the number of species at the end of the trial.

To estimate the overall reliability of the microworld tasks, we computed a total penalty score for each block of trials for each participant equal to the sum of all penalty scores for all trials in that block. Reliability across the 12 blocks is α = .55.



Welcome to the Island!
You have been put in charge of managing the number of species on an island.
At the start, the island population is 105 species. The target population level for the island to function optimally is 100 species.

Objective
Your task is to regulate the ecosystem by introducing and relocating (back to the mainland) species to achieve equilibrium.
You will be given 10 trials to achieve this goal. Add new species by entering a positive number in the box below. Relocate species by entering a negative number in the box below.

How many species do you want to introduce or relocate for this period? 10

Figure 5.1. Microworld simulation interface (Birney et al., 2018).

Table 5.1. Microworld variants (participants completed each variant 3 times)

	Outflows	
Delays	*Fixed* (20 species per trial)	*Random* (between 10 and 30 species per trial)
Absent (no lag)	Variant 1 (x 3)	Variant 2 (x 3)
Present (3 trial lag)	Variant 3 (x 3)	Variant 4 (x 3)

Cognitive Ability. Cognitive ability was assessed using the ICAR (ICAR, 2014), an open-access cognitive ability assessment that is strongly correlated with longer and more complex measures of general intelligence (Dworak et al., 2021; Young & Keith, 2020). We used the 16-item version of the ICAR, which includes four items each on verbal reasoning, letter-number series, matrix reasoning, and 3D object rotation. This version of the ICAR is different to the items administered in Study 1. The presentation order of the four ability domains was randomised and the order of items within each domain was also randomised. The 16-item ICAR had good Cronbach's α reliability = .80.

Need for Cognition (NFC). An 18-item scale developed by Cacioppo and Petty (1982) was used to assess Need for Cognition and had good Cronbach's α reliability of α = .851.

Goal Orientation (GO). A 16-item scale developed by VandeWalle (1997) was used to assess Goal Orientation. The scale had good Cronbach's α reliability of α = .804 for the Learning Goal Orientation (LGO) sub-scale, acceptable reliability of α = .763 for the Performance-Avoid Goal Orientation (PGA) sub-scale, but relatively poor reliability of α = .574 for the Performance-Prove Goal Orientation (PGP) sub-scale.

Personality. Personality was assessed using the Big Five Aspect Scale (BFAS) (DeYoung et al., 2007), which captures the big five domains of personality and two distinct aspects for each domain. The BFAS had excellent Cronbach's α reliability for Neuroticism (α = .912), and good reliability for Agreeableness (α = .815), Conscientiousness (α = .850), Extraversion (α = .884), and Openness/Intellect (α = .805). All aspects had acceptable to excellent reliability; Neuroticism: Volatility (α = .902), Neuroticism: Withdrawal (α = .838), Agreeableness: Compassion (α = .809), Agreeableness: Politeness (α = .719), Conscientiousness: Industriousness (α = .834), Conscientiousness: Orderliness (α = .770), Extraversion: Enthusiasm (α = .826), Extraversion: Assertiveness (α = .826), Openness/Intellect: Intellect (α = .805), Openness/Intellect: Openness (α = .709).

Academic Performance. Academic performance was computed as the principal component extracted from a Principal Components Analysis of first year average mark and performance in two first year introductory psychology courses.

1.3. Procedure

Phase 1: Participants firstly completed a short demographics questionnaire, the non-cognitive measures, and the ICAR. Following this, they were given detailed instructions on the EMA component, including information on how and when the pulses would be sent, and a short walkthrough of the microworld task. Participants then completed a 15-trial practice block on variant 1 of the microworlds task.

Phase 2: Using the EMA method, participants were sent 12 microworld pulse sessions over a 3-week period by email and given six hours to complete. Pulses were sent Monday to Friday between 10am and 3pm. In each session, participants indicated what they were doing immediately before the session and provide an indication of their current mood, neither of which are analysed in this paper. Participants were recruited in groups on a weekly basis, and the order in which variants were sent was randomised between groups. At the conclusion of the EMA component, participants were sent a short feedback survey to gauge their experiences of completing the study, the results of which are not the focus of this paper.

1.4. Analysis Approach

Data analysis was conducted using RStudio (2020) and JASP (2023). Figures were generated using ggplot2 (Wickham, 2016) and sjPlot (Lüdecke, 2023). Trial-level penalty score was used as the dependent variable in our analyses and was log transformed to attenuate a right-tailed skew due to the fixed lower bound (0 species) and lack of upper bound in the number of species that could exist on the island (i.e., the task design of the microworld).

The proposed parameters in Chapter 3 were computed as follows. Trial-level penalty score was used to calculate all within-person parameter estimates except for the best and worst scores. This approach differed to Study 1 because microworld tasks demand a strategic problem-solving capacity that is manifest across trial-on-trial performance. Unlike reasoning tasks where items are functionally independent, on microworld tasks we expect participants to leverage various strategies to control the microworld and that this will be reflected in trial-on-trial penalty score dependencies. Best score was calculated as the mean performance on the block with the lowest average penalty score and worst score was calculated as the mean on the block with the highest average penalty score.

For contingency parameters we used trial-level penalty scores as the basis for models run using the lme4 package in RStudio (Bates et al., 2015). The basic multilevel model for extracting experience and difficulty contingencies was similar to that used in Study 1. In Study 2, difficulty manipulations, outflow and delay, were contrast coded and included as random-effects. Fixed effects for the interaction between trial and block and the interaction between outflow and delay were also included. Trial order and block order were centered. As for Study 1, we included both experience and difficulty contingencies in the same multilevel model to control for the effects of one contingency on the other. This results in the below model:

Level 1:

$$Y_{ij} = \pi_0 + \pi_1 * trial + \pi_2 * block + \pi_3 * outflow + \pi_4 * delay + \pi_5 * trial * block + \pi_6 * outflow * delay + e_{ij}$$

Level 2:

$$\pi_{0j} = \beta_{00} + r_{0j}$$

$$\pi_{1j} = \beta_{10} + r_{1j}$$

$$\pi_{2j} = \beta_{20} + r_{2j}$$

$$\pi_{3j} = \beta_{30} + r_{3j}$$

$$\pi_{4j} = \beta_{40} + r_{4j}$$

$$\pi_{5j} = \beta_{50}$$

$$\pi_{6j} = \beta_{60}$$

Where i = trial (1-15) and j = person (j to N); block = block-order (1-12); delay (no-delay = -.5, delay = .5); outflow (constant = -.5, random = .5)

The random slopes were then extracted for each individual and used as estimates of within-person variation in within-block experience trajectories (π_{1j}), between-block experience trajectories (π_{2j}), outflow trajectories (π_{3j}), and delay trajectories (π_{4j})[8].

2. Results

2.1. Participant Engagement with the EMA Approach

We examined engagement across the full 101 participant sample as in Study 1. The mean number of tasks completed was 9.18 tasks (76.5%, SD = 3.47 tasks) and the median was 10 tasks. This is in line with other EMA studies which have a 65-90% task completion rate (Csikszentmihalyi & Larson, 2014; Fleeson, 2007; Rauthmann et al., 2016). For tasks that were completed, the mean delay between receiving and completing the task was 3.43 hours (SD = 9.93 hours) and the median delay was 0.99 hours. Unlike Study 1, there was a significant correlation between number of tasks completed and mean completion delay such that the more tasks a participant completed, the lower their average completion delay (r = -.251, p < .05).

As we collected individual differences measures in this study, we also tested the relationship between engagement and individual differences variables. The results are shown in Tables 5.2 and 5.3. There was a significant positive relationship between Conscientiousness and number of tasks completed (r = .241, p < .05) and proportion of tasks completed within the allotted six hour time window (r = .204, p < .05). The personality aspect correlations in Table 7 suggest completing more tasks is driven by both aspects of Conscientiousness but completing tasks more rapidly is primarily driven by Conscientiousness-Orderliness. Several non-Conscientiousness personality aspects were also correlated with various engagement metrics. Neuroticism-Withdrawal was negatively correlated with mean completion delay (r = -.242, p < .05). Agreeableness-Compassion was positively correlated with number of tasks completed (r = .219, p < . 05) and proportion of tasks completed within the allotted six hours (r = .225, p < . 05). Extraversion-Enthusiasm was positively correlated with number of tasks completed (r = .196, p < .05).

We also obtained feedback from participants to understand their experiences completing the study. Feedback data was missing for 19 participants who did not complete the minimum number of tasks to meet the course credit requirements. Feedback results are

---

[8] This was computed using lme4 with the function: lmer(penalty ~ 1 + trial + block + outflow + delay + trial*block + outflow*delay + (1 + item + block + outflow + delay | subject)).

summarised in Table 5.4. Participants reported positive experiences with the study interface and technical experience, as well as the time given to complete each task and the time spent on each task.

Table 5.2. Correlations between individual differences variables and completion metrics (N=101).

	Age	E	A	C	O	N	NFC	LGO	AGO	PGO
Number of tasks completed	.107	.194	.152	**.241***	.013	-.122	-.034	.028	.018	.089
Mean completion delay	-.068	.121	.005	.044	-.098	-.180	-.006	-.004	-.115	-.143
Proportion completed within 6 hours	.126	.181	.136	**.204***	.051	-.070	-.017	.035	.047	.128

*Correlation is significant at the 0.05 level (2-tailed).

Key: E = Extraversion, A = Agreeableness, C = Conscientiousness, O = Openness/Intellect, N = Neuroticism, NFC = Need for Cognition, GMA = General Mental Ability, LGO = Learning Goal Orientation, AGO = Perform-Avoid Goal Orientation, PGO = Perform-Prove Goal Orientation.

Table 5.3. Correlations between BFAS facets and completion metrics (N=101).

	N-VOL	N-WDR	A-COMP	A-POL	C-IND	C-ORD	E-ENTH	E-ASRT	O-INT	O-OPE
Number of tasks completed	-.106	-.114	**.219***	.034	**.211***	**.197***	**.196***	.152	.027	-.009
Mean completion delay	-.089	**-.242***	-.007	.015	.146	-.084	.086	.128	-.079	-.078
Proportion completed within 6 hours	-.084	-.040	**.225***	.001	.147	**.202***	.182	.143	.067	.012

*Correlation is significant at the 0.05 level (2-tailed).

Key: N-VOL = Neuroticism-Volatility, N-WDR = Neuroticism-Withdrawal, A-COMP = Agreeableness-Compassion, A-POL = Agreeableness-Politeness, C-IND = Conscientiousness-Industriousness, C-ORD = Conscientiousness-Orderliness, E-ENTH = Extraversion-Enthusiasm, E-ASRT = Extraversion-Assertiveness, O-INT = Openness/Intellect-Intellect, O-OPE = Openness/Intellect-Openness.

Table 5.4. Proportion of participants who somewhat agreed, agreed, or strongly agreed with statements in Study 2 (N=82).

Statement	Study 2
"I was given enough time to complete each task"	.854
"The time I spent on each task was reasonable"	.915
"The interface of the task was easy to use"	.890
"I was able to complete this study without any technical difficulties"	.866

2.2. Preliminary Analysis

Following the procedure used by Birney et al. (2017), analyses were conducted on the combined variant data to test the relationship between task experience and task difficulty. There was a significant main effect of trial order on performance (β = 0.11, 95% CI = 0.07 – 0.14, p < .001). While there was no main effect of block order on performance (β = 0.01, 95% CI = -0.02 – 0.04, p = .615), there was a significant interaction between trial order and block order (β = 0.02, 95% CI = 0.00 – 0.03, p = .008). There were significant main effects of both outflow (β = 0.07, 95% CI = 0.04 – 0.11, p < .001) and delay (β = 0.48, 95% CI = 0.41 – 0.55, p < .001) on performance, and a significant interaction between outflow and delay (β = -0.07, 95% CI = -0.08 – -0.06, p < .001). This model explained 48.7% of the variation in microworlds performance.

2.3. Predictive Utility of Cognitive Dynamics

Correlations between the cognitive dynamics parameters and academic performance are displayed in Table 5.5. The point of comparison was with the single session cognitive assessment (ICAR-16), which showed a moderate positive correlation of *r* = .287 with academic performance. Full intercorrelations between the cognitive dynamics parameters are contained in the supplementary materials.

As in Study 1, we computed a series of hierarchical regression models with academic performance as the dependent variable, the single session cognitive assessment and number of EMA pulses completed entered as independent variables in the first block, and the cognitive dynamics parameter entered as an independent variable in the second block. We included number of EMA pulses completed in the first block as we wanted to control for engagement with the EMA component so that the interpretation of the cognitive dynamics parameters is over and above individual differences in engagement. The results of these regressions are shown in Table 5.6.

Table 5.5. Correlation between cognitive dynamics and academic performance.

Parameter	Combined	Variant 1 (O,D)	Variant 2 (O^+,D)	Variant 3 (O,D^+)	Variant 4 (O^+,D^+)
Central Tendency Parameters					
Mean	-.203	**-.405**	-.252	.108	.046
Median	-.260	**-.417**	**-.297**	.009	-.070
Dispersion Parameters					
SD	.259	.061	.007	.146	.110
MSSD	.194	.046	.142	.180	.169
Skewness	.226	**.383**	.091	.202	**.317**
Best Run	**-.335**	**-.380**	-.224	.092	.024
Worst Run	.119	**-.366**	-.222	.136	.048
Contingency Parameters					
Trial Slope	-.016				
Block Slope	-.124				
Outflow Slope	.193				
Delay Slope	**.331**				
Pre-Change Point Slope	-.167	**-.288**	-.202	.027	.126
Post-Change Point Slope	.210	**.324**	.173	.176	-.065
Change Point	.110	.084	.125	.059	-.098

Note. O = constant outflow; O^+ = random outflow; D = no delay; D^+ = 3 trial delay

Table 5.6. Additional predictive utility of cognitive dynamics.

	Combined		Variant 1 (O,D)		Variant 2 (O^+,D)		Variant 3 (O,D^+)		Variant 4 (O^+,D^+)	
Parameter	$R^2\Delta$	p(FΔ)	$R^2\Delta$	p(FΔ)	$R^2\Delta$	p(FΔ)	$R^2\Delta$	p(FΔ)	$R^2\Delta$	p(FΔ)
Central Tendency Parameters										
Mean	.013	.286	**.096**	**.003**	.021	.174	.010	.359	.002	.693
Median	.029	.110	**.102**	**.002**	.029	.106	.000	.867	.001	.762
Dispersion Parameters										
SD	.037	.067	.000	.984	.001	.815	.019	.198	.007	.430
MSSD	.031	.094	.002	.642	.007	.449	.029	.108	.019	.201
Skewness	.037	.069	**.081**	**.006**	.000	.929	.017	.223	**.077**	**.008**
Best Run	**.046**	**.042**	**.077**	**.008**	.009	.374	.007	.436	.001	.765
Worst Run	.010	.337	**.076**	**.009**	.020	.182	.013	.292	.001	.745
Contingency Parameters										
Trial Slope	.005	.523								
Block Slope	.035	.079								
Outflow Slope	.019	.190								
Delay Slope	**.058**	**.022**								
Pre-Change Point Slope	.010	.343	**.046**	**.043**	.023	.157	.000	.878	.013	.289
Post-Change Point Slope	.037	.072	.051	.053	.018	.250	.034	.093	.000	.986
Change Point	.001	.776	.002	.678	.007	.492	.004	.563	.017	.233

Note. Compared against a base model including single session cognitive assessment score as a predictor and number of microworlds tasks completed as a covariate, which had an $R^2 = .138$. O = constant outflow; O^+ = random outflow; D = no delay; D^+ = 3 trial delay

To account for potential multicollinearity of the cognitive dynamics parameters, we ran a further OLS stepwise regression model using the same approach as for Study 1. To simplify the number of models, we examined the cognitive dynamics parameters from the combined data and variant 1 data only as these showed the strongest relationships to academic performance. In the first model using combined data, academic performance as the dependent variable, single session cognitive assessment score entered as an independent variable in the first block, and all 14 cognitive dynamics parameters from Table 5.5 (Combined column) entered as independent variables in the second block. Stepwise estimation was used for the second block to determine which cognitive dynamics parameter or parameters were most prominently predictive of academic performance. Delay slope was the only parameter in the second block that was a significant predictor of academic performance over and above single session cognitive assessment score ($\beta = 0.249$, $R^2_{model} = .146$, $R^2\Delta = .053$, $p(F\Delta) = .036$). We then repeated this model examining variant 1 data using the 10 cognitive dynamics parameters from Table 5.5 (Variant 1 column) for the second block. Median penalty score was the only parameter in the second block that was a significant predictor of academic performance above and beyond single session cognitive assessment score ($\beta = -0.388$, $R^2_{model} = .204$, $R^2\Delta = .104$, $p(F\Delta) = .006$).

2.4. Mediation Analysis

Finally, we conducted a series of mediation analyses replicating Study 1 (see Figure 4.2 for model representation). The only change from Study 1 was that we included number of pulses completed as a background confounder to control for individual differences in engagement. We selected only cognitive dynamics parameters that predicted additional variation in academic performance over and above the single session cognitive assessment. The results of these models can be found in Tables 5.7 and 5.8. Except for skewness of variant 4 and pre-change point slope of variant 1, all models tested demonstrated that the cognitive dynamics parameter fully and significantly mediated the relationship between single session cognitive assessment score and academic performance and that this relationship was unidirectional. These findings echo those of Study 1 and imply that the cognitive dynamics parameters are accounting for a broader construct than traditional single session assessments of cognitive ability. This broader construct is what is important for ecological cognitive performance in an academically demanding university context, rather than what is captured by cognitive ability as traditionally conceptualised and assessed.

Table 5.7. Single session score predicting academic performance, mediated by cognitive dynamics parameters.

	Combined						Variant 1 (O,D)						Variant 4 (O^+,D^+)			
	Direct Effect		Indirect Effect		Significance		Direct Effect		Indirect Effect		Significance		Direct Effect		Indirect Effect	
Parameter	Estimate	p	Estimate	p	Sobel	p	Estimate	p	Estimate	p	Sobel	p	Estimate	p	Estimate	p
Central Tendency Parameters																
Mean	-	-	-	-	-	-	.018	.536	**.050**	**.006**	3.20	**.001**	-	-	-	-
Median	-	-	-	-	-	-	.015	.609	**.054**	**.004**	3.30	**<.001**	-	-	-	-
Dispersion Parameters																
Skewness	-	-	-	-	-	-	.043	.108	**.026**	**.033**	2.40	**.016**	**.062**	**.015**	.007	.412
Best Run	.037	.214	**.032**	**.049**	2.67	**.007**	.033	.242	**.036**	**.018**	2.81	**.005**	-	-	-	-
Worst Run	-	-	-	-	-	-	.025	.407	**.044**	**.013**	2.95	**.003**	-	-	-	-
Contingency Parameters																
Delay Slope	.044	.112	**.025**	**.046**	2.35	**.019**	-	-	-	-	-	-	-	-	-	-
Pre-Change Point Slope	-	-	-	-	-	-	.191	.076	.078	.076	-	-	-	-	-	-

Note. Number of tasks completed included as a background confounder.

Table 5.8. Cognitive dynamics parameters predicting academic performance, mediated by single session score.

	Combined				Variant 1 (O,D)			
	Direct Effect		Indirect Effect		Direct Effect		Indirect Effect	
Parameter	Estimate	p	Estimate	p	Estimate	p	Estimate	p
Central Tendency Parameters								
Mean	-	-	-	-	**-.267**	**.002**	-.028	.539
Median	-	-	-	-	**-.232**	**.001**	-.020	.610
Dispersion Parameters								
Skewness	-	-	-	-	**.285**	**.004**	.053	.152
Best Run	**-.200**	**.034**	-.059	.227	**-.221**	**.005**	-.042	.257
Worst Run	-	-	-	-	**-.213**	**.006**	-.034	.412
Contingency Parameters								
Delay Slope	**.190**	**.017**	.047	.146	-	-	-	-
Pre-Change Point Slope	-	-	-	-	**-.227**	**.035**	-.066	.119

Note. Number of tasks completed included as a background confounder.

3. Study 2 Discussion

As for Study 1, our goal was driven by the theory that cognitive dynamics parameters provide a more symmetrical conceptualisation and operationalisation of an important ecological cognitive performance outcome, academic performance. In Study 2, we administered a different type of cognitive task to understand the generalisability of the effects observed in Study 1. Overall, a range of cognitive dynamics parameters captured variation in academic performance beyond what is captured by a single occasion test of cognitive ability. In addition, participants engaged with the EMA approach at similar rates to EMA studies in related fields and reported positive experiences with the task interface and study method. However, our results also differed in some ways from Study 1.

Similar to Study 1, we found evidence for the incremental predictive utility for parameters derived using EMA over several weeks reflecting central tendency (mean and median), dispersion (skewness, best run, and worst run), and contingencies (delay slope and pre-change point slope). Although cognitive ability as traditionally assessed was moderately correlated with academic performance, this relationship was fully and significantly mediated by cognitive dynamics parameters in a unidirectional way. Once again, we interpret these findings as evidence for the usefulness of cognitive dynamics in capturing a capacity for time and context-variant ecological cognitive performance that is critical for success in cognitively demanding real-world environments.

Most of the incrementally predictive cognitive dynamics parameters were derived from the least complex variant of the microworlds task, the exceptions being best run overall (in addition to best run on the easiest variant), skewness in the most complex variant (which did not capture something distinct from cognitive ability as traditional assessed in our mediation analyses), and delay slope. We note that best run overall is likely to reflect best run for the easiest variant for many participants as this was the simplest variant of the task. This suggests that between-person differences in within-person variation on microworld tasks may only emerge when the rules governing the microworld can easily be deduced. It is important to note, relative to Birney et al. (2018), our task only presented 15 trials, which may not have been enough for between-person differences in within-person variation to emerge, especially in the three-trial delay variant.

VI. GENERAL DISCUSSION OF STUDY 1 AND STUDY 2 FINDINGS

1. General Discussion

In Chapter 3, we proposed a series of cognitive dynamics parameters to capture ecological cognitive performance. In Chapters 4 and 5, we implemented these cognitive dynamics across two empirical studies. We found that a range of these parameters were predictive of university performance over and above general cognitive ability as it is traditionally operationalised and assessed. We also found that participants engaged with the EMA at similar rates to EMA studies on personality and affective dynamics, even in this low stakes setting, and that the EMA was acceptable to participants from a technical and practical standpoint. In this chapter, we synthesise and discuss these findings in detail.

To our knowledge these are the first studies to a) develop and test a set of parameters that are grounded in a firm theoretical conceptualisation of ecological cognitive performance, b) understand whether they have predictive utility beyond traditional single session cognitive ability measures, and c) examine whether the predictive utility is greater for any proposed class of parameters. That additional predictive utility was observed in Study 1 where the sole modification was the method of administration (i.e., the quantity and format of the items remained constant across both the EMA and traditional single session assessment) is particularly noteworthy. It suggests that cognitive dynamics parameters representing ecological cognitive performance tell us something important about an individual's application of cognitive ability in the real world that we are not currently capturing in traditional single occasion assessments of cognitive ability. We propose that cognitive dynamics capture an individual's ecological cognitive performance across time and contexts, and it is that within-individual variability across time and contexts that is currently missing from traditional conceptualisations of cognitive ability and methods of assessment.

1.1. Participant Engagement with the EMA Approach

Our results suggest that when the duration and spacing of cognitive EMA are reasonable, participants will complete a similar proportion of tasks within the allotted time windows as for other EMA studies. In personality and affect EMA studies, participants are usually instructed to respond to tasks within an hour and research suggests that around 65-90% of tasks are completed within an hour of the task being sent (Csikszentmihalyi &

Larson, 2014; Fleeson, 2007; Rauthmann et al., 2016)[9]. In our studies, participants were given longer to complete each task than in personality and affect EMA studies. This methodological decision was made because cognitive tasks are substantially more demanding than personality and affect measures as they require disengagement from whatever the participant is doing when the task is received and reallocation of cognitive resources to the cognitive task. In both studies the task completion rates (80.4% in Study 1 and 76.5% in Study 2) were in line with other EMA studies.

Our results from Study 2 also provided some insights into the individual differences related to engagement with cognitive EMA. Several traits and facets predicted engagement with the study, particularly number of tasks completed and adherence to the study requirements (to complete a minimum number of tasks): Conscientiousness, Conscientiousness-Industriousness, Conscientiousness-Orderliness, Agreeableness-Compassion, and Extraversion-Enthusiasm. Conscientiousness is likely to drive engagement with EMA, as it drives individuals to diligently and responsibly complete the tasks as specified (DeYoung et al., 2007). Agreeableness-Compassion and Extraversion-Enthusiasm may have had a similar facilitating role because both traits reflect a tendency towards positive and active engagement with the world around them (DeYoung et al., 2007). This finding is interesting from an applied perspective, as it suggests that EMA may capture not just individuals with the cognitive ability to flexibly and adaptably respond to cognitive challenges but may also lend itself to filtering out individuals lower in motivational and engagement traits necessary for educational and occupational success.

Overall, participants in Study 2 reported more favourable experiences with the study than participants in Study 1. This was contrary to our expectations, as participants in Study 1 were given more opportunities and longer to complete the tasks, and the type of task should be relatively familiar for university students. We have several hypotheses as to why this was the case. Firstly, the different nature of the microworlds and reasoning tasks may be a contributing factor. In the microworlds tasks, participants could see their performance and (theoretically) self-correct if deviating from the goal state. The microworlds task was also presented in a more engaging format, with multiple graphs and changes in the interface showing progress with each decision. No performance information was available in the reasoning task and items were presented in static, black and white formats. Moreover, the

---

[9] Substantial variability has been reported in response times, with the three cited studies showing quite different response rates. However, both of our studies fell comfortably within the cited boundaries.

reasoning tasks may have been more difficult or frustrating for participants than the microworlds tasks. Despite being multiple choice, if participants did not know the answer they would have to guess. Conversely, in the microworlds task, participants were given information that could have enabled them to adopt a good control strategy, or at least a good enough control strategy to ensure the system did not spiral out of control. Finally, the microworlds task was likely relatively novel for university student participants, whereas the reasoning task was likely to be somewhat familiar – either through exposure in undergraduate psychology courses or exposure in previous selection and assessment scenarios. While we anticipated this exposure may make participants view this study more favourably than Study 2, it may have worked in the opposite direction if participants were more familiar with the reasoning tasks as a measure of cognitive ability, which may have affected their engagement and motivation. This reasoning may also explain why participants found the time allowance less reasonable in Study 1 despite only having 9 multiple choice decisions per task in Study 1 compared to 15 theoretically infinite scale decisions to make per task in Study 2.

Despite these reports, participants generally encountered few technical difficulties and reported the interface as being easy to use. This finding is of note because mass psychometric testing is common in graduate recruitment of which university students make up the bulk of applicants, at least in a white-collar context (Carless, 2007, 2009). When recruiting graduates, applicants usually cannot show work history to demonstrate their suitability, and efficient methods are required to narrow the applicant pool (Carless, 2007, 2009). Our findings suggest that there is willingness to complete an EMA cognitive assessment from a university student sample.

1.2. Central Tendency Parameters

Overall, the cognitive dynamics parameters that displayed the strongest and most consistent effects were measures of central tendency: the mean and median of within-person performance collected over time. The central tendency of EMA pulse scores represents an individual’s typical performance level over a short period of time and are better predictors of academic performance than a traditionally administered single session cognitive assessment. This is because they capture a type of cognitive process important for ecological cognitive performance that traditional notions of cognitive ability do not.

Comparing these EMA-derived central tendency parameters between individuals provides information about between-person differences in how individuals generally apply their cognitive ability across time and in a variety of contexts with different task demands and situational characteristics. This information is not captured by traditional single session assessments often collected under static, high stakes conditions (von Stumm et al., 2011). Academic performance is typically assessed across an extended period of time and a range of tasks to provide a comprehensive picture of an individual’s knowledge, skills, and development over time, reflecting not just isolated instances of success or failure but an extended period of performance under varied conditions. Likewise, in a workplace context, performance is often evaluated through a diverse set of metrics including individual task completion, project outcomes, teamwork, leadership skills, and problem-solving abilities over an extended period of dynamically changing situations. Both of these contexts demand a sustained level of performance and adaptability across time and contexts, which is not reflected in single occasion cognitive assessments but might be reflected in EMA-derived central tendency parameters.

1.3. Dispersion Parameters

Previous studies have found between-person variation in ecological cognitive performance on complex cognitive tasks using various distributional parameters (e.g., Allaire & Marsiske, 2005; Salthouse, 2007, 2012), however, these studies did not investigate the implications of this variation for educational or workplace performance. Some of our proposed cognitive dynamics dispersion parameters were consistent incremental predictors of academic performance. Best and worst performance scores capture an individual's maximal and minimal performance, which reflects how individuals adaptively apply their cognitive ability under (presumably) the most and least ideal conditions (Coyle, 2001, 2003a, 2003b). That such predictors captured were implicated in real-world cognitive performance makes sense; performing well at university requires performing well under a variety of conditions and being able to sustain a good level of best and worst performance is undoubtedly beneficial for success. It is plausible that such parameters may be even stronger predictors where success relies on extremely high and/or precise performance.

None of our other proposed distributional parameters (SD, MSSD, and Skewness) appeared to capture meaningful between-person differences in ecological cognitive performance. For Study 1, this may be due to the relatively small number of sampling occasions, which did not allow for between-person differences in these parameters to emerge. For Study 2, this may be because people generally displayed relatively similar patterns of performance within variants. For both studies, such parameters may also simply not be effective ways of capturing ecological cognitive performance. Regardless, replication with more measurement occasions is required to ascertain this result.

1.4. Contingency Parameters

Finally, we turn to our contingency parameters. The only contingency parameter that consistently incrementally predicted academic performance was difficulty. Difficulty contingencies capture an individual's capacity to deal with increased task complexity controlling for their experience with the task (Birney et al., 2017; Birney et al., 2018). Success at university requires the capacity to perform well when faced with learning increasingly complex concepts and skills. Given this, it is perhaps unsurprising that this capacity predicted academic performance above and beyond cognitive ability as traditionally assessed.

In Study 1, difficulty contingencies were based on a Rasch measurement model conceptualisation of difficulty in line with Birney et al. (2017). In Study 2, difficulty was manipulated in terms of outflow and delay within the microworld task system, and it was only the delay contingency that incrementally predicted variation in academic performance above and beyond cognitive ability as traditionally assessed. As argued by Birney et al. (2018), the delay manipulation is cognitively substantive in that keeping track of the likely impact of previous decisions places a working memory burden on current decision making capacity. Thus, it is not surprising that we see individuals' delay-performance contingency being predictive of academic performance, nor that it fully mediates the relationship between cognitive ability as traditionally conceived and academic performance, similar to Study 1. Importantly, this was not the case for the outflow manipulation. While penalty scores were higher in random outflow variants than constant variants, this source of difficulty was not related to academic performance. Birney et al. (2018) observed similar findings, with outflow-performance contingencies being predictive of performance goal orientations and not reasoning abilities.

That experience contingencies did not capture meaningful between-person differences in ecological cognitive performance in Study 1 is not entirely unexpected; on a reasoning task with nine independent items and only five measurement occasions people are not likely to exhibit significant learning effects. For Study 2, however, this finding was somewhat unexpected as microworld tasks demand participants learn the rules governing the system and apply them to control it (Birney et al., 2018). While we did observe overall within-task learning across participants, within-person experience contingencies were not significant incremental predictors of academic performance.

2. Limitations and Future Directions

While this research has advanced our understanding of ecological cognitive performance and its measurement, there are limitations to be addressed in future studies. First and foremost, replication is crucial to ascertain the robustness and generalisability of our findings. We recommend replication in more age and gender diverse samples (e.g., as in Salthouse, 2012), as well as in professional samples using workplace performance as the criterion outcome (e.g., as in Beckmann et al., 2020). We also recommend future studies should seek to increase the number of measurement occasions. While our findings are encouraging, the number of measurement occasions may not have been sufficient to fully capture the complexity and dynamic nature of within-person variation, especially in Study 1 where we only implemented five measurement occasions. Increasing the number of measurement occasions would also increase the confidence in the observed results for the contingency models derived from the multilevel modelling approach (Diallo et al., 2014). Despite this, that we see such compelling incremental effects for Study 1 with the small number of observations suggest further research on cognitive dynamics parameters derived from the EMA is worthwhile.

Furthermore, we recommend future researchers continue to explore task, situation, and person contingencies that impact between-person differences in ecological cognitive performance (Beckmann et al., 2020; Beckmann et al., 2021; Birney & Beckmann, 2022; Vaughan & Birney, 2023). In this study we examined only difficulty and experience contingencies. However, other contingencies are likely to uniquely and interactively influence within-person variation (Beckmann et al., 2021). These contingencies are likely to support further explanatory accounts for our results, that is, they may help us understand why ecological cognitive performance predicts real-world outcomes beyond single session assessments. For example, differences in situational characteristics may explain between-person differences in ecological cognitive performance such that some individuals are more resilient to environmental changes than others and these individuals also have a propensity to perform better at university and in the workplace. Developing such explanatory process accounts is important for understanding the nature of ecological cognitive performance and for developing assessments that faithfully capture it. Further, it aids in the interpretation of cognitive dynamics parameters in practical contexts because it allows for individuals to be more accurately assessed, selected, and trained in educational and organisational settings (e.g., maximising person-job fit via matching individual profiles to appropriate roles, identifying areas for training for individuals).

Finally, we recommend that future researchers consider manipulating the stakes of the situation in which the EMA tasks are conducted. Our participants were university students given no incentive to complete the tasks to the best of their ability. However, given what is known about how stakes of the situation affecting cognitive ability test performance (von Stumm et al., 2011), future studies are needed to understand the stability of our findings under different stake conditions.

3. Conclusion

In this set of studies we presented preliminary evidence that ecological assessment of cognitive ability predicts academic performance better than traditional assessments of cognitive ability. Cognitive ability as it is traditionally conceived of and operationalised is a point in time assessment that does not account for ecological variation in the application of cognitive ability across time and contexts. This is fundamentally asymmetrical with the criterion outcomes it is commonly used to predict, chiefly academic and workplace performance, which demand adaptability, flexibility, and ongoing engagement with cognitive challenges. Using the EMA we captured cognitive performance across several weeks and derived a set of cognitive dynamics parameters intended to capture between-person differences in ecological cognitive performance across time and contexts. Many of these parameters incrementally predicted university performance over and above performance on a traditional single session cognitive assessment, and our analysis suggested that this was because these parameters capture a substantive construct beyond cognitive ability as it is traditionally assessed. While our findings require replication, we are enthusiastic about their implications for the way in which we conceptualise and measure cognitive ability. Explanatory accounts of how and why individuals' ecological cognitive performance varies are needed to theoretically unravel the implications for theory development and the methods we use to study and assess cognitive ability. In the following chapter, we aim to begin building this explanatory process account by understanding how individuals' cognitive performance changes depending on their situation.

VII. STUDY 3: A SITUATION CONTINGENCY APPROACH TO ECOLOGICAL COGNITIVE ASSESSMENT

1. Introduction

In his seminal book, *The Abilities of Man*, Charles Spearman (1927) theorised that cognitive ability could be captured by a general 'g' factor, which is influenced by multiple underlying factors. One of these factors, which Spearman termed the 'oscillation factor', was purported to capture the extent to which an individual's cognitive performance varied over time and contexts. In the century that has followed, a significant body of research has supported Spearman's early theorising and shown that individuals' cognitive performance does vary over time and context, rather than being entirely stable (see e.g., Allaire & Marsiske, 2005; Bunce et al., 2004; Li et al., 2004; Lovden et al., 2007; MacDonald et al., 2003; MacDonald et al., 2009; Salthouse, 2007, 2012; Salthouse & Berish, 2005; Salthouse et al., 2006; Schmiedek et al., 2013). This has paradoxically been observed even though Spearman's oscillation factor has been long rejected by the field of cognitive ability research. Predominant theories of cognitive ability assume that people's ecological cognitive performance is relatively stable across time and contexts and, as a result, prevalent methods of cognitive assessment are conducted as single-session, context-invariant assessments from which performance scores are extracted and used to predict various outcomes (Birney et al., 2019; Cronbach, 1957). However, such an approach neglects the evidence that ecological cognitive performance varies across time and contexts.

In this final study, we aim to begin building an explanatory process account for how and why individuals' ecological cognitive performance varies as a follow up to our findings in the previous chapter. Undergraduate university student participants complete 10 measures of cognitive ability over three weeks administered via an Ecological Momentary Assessment (EMA) paradigm. At each measurement occasion, they complete questions about their current context. Further, participants complete measures of cognitive ability as we traditionally conceptualise and measure it (i.e., a single occasion, context-invariant assessment) and personality and motivational traits implicated in cognitive engagement. We derive within-person situation contingencies (i.e., situation-performance relationships) and use these to predict an important real-world outcome, university performance, over and above cognitive ability as traditionally conceptualised and personality and motivational traits associated with cognitive engagement. In doing so, we demonstrate that ecological cognitive performance variation appears to be an adaptive process that facilitates resilience to contextual changes. This adaptive process is critical for success in cognitively demanding real-world environments, such as university, and represents a distinct aspect of cognitive ability not currently accounted for by predominant theories and assessment methods.

2. Ecological Cognitive Performance as an Adaptive Contingency

Across Study 1 and Study 2, we found preliminary evidence that ecological assessment of cognitive ability predicts an important real-world outcome, university performance, better than traditional assessments of cognitive ability. We proposed that this is because there is a fundamental asymmetry between traditional assessment methods and analysis approaches and the criterion outcomes they are used to predict. The traditional conceptualisation and measurement of cognitive ability assesses individuals at a single point in time under a static context, whereas criterion outcomes like academic and professional success demand adaptability and flexibility in response to contextual changes over time. The EMA facilitates an understanding of how and why individuals' performance changes over time in response to changes in context because it takes short-term repeated measures of cognitive performance under different contexts. This short-term repeated measures assessment method yields rich within-person performance data that can be used to derive ecological within-person cognitive performance metrics. However, while our previous work showed that these metrics are predictive of university performance over and above a single occasion context-invariant assessment of cognitive ability, further evidence is required to establish that this is a systematic and meaningful process.

For ecological cognitive performance variation to be considered systematic and meaningful, it must be shown to be systematically contingent on a situation, person, or task factor and serve an adaptive function in the environment. Drawing from prior work in personality research, Birney and Beckmann (2022) term this an *adaptive contingency*. Mathematically, an adaptive contingency is quantified as the slope of an individual's regression line with performance as the dependent variable and situation, person, or task factors as the independent variable(s). They represent the conditional within-person relationship between relevant factors and cognitive performance, and may vary in direction and magnitude. Between-person variation in the direction and magnitude of these contingencies captures between-person differences in individuals' flexibility, or lack thereof, in dealing with the range of changes we encounter on a day-to-day basis. These adaptive contingencies therefore allow us to build a process account of why individuals differ in their performance over relatively short time periods and translate this process account to our theories of cognitive ability and our methods of cognitive assessment.

Adaptive contingencies can be derived from a range of factors. Situation factors are features of the context in which the task is completed. Person factors are trait and transient factors such as cognitive ability, personality, motivational dispositions, neuromodulatory processes, stress, and motivation. Finally, task factors are aspects of task demands and design that can change both between and within tasks. These situation, person, and task factors will individually and interactively shape an individual’s ecological cognitive performance variation. The EMA approach allows for a comprehensive understanding of the unique and interactive role of these factors by allowing us to keep certain factors constant while observing how individuals respond to changes in other factors across repeated administrations.

2.1. The Role of the Situation in Cognitive Performance

In this study, we focus on the role of the situation. We administer cognitive tasks of the same task design and difficulty to the same group of participants on 10 measurement occasions over two to three weeks and collect data on the situation in which each task is completed. Consequently, we can examine the role of the situation as an adaptive contingency factor in the absence of changes in task factors or person factors, at least to a certain extent. We choose to focus on situation factors in this study because of their relevance in an applied context. As process-based occupations are replaced by automation and artificial intelligence, higher education and professional occupations are becoming increasingly reliant on critical thinking, problem-solving, and interpersonal skills relevant for dealing with ambiguity and complexity in our rapidly changing environment. Situation contingencies may therefore serve an adaptive function in our environment by facilitating adaptivity and flexibility in response to the vast array of contextual changes we encounter on a daily basis in our education and work.

Various measures have been developed to capture situational features that systematically affect behaviour. One of the more recent measures is the DIAMONDS scale (Rauthmann et al., 2014; Rauthmann & Sherman, 2016; Rauthmann et al., 2015), which captures eight situational features: Duty, Intellect, Adversity, Mating, Positivity, Negativity, Deception, and Sociality. Research has shown substantial within-person variation in self-reported experiences of DIAMONDS features over several weeks (Rauthmann et al., 2014; Rauthmann & Sherman, 2016; Rauthmann et al., 2015). As a result, the DIAMONDS may be a useful tool for understanding how variation in contexts impacts ecological cognitive performance. However, not all of the DIAMONDS features may be theoretically linked to cognitive performance. Therefore, the situation factors we capture are adapted from the DIAMONDS with additional theoretical consideration for other factors likely to influence ecological cognitive performance.

2.2. Capturing Situational Influences on Cognitive Performance

We retained items relevant to Duty, Intellect, Positivity, Negativity, and Sociality, as all of these features have established theoretical links with cognitive performance. Duty and Intellect[10] capture the extent to which a situation demands cognitive engagement. Duty captures an external appraisal of the need for cognitive engagement – that work has to be done – whereas Intellect captures an internal appraisal of the demands of this cognitive engagement – that deep thinking is required. Individuals differ in their appraisal of cognitive challenges and their motivation and reasons for engaging with cognitive challenges. High levels of traits such as Conscientiousness, Openness/Intellect, Need for Cognition, and Goal Orientation are robust predictors of both educational and workplace performance because they reflect a general propensity to enjoy and engage with cognitive activities in the face of challenges (Schmidt & Hunter, 1998; Steinmayr et al., 2011; Strobel et al., 2019).

Positivity and Negativity[11] are also likely to impact ecological cognitive performance because there is a demonstrated link between affect and cognitive performance (Harmon-Jones et al., 2013; Tyng et al., 2017; Yang et al., 2013). Positive affect is linked to improved cognitive performance (Yang et al., 2013) and broadened cognitive processing, which facilitates higher order cognitive processes such as problem-solving and creativity (Harmon-Jones et al., 2013). Conversely, negative affect is linked to impaired cognitive performance (Paino et al., 2018) and narrowed cognitive processing, which facilitates a focus on task completion suited to process-oriented tasks (Harmon-Jones et al., 2013).

Sociality[12] describes the extent to which social interaction is possible in the current situation. Social interaction may impact cognitive performance, particularly in a young adult sample where it may serve as a distraction that inhibits cognitive performance (Blasiman et al., 2018; Blatchford et al., 2003). Taken together, we expect that individuals who are relatively more resilient to fluctuations in their environment – be they Duty, Intellect,

---

[10] Captured by items "work has to be done" and "deep thinking is required" respectively.

[11] Captured by items "the situation is pleasant" and "the situation contains negative feelings (e.g., stress, anxiety, guilt, etc.)" respectively.

[12] Captured by item "social interactions are possible or required".

Positivity, Negativity, or Socialty – may be better adapted to perform well in cognitively demanding real-world environments that are dynamic and ever-changing.

However, Adversity, Mating, and Deception[13] may not be as influential because none have particularly apparent theoretical or empirical links with cognitive performance. As a result, we do not include these items. In place of the remaining three features, we included items describing Distractions, Environment, and Fatigue. The Distractions item captures the impact of non-social distractions on cognitive performance. Like Sociality, non-social distractors impact cognitive performance (Blasiman et al., 2018; Dontre, 2021; May & Elder, 2018). This may be exacerbated if the non-social distractors act as reward-based cues that detract from the task at hand, which young adults in particular have heightened sensitivity to (Cools, 2008; Geier et al., 2010; Prencipe et al., 2011). The Environment item captures the impact of the physical environment. This item was included because the comfort of the physical environment has been demonstrated to impact cognitive performance, including room temperature (Zhang et al., 2019), artificial lighting (Keis et al., 2014), and natural light (Ko et al., 2020). Finally, the Fatigue item captures the impact of physical and mental fatigue on cognitive performance, which has been empirically demonstrated to have a detrimental impact on cognitive performance (Schmiedek et al., 2010). Again, our expectation is that individuals who are relatively more resilient to fluctuations in these factors may be better adapted to perform well in cognitively demanding real-world environments.

This resulted in an 8-item measurement capturing Duty, Intellect, Positivity, Negativity, Sociality, Distractions, Environment, and Fatigue. These factors can be broadly segmented into two categories: extrinsic and intrinsic. Extrinsic situation factors are those factors which require an appraisal of an individual’s external environment and which an individual sometimes has relatively less control over (Duty, Sociality, Distractions, and Environment). Intrinsic situation factors are those factors which require an appraisal of an individual’s internal state and which an individual has relatively more control over, at least to a certain extent (Intellect, Positivity, Negativity, and Fatigue).

2.3 The Present Study

This work makes a novel contribution to the field of cognitive ability research by attempting, for the first time, to elucidate whether adaptive situation contingencies provide an explanatory process account of why ecological cognitive assessment predicts university

---

[13] Captured by items “somebody is being threatened, accused, or criticized”, “potential romantic partners are present”, and “somebody is being deceived” respectively.

performance over and above traditional cognitive assessment. Participants will complete 10 measures of cognitive ability across three weeks using an EMA paradigm and, at each measurement occasion, will provide information about the context in which they complete the cognitive ability measure. Participants will also complete measures of cognitive ability (a single-session context-invariant assessment), personality, and goal orientation, which are all implicated in between-person differences in cognitive performance. We will derive adaptive situation contingencies and examine whether these are systematic and meaningful predictors of real-world cognitive performance (in this study, university performance) over and above cognitive ability as we traditionally conceptualise and assess it, and traits implicated in cognitive engagement.

2.4 Hypotheses

We hypothesise that ecological cognitive performance variation is an adaptive process that equips individuals with the resilience to deal with environmental changes. We propose that this process is important for real-world cognitive performance over and above baseline cognitive ability as traditionally conceptualised and assessed, and motivational traits because it accounts for the role of contextual changes in cognitive performance, which context-invariant cognitive ability assessments do not. If our hypothesis is correct, individuals with relatively flat situation-performance contingencies will perform better at university controlling for cognitive ability and motivation, whereas individuals with relatively steeper situation-performance contingencies will perform worse at university controlling for cognitive ability and motivation. Empirical evidence suggests that extrinsic and intrinsic situation factors can significantly impact cognitive performance both positively and negatively. We hypothesise that people who exhibit greater capacity to regulate their cognitive performance in response to oscillating extrinsic or intrinsic cues are likely to be better equipped to cognitively perform under a variety of situations.

3. Methods

3.1. Participants

Participants were second-year undergraduate psychology students (n = 85, M_{age} = 21.31 years, SD_{age} = 4.13 years, N_{female} = 74). All participants received course credit in return for their participation (1 credit for 1 hour of time).

3.2 Materials

Reasoning Tasks. Reasoning items were sourced from the International Cognitive Ability Resource (ICAR, 2014), an open-access database of cognitive ability tests. We drew items from three tests in this database: *verbal reasoning*, *numerical reasoning*, and *abstract reasoning*. *Verbal reasoning* items required participants to read a paragraph of text and respond to a question about the text. *Numerical reasoning* items asked participants to examine numerical data in the form of graphs or tables and answer a question about the information presented. *Abstract reasoning* items required the participant to observe a sequence of figures and deduce the next figure in the sequence. All items presented five multiple choice options, with only one correct response per item.

From the item pool, 30 items were selected from each reasoning domain. These were the same 90 items used in Study 1 of the previous chapter and sample items can be found in Figure 4.1 in Chapter 4. For this study, these items were then pseudo-randomised across the EMA pulse tasks, with three items assigned from each domain to each pulse, resulting in a total of nine items per pulse. All participants received the pulses in the same randomised sequence, however, because some participants missed pulses when initially sent, not all pulses were completed in the intended order. The EMA pulse component had good Cronbach's alpha reliability of $\alpha = .88$. Reliability of between-person differences averaged over time was RkRn = .85 and reliability of within person variation averaged over items was Rcn = .02 (mlr function from psych R package: Revelle, 2023).

Situation Measures. At each EMA pulse, participants were asked to respond to an 8-item scale describing their current situation. This scale was adapted from the brief DIAMONDS scale (Rauthmann & Sherman, 2016), as described in Chapter 7, Section 1.2. Participants were asked to rate "to what extent do the following phrases apply to your current situation?" with responses on a 0-100 sliding scale. Situation items and associated descriptive statistics are contained in Table 7.1 and show that participants reported substantial within-person variation in the range of situations experienced during the testing period.

Table 7.1. Situation Feature Items (adapted from Rauthmann & Sherman, 2016).

Domain	Item	Overall mean rating (0-100)	Between-person SD	Average within-person SD	SD of within-person SD	Average within-person SD/between-person SD	Intraclass Correlation Coefficient (ICC)[14]
Duty	Work has to be done	62.33	29.89	19.01	9.52	.636	.498
Intellect	Deep thinking is required	45.25	29.75	19.49	8.88	.655	.486
Positivity	The situation is pleasant	55.69	24.40	17.63	7.27	.723	.393
Negativity	The situation contains negative feelings	29.77	25.09	16.90	8.66	.674	.431
Sociality	Social interactions are possible or required	34.21	28.85	21.97	10.57	.762	.290
Distractions	Non-social distractions are present	44.39	28.58	20.83	8.76	.729	.379
Environment	The physical environment is comfortable	65.96	22.91	17.16	8.10	.749	.318
Fatigue	I feel fatigued	48.27	28.91	20.67	7.42	.715	.426

---

[14] The ICC was derived from a series of unconditional multilevel models, one for each situation characteristic:
Level 1: $\text{situation}_{ij} \sim \pi_{0j} + e_{ij}$
Level 2: $\pi_{0j} = \beta_{00} + r_{0j}$
Where i = block (measurement occasion) and j = person
This was computed using lme4 with the function: lmer(situation ~ 1 + (1 | subject)).

Cognitive Ability. Cognitive ability as it is traditionally conceived of and assessed was measured using a 16-item version of the ICAR that is strongly correlated with longer measures of general cognitive ability (Dworak et al., 2021; Young & Keith, 2020). This test includes four items each for verbal reasoning, letter-number series, matrix reasoning, and 3D object rotation. The four domains were randomly presented to participants, with the four items in each domain also randomly presented to participants. The ICAR had acceptable Cronbach's α reliability of α = 75.

Goal Orientation. GO was measured with the 16-item scale from VandeWalle (1997). This measure had good Cronbach's α reliability of α = .81 for the Learning Goal Orientation (LGO) sub-scale, good reliability of α = .83 for the Performance-Avoid Goal Orientation (PGA) sub-scale, and acceptable reliability of α = .67 for the Performance-Prove Goal Orientation (PGP) sub-scale.

Personality. Personality was measured with a 100-item version of the BFAS (DeYoung et al., 2007), which generates scores for each of the big five domains of personality and two distinct aspects under each domain. This measure had good Cronbach's α reliability for all sub-scales (α = .88 for Neuroticism, α = .85 for Agreeableness, α = .87 for Conscientiousness, α = .89 for Extraversion, and α = .84 for Openness/Intellect). All aspects had acceptable to good reliability (Neuroticism – Volatility (α = .86), Neuroticism – Withdrawal (α = .82), Agreeableness – Compassion (α = .86), Agreeableness – Politeness (α = .71), Conscientiousness – Industriousness (α = .86), Conscientiousness – Orderliness (α = .80), Extraversion – Enthusiasm (α = .84), Extraversion – Assertiveness (α = .89), Openness/Intellect – Intellect (α = .85), Openness/Intellect – Openness (α = .76)).

Academic Performance. Academic performance was computed as the principal component extracted from a Principal Components Analysis of first year average mark and performance in two first year introductory psychology courses.

3.3. Procedure

On sign-up, participants completed a short demographics questionnaire, and the ICAR, GO, and BFAS. They were then given detailed instructions about the EMA pulse component, which began the Monday following completion of the demographics and individual differences measures. In the EMA pulse component participants were sent 15 reasoning tasks over three weeks and were informed they had to complete 10 tasks for the study. These tasks were delivered to the participants' email address and they were allowed a 24-hour window to complete each task. One task was sent each weekday (Monday to Friday) between 10 am and 3 pm throughout the three-week testing period. Once participants finished all 10 tasks, they ceased receiving additional tasks.

4. Results

4.1. Preliminary Analysis

Data analysis was conducted using RStudio (2020) and JASP (2023). Figures were generated using ggplot2 (Wickham, 2016) and sjPlot (Lüdecke, 2023). Full within-person performance trajectories are shown in Figure 7.1.

4.2. Replication of Prior Findings

Firstly, we replicated the analysis conducted in the previous chapter to ascertain the robustness of our prior findings regarding the predictive utility of ecological cognitive assessment. We extracted the same set of ecological cognitive performance parameters and examined their predictive utility compared to cognitive ability as traditionally assessed (i.e., performance on the 16-item ICAR) [15]. The correlation between academic performance and cognitive ability as traditionally assessed was $r = .116$ (p = .280). Correlations between the ecological cognitive performance parameters and academic performance are shown in Table 7.2 and demonstrate that a range of parameters were stronger numerical correlates with academic performance than cognitive ability as traditionally assessed. We then ran a series of hierarchical OLS regression models to examine whether these parameters predicted variation in academic performance over and above cognitive ability as traditionally assessed. The results of these regressions are shown in Table 7.3. All variables that were stronger numerical correlates of academic performance than cognitive ability as traditionally assessed also predicted significant incremental variation in academic performance, replicating the previous chapter's findings[16]. Full intercorrelations between within-person cognitive performance parameters and individual differences variables are contained in the supplementary materials.

To account for potential multicollinearity of the cognitive dynamics parameters, we ran two further OLS stepwise regression models using the same approach as for Study 1 and 2. In the first model, academic performance was the dependent variable, single session

---

[15] Central tendency and dispersion parameters were computed using the 10 pulse total scores. Contingency parameters were extracted from a multilevel model using item-level scores run in the lme4 package in R (Bates et al., 2015). Following Birney et al. (2017), item difficulty was based on Rasch calibrated item-difficulty estimates based on items within each pulse using the ltm package in R (Rizopoulos, 2006). For further details see the previous chapter.

[16] We do not replicate the mediation analyses from the previous as single session cognitive ability score was not a significant predictor of academic performance in this study, violating the assumptions of mediation analysis.

cognitive assessment score was entered as an independent variable in the first block, and all cognitive dynamics parameters from Table 7.2 were entered as independent variables in the second block. Stepwise estimation was used for the second block to determine which cognitive dynamics parameter or parameters were most prominently predictive of academic performance. Mean pulse score was the only parameter entered in the second block as a significant predictor of academic performance over and above single session cognitive assessment score ($\beta = 0.543$, $R^2_{model} = .192$, $R^2\Delta = .178$, $p(F\Delta) < .001$) (c.f., Study 1, in which median pulse score was the only parameter in the second block that was a significant predictor of academic performance over and above single session cognitive assessment score).

Given individual differences in personality and goal orientation on performance may also have influenced cognitive performance over time and contexts, we ran a second model including these variables in the first block. In the second model, academic performance was the dependent variable, single session cognitive assessment score, all five BFAS domains, and all three GO facets were entered as independent variables in the first block, and all cognitive dynamics parameters from Table 7.2 were entered as independent variables in the second block. Stepwise estimation was again used for the second block. Again, mean pulse score was the only parameter entered in the second block as a significant predictor of academic performance over and above single session cognitive assessment score ($\beta = 0.597$, $R^2_{model} = .278$, $R^2\Delta = .181$, $p(F\Delta) < .001$).

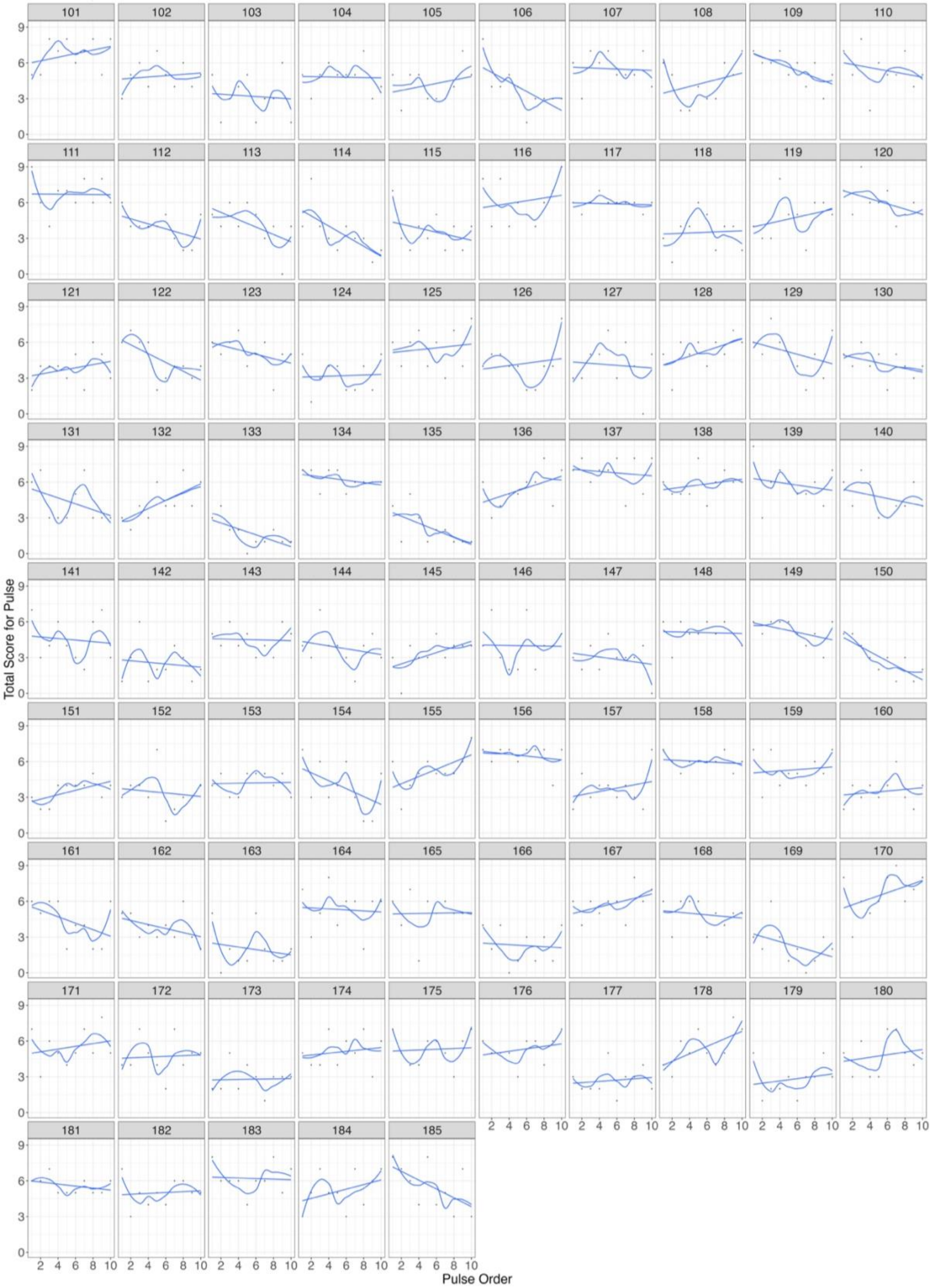


Figure 7.1. Within-person performance trajectories (with loess smoothing and linear model).

Table 7.2. Correlation between ecological cognitive performance parameters and academic performance.

	Academic Performance
Central Tendency Parameters	
Mean	**.402**
Median	**.360**
Dispersion Parameters	
SD	.040
Skewness	-.150
Best Score	**.372**
Worst Score	**.292**
Contingency Parameters	
Trial Slope	**.393**
Block Slope	.212
Difficulty Slope	**-.386**

Table 7.3. Additional predictive utility of ecological cognitive performance parameters. *Note.* Compared against a base model including single session score as a predictor, which had an R^2 = .014.

	R^2_{model}	$R^2\Delta$	p(FΔ)
Central Tendency Parameters			
Mean	**.192**	**.178**	**<.001**
Median	**.145**	**.131**	**<.001**
Dispersion Parameters			
SD	.017	.003	.631
Skewness	.028	.014	.273
Best Score	**.141**	**.127**	**<.001**
Worst Score	**.089**	**.074**	**.011**
Contingency Parameters			
Trial Slope	**.183**	**.168**	**<.001**
Block Slope	.048	.034	.091
Difficulty Slope	**.172**	**.158**	**<.001**

4.3. Adaptive Contingencies

We then turned to the primary purpose of this study: investigating whether ecological cognitive performance explains variation in academic performance beyond context-invariant cognitive assessments because it captures a systematic and meaningful adaptive process that facilitates resilience to contextual changes. We extracted situation contingencies from a multilevel model computed in the lme4 package in RStudio (Bates et al., 2015)[17]. The multilevel model used to derive contingencies was as follows:

Level 1:

$$\text{score}_{ij} \sim \pi_{0j} + \pi_{1j}.\text{Duty} + \pi_{2j}.\text{Intellect} + \pi_{3j}.\text{Positivity} + \pi_{4j}.\text{Negativity} + \pi_{5j}.\text{Sociality} + \pi_{6j}.\text{Distractions} + \pi_{7j}.\text{Environment} + \pi_{8j}.\text{Fatigue} + e_{ij}$$

Level 2:

$$\pi_{0j} = \beta_{00} + r_{0j};\ \pi_{1j} = \beta_{10} + r_{1j};\ \pi_{2j} = \beta_{20} + r_{2j}\ ;\ \pi_{3j} = \beta_{30} + r_{3j};\ ;\ \pi_{4j} = \beta_{40} + r_{4j};\ \pi_{5j} = \beta_{50} + r_{5j}\ ;\ \pi_{6j} = \beta_{60} + r_{6j};\ ;\ \pi_{7j} = \beta_{70} + r_{7j};\ \pi_{8j} = \beta_{80} + r_{8j}$$

Where i = block (pulse) and j = person.

Situation characteristic ratings were person-centred. The multilevel model included pulse total scores as the dependent variable and each person-centred situation characteristic rating simultaneously entered as fixed and random effects. We entered all situation characteristics simultaneously rather than computing separate models because situations do not occur in a vacuum. At any one time, people experience a range of different situational features which interactively shape their experience. Moreover, as our situation questions were answered at the same time, people's responses are likely to be interdependent. From this model we extracted each participant's random slopes for every situation characteristic and used these slopes as correlates and predictors in subsequent analysis. Within-person random slopes for each situation characteristic are shown in Figure 7.2.

---

[17] This was computed using lme4 with the function: lmer(score ~ 1 + Duty + Intellect + Positivity + Negativity + Sociality + Distractions + Environment + Fatigue + (1 + Duty + Intellect + Positivity + Negativity + Sociality + Distractions + Environment + Fatigue | subject)).

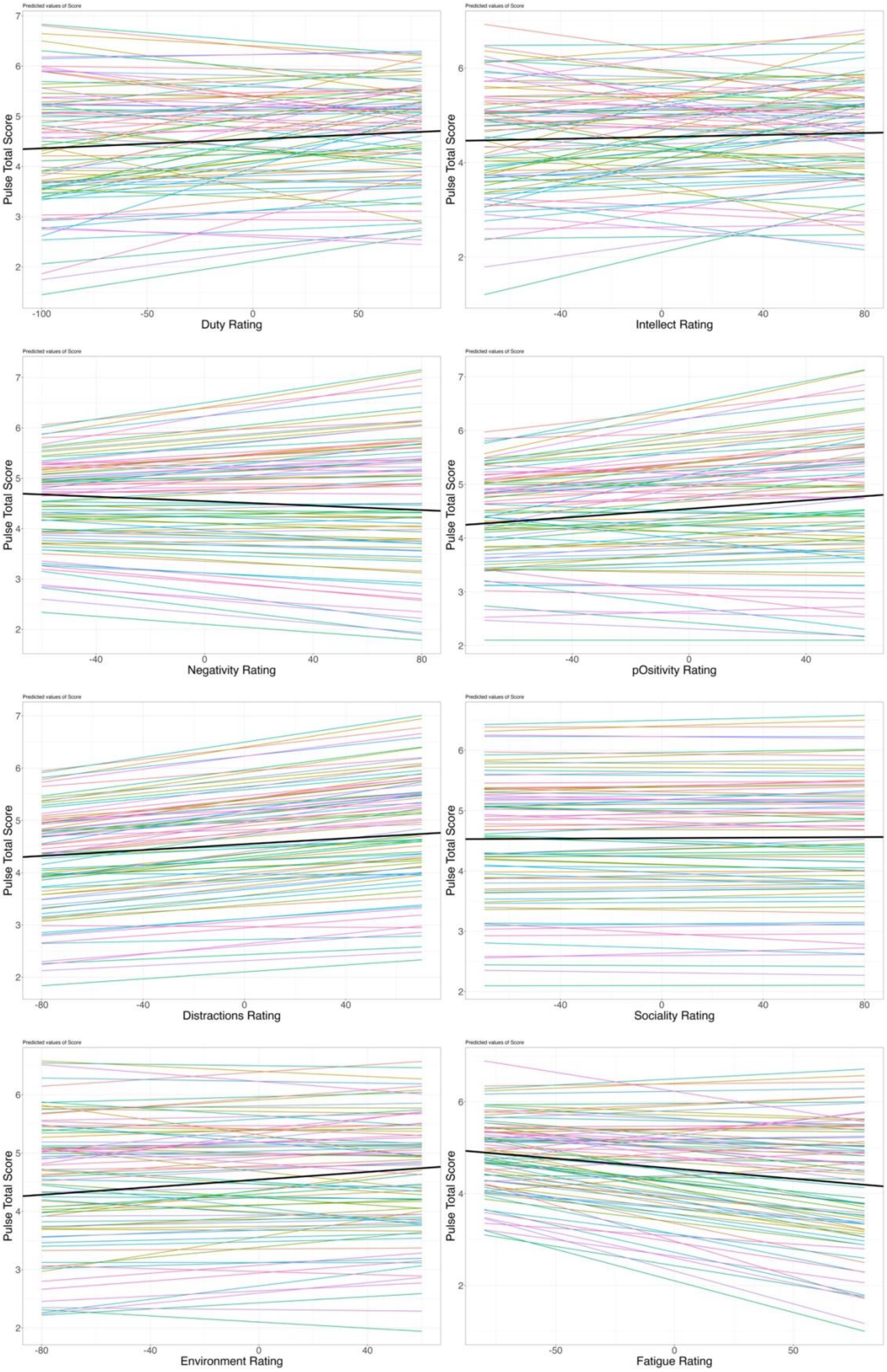


Figure 7.2. Within-person situation contingencies (situation ratings are person-centred).

Note: Fixed effects are indicated by the black line and random effects are indicated by the coloured lines.

Within-person contingencies for three characteristics were moderately correlated with academic performance: Positivity, Negativity, and Fatigue (see Table 7.4). We then examined whether these contingencies reflected a flexible and adaptive resilience to situational changes or whether they merely reflected individual differences in cognitive ability as traditionally assessed and motivation and cognitive engagement style traits. To do this we ran a hierarchical regression analysis with cognitive ability as traditionally assessed, the Big Five domains, and the three Goal Orientation domains entered in the first block and the situation contingencies were entered in the final block. The first block The results of this regression analysis are contained in Table 7.5. Positivity, Negativity, and Fatigue contingencies all predicted significant variation in academic performance above and beyond all of these first block predictors, suggesting that situation contingencies capture an aspect of academic performance above and beyond individual differences in cognitive ability, personality, and motivational traits.

We then conducted a series of mediation analyses to investigate the direction of this relationship. For each of the cognitive dynamics parameter that was a significant incremental predictor in Table 7.5, the model tests whether the relationship between cognitive ability as traditionally conceptualised (single session performance) and academic performance is mediated by ecological cognitive performance (situation contingency parameters). The Big Five domains and the three Goal Orientation domains were included as background confounders. The results of this analysis are shown in Tables 7.6 and 7.7. Positivity, Negativity, and Fatigue contingencies were all unidirectional mediators of the relationship between cognitive ability as traditionally assessed and academic performance. This implies that situation contingencies are important for real-world cognitive performance and do not merely reflect individual differences in static cognitive ability, personality, or motivation.

Table 7.4. Correlation between situation contingencies and academic performance.

Situation Contingencies	Academic Performance
Duty	-.046
Intellect	.001
Positivity	**.259**
Negativity	**.366**
Sociality	.034
Distractions	.172
Environment	-.126
Fatigue	**.296**

Table 7.5. Additional predictive utility of situation contingencies.

	β	R^2_{model}	$R^2\Delta$	$p(F\Delta)$
Step 1: Baseline Model		.097	.097	.535
Single session cognitive performance (ICAR)	0.177			
BFAS: Extraversion	0.112			
BFAS: Agreeableness	0.060			
BFAS: Conscientiousness	0.123			
BFAS: Openness/Intellect	-0.153			
BFAS: Neuroticism	0.158			
Learning Goal Orientation	0.173			
Perform-Avoid Goal Orientation	-0.080			
Perform-Prove Goal Orientation	-0.074			
Step 2: Additional Predictive Utility of Situation Contingencies				
Duty	-0.022	.097	.000	.849
Intellect	0.017	.097	.000	.881
Positivity	**0.290**	**.152**	**.055**	**.031**
Negativity	**0.526**	**.241**	**.144**	**<.001**
Sociality	0.030	.098	.001	.802
Distractions	0.163	.117	.020	.196
Environment	-0.124	.110	.013	.293
Fatigue	**0.329**	**.174**	**.077**	**.010**

Note. Step 2 variables compared against Step 1 baseline model, which included single session score, BFAS domains, and goal orientation domains as predictors, which had an $R^2 = .097$.

Table 7.6. Single session score predicting academic performance, mediated by situation contingencies.

	Direct Effect		Indirect Effect		Significance	
Situation Contingency Mediator	Estimate	p	Estimate	p	Sobel	p
Positivity	.018	.625	**.036**	**.041**	**2.16**	**.035**
Negativity	-.034	.376	**.087**	**<.001**	**3.11**	**.002**
Fatigue	.005	.886	**.048**	**.015**	**2.49**	**.013**

Note. BFAS domains and goal orientation domains included as background confounders.

Table 7.7. Situation contingencies predicting academic performance, mediated by single session score.

	Direct Effect		Indirect Effect	
Situation Contingency Predictor	Estimate	p	Estimate	p
Positivity	**72.452**	**.019**	6.283	.627
Negativity	**148.232**	**<.001**	-18.405	.381
Fatigue	**67.012**	**.005**	1.597	.886

Note. BFAS domains and goal orientation domains included as background confounders.

5. Discussion

The purpose of this study was to begin building an explanatory process account for how and why individuals' ecological cognitive performance varies across time and contexts. Firstly, we validated our prior findings from Chapters 4 and 5. We then extended upon these findings by showing that ecological cognitive assessment captures an adaptive process that allows individuals to be resilient to changes in the environment. This adaptive process is important for university performance over and above cognitive ability as we traditionally assess it, and personality and motivational traits implicated in cognitive engagement.

5.1. Validation of Prior Findings

Our results replicated the findings from Study 1 and 2 (Chapters 4 and 5), affirming that ecological assessment predicts incremental variation in academic performance over and above cognitive ability as traditionally assessed. As for Study 1, which used the same reasoning tasks as this study, central tendency parameters were the most robust predictors of academic performance. Furthermore, this relationship emerged over and above personality and motivational traits that drive cognitive engagement and performance.

5.2. Adaptive Situation Contingencies

Three adaptive situation contingencies emerged as predictors of university performance – Positivity, Negativity, and Fatigue. The Positivity contingency reflects the extent to which an individual's cognitive performance changes as the self-rated pleasantness of the situation changes. The Negativity contingency reflects the extent to which an individual's cognitive performance changes as the self-rated level of negative feelings in the situation changes. Lastly, the Fatigue contingency reflects the extent to which an individual's cognitive performance changes as their self-rated level of fatigue changes. All three predictive relationships were in a positive direction. An examination of the within-person slopes plots in Figure 7.2 adds clarification to these findings; as most people's slopes were negative, the positive relationship means that individuals whose slopes were relatively flatter or slightly positive as the valence of the situation increased performed better at university. This may be construed as an adaptive contingency because these individuals demonstrated consistent performance regardless of the situation.

All three contingencies are internally focused; that is, they require an appraisal of one's internal state in the situation. This is in contrast to the other situation contingencies we captured, which predominantly tap into an external focus (with the exception of Intellect, which requires an internal appraisal of the cognitive demands of the situation). Why did resilience to changes in these internal situation factors emerge as predictors and not resilience to changes in other, predominantly external, situational factors? As the first study to explore these relationships, we preface our explanation by stating that these results require replication to ensure that they are robust. However, we do propose some reasons as to why resilience in response to changes in these factors, and not others, facilitates academic performance.

Empirical evidence suggests that psychosocial factors and affective states influence cognitive processes (Harmon-Jones et al., 2013; Royall et al., 2012; Tyng et al., 2017; Yang et al., 2013; Zahodne et al., 2014). In this study our sample was university students, most of whom were in late adolescence or early adulthood. This stage of life is marked by major social, emotional, and environmental changes, as well as heightened incidence of mental health conditions (Brown, 2018). Therefore, resilience to positive and negative affect and fatigue may be relatively more important for success at university than other factors because, at this stage of life, these internal factors are more impactful on behaviour than at other stages of life. Adolescents and young adults who are relatively resilient to the variety and intensity of these dynamic and changeable affective and internal cues may be better equipped to succeed at cognitive challenges. Conversely, adolescents and young adults who constantly have to recalibrate their internal state may have struggled to perform consistently on cognitive challenges. This hypothesis could be confirmed in future studies with broader sample demographics or with the inclusion of measures of self-regulation (for example, intrinsic emotion regulation).

5.3. Implications

Our findings have significant implications for our theories of cognitive ability and methods of cognitive assessment. They suggest that ecological within-person variation in cognitive performance is not noise but a systematic and meaningful component of cognitive ability that is not integrated into our theories and methods at present. Further, our findings imply that cognitive ability is meaningful for real-world cognitive performance not just because it allows us to succeed in constrained cognitive tasks but because it allows us to flexibly, adaptably, and resiliently perform under a range of situational cues.

Our findings also have practical implications for cognitive assessment practices. Most users of cognitive assessments, at least in assessment and selection scenarios, would intuitively prefer to select candidates whose behaviour and performance is consistent across time and contexts as these individuals will presumably be more reliable in future scenarios. The results of this study support this notion and raise interesting future pathways for cognitive assessment methods. Our results demonstrate a viable method for measuring cognitive performance consistency in addition to relative levels of mean cognitive performance, which could complement and enhance existing methods of cognitive assessment in high volume assessment and selection scenarios where selectors have little other data to go by, for example, graduate recruitment.

5.4. Limitations and Future Directions

We acknowledge several limitations of the present study and suggest some future directions for research. Our sample was skewed young, female, and highly educated, thus replication in more diverse samples is essential for understanding the generalisability of our results. We also recommend replication and extension using a different criterion outcome, such as workplace performance, to explore the robustness of these findings in other cognitively demanding real-world environments. Furthermore, in this study we construed personality and motivational variables as traits, however, there is ample evidence that such traits also vary in the short-term (Fleeson, 2001; Locke & Braver, 2008; Wigfield, 1994). Future studies could capture how changes in state personality and/or motivation also influence short-term within-person performance, as well the aforementioned inclusion of fluctuating self-regulation variables such as intrinsic emotion regulation. Our contingency models were also constrained to be linear but, in actual fact, may not be. Future research could examine non-linear adaptive contingencies to confirm or add nuance to our findings.

In addition, we encourage researchers to consider implementing app-based cognitive EMA with restricted time windows for completion. We sent tasks via email and allowed participants to delay EMA pulse completion (i.e., by allowing them to complete it within a 24 hour window), as we reasoned that cognitive tasks require time and headspace to engage with. However, this may have resulted in participants delaying until a more optimal time to complete the pulse tasks. App-based functionality could offer a practical solution to enabling rapid task completion, while also introducing a gamification element to engage participants. Moreover, app-based cognitive EMA could lend itself to practical applications of our findings in assessment and selection contexts.

Finally, we make comment on a notable situation factor that has been extensively explored in the literature but which we did not explore here: the stakes of the situation in which a task is completed (Goff & Ackerman, 1992; von Stumm et al., 2011). In high stakes situations, where the perceived importance of successful task completion is elevated, individuals strive for maximal performance and their performance is likely to reach its peak (i.e., performance will be around the upper tail of their within-person performance distribution). However, as the stakes and perceived importance decrease, people tend to exhibit more typical performance (i.e., performance will be towards the central tendency or even the lower tails of their within-person performance distribution). This is because lower stakes activities require greater discretionary effort and motivation to voluntarily engage cognitive resources, as the expectancy of a valued outcome is diminished (Wigfield, 1994). How an individual responds to differences in the stakes of a situation is likely to manifest in between-individual differences in within-person performance variation because individuals will be differentially susceptible to the influence of situational stakes. Nevertheless, unlike a simple measure like the DIAMONDS, eliciting the stakes-within-person performance relationship would require the experimental manipulation of the stakes in which within-person performance data is gathered, e.g., by offering valued performance incentives and/or identity priming. We recommend this as an important avenue for future research, not only to investigate the role of stakes in within-person cognitive performance variation but to understand whether our observed results generalise to higher stakes contexts. This is especially important given that there is evidence that cognitive processes vary interactively depending on the motivational intensity of the affect experienced in a situation (Harmon-Jones et al., 2013; Locke & Braver, 2008).

6. Conclusions

In this work we have demonstrated that short-term within-person variation in cognitive performance is not just systematic and meaningful but the key driver of success in cognitively demanding real world environments. Short-term within-person variation is an adaptive process that facilitates resilience to changes in internal states, which enables individuals to perform well under varied situational cues. Our work suggests that cognitive ability is a meaningful predictor of university performance because it can be used in this flexible and adaptive manner, and it is this flexibility and adaptivity that determines success in cognitively demanding real world scenarios rather than cognitive ability as it is traditionally conceptualised and operationalised. We suggest that future research builds on these findings by furthering exploring within-person contingencies and by replicating these findings in other contexts, such as workplace samples and broader demographics.

VIII. GENERAL CONCLUSIONS

The central aim of this thesis was to explore whether the observed within-person variation in cognitive performance in prior empirical research was noise, or something systematic and meaningful for our theories of cognitive ability and methods of cognitive assessment. To do this, we leveraged ecological assessment methods and analysis approaches that allowed us to more comprehensively understand if, how, and why cognitive performance varies over time within the individual. This thesis was structured around four research questions. We now summarise the general findings of this thesis as they pertain to each research question.

1. Summary of Findings

Research Question 1: Why is understanding within-individual variability in cognitive performance important for our theories of cognitive ability?

We argued that cognitive ability as it is traditionally conceived of and operationalised, as a point in time assessment from which sum scores reflecting a static underlying ability are derived and compared between individuals, is fundamentally asymmetrical with the criterion outcomes it is commonly used to predict. These outcomes, such as academic and workplace performance, encompass within-person variation in the application of cognitive ability across time and contexts. We proposed the use of the EMA and within-person analysis techniques to study within-person variation in cognitive performance and understand whether it is just noise, or whether it is systematic and meaningful.

Research Question 2: How can we assess and analyse within-individual variability in cognitive performance?

Having proposed the use of the EMA to capture ecological within-individual variability in cognitive performance, we then turned to the question of how to quantify this observed variation. We proposed three classes of quantitative parameters that leverage the short-term repeated measures data generated from an EMA approach to model the extent, magnitude, and importance of ecological within-individual variation in cognitive performance. We referred to these parameters collectively as “cognitive dynamics” to highlight they are indicators of performance derived across different repeated measures and settings more so than is typically the case. These classes of parameters were a) central tendency (mean and median of repeated measures data), b) distributional parameters (SD, MSSD, skewness, best performance and worst performance), and c) contingency parameters (within-person random slopes contingent on some person, task, or situation factor).

Research Question 3: Does this assessment and analysis approach lead to better prediction of academic performance than cognitive ability as traditionally conceptualised and assessed?

We then generated these cognitive dynamics parameters for three empirical studies across Chapters 4 to 7 and compared the predictive utility of the cognitive dynamics with a measure of cognitive ability as traditionally conceptualised and assessed: a single occasion, context-invariant assessment where performance was measured as a sum score of correct responses. Across all three studies, our findings were consistent; many of these cognitive dynamics incrementally predicted university performance over and above performance on a traditional single session cognitive assessment. Specifically, central tendency cognitive dynamics (mean and median) and some distributional cognitive dynamics (best and worst performance) emerged as the most consistent and robust predictors. Further analyses revealed that these parameters predicted additional variation in university performance because they captured something broader than, or different to, cognitive ability as it is traditionally conceptualised and assessed.

Research Question 4: Why might this assessment and analysis approach lead to better prediction of academic performance than cognitive ability as traditionally conceptualised and assessed?

Finally, having established that within-person variation in cognitive performance appears to capture an important aspect of cognitive ability not accounted for by cognitive ability as it is traditionally conceptualised and assessed, we turned to the question of why this relationship may emerge. In Chapter 7, we began building and testing an explanatory process account for why ecological within-individual variability in cognitive performance predicts academic performance beyond cognitive ability as traditionally conceptualised and assessed. We proposed that within-person variation, or lack thereof, may reflect individual differences in performance across time and situations. To explore this hypothesis, we derived within-person situation contingencies (i.e., situation-performance relationships) that demonstrate individual differences in the impact of the situation on cognitive performance. We demonstrated that some of these situation contingencies, but not others, predicted university performance over and above cognitive ability as traditionally conceptualised and personality and motivational traits associated with cognitive engagement. Specifically, individuals who were relatively more resilient to changes in positive and negative affect, and fatigue (i.e., who had flatter situation-performance contingencies) performed better at university. We suggested that cognitive ability is a meaningful predictor of university performance because it can be used in this flexible and adaptive manner, and it is this flexibility and adaptivity that leads to behavioural consistency that determines success in cognitively demanding real world scenarios rather than cognitive ability as it is traditionally conceptualised and operationalised.

2. Implications for Theories

Our findings challenge the construct validity of cognitive ability as it is currently conceptualised by demonstrating that short-term within-individual variability in cognitive performance is a systematic and meaningful component of our real-world application of cognitive ability, and not just noise. Most dominant theories of cognitive ability assume that cognitive ability differs between individuals but remains static within individuals and, as a result, any observed within-individual variability is construed as measurement error (Birney et al., 2019; Cronbach, 1957). However, our findings suggest that short-term within-person variability in cognitive performance is systematic and meaningful, and that ecological cognitive assessment captures short-term within-person variability in cognitive performance that has predictive validity for real-world cognitive performance. Further, our results demonstrate that within-person variability represents a distinct but related component of cognitive ability over and above this more static form of cognitive ability. We suggest that this distinct but related component is what is not captured by existing theories of cognitive ability: flexible, adaptable, and resilient performance under a range of situational cues that is critical for success in increasingly complex and ambiguous academic and workplace settings.

3. Implications for Cognitive Assessment

Our findings also have implications for how we use and interpret results from cognitive assessments in a range of applied settings. Current single-session cognitive assessment approaches do not allow for any understanding of how individuals vary in their performance across time and contexts. In contrast, ecological assessment approaches and analysis techniques enable us to understand the whole person across time and contexts, and the degree to which they can flexibly and adaptably use their cognitive ability. Using this ecological assessment approach has the potential to greatly enhance the utility of cognitive assessment by more closely aligning assessment methods and analysis approaches with criterion outcomes. Furthermore, that participants were receptive to the use of this assessment method and engaged with it at similar rates to ecological assessment in related fields suggests that using this assessment method would likely be acceptable to participants.

Ecological assessment has a range of potential applications across organisations. In graduate recruitment, candidates have little work experience and are often selected on academic performance and cognitive performance alone. Incorporating ecological assessment would allow recruiters to select candidates who are both high performers, as shown by their academic performance, but also consistent performers, as shown by their ecological cognitive performance variation across time and contexts. Ecological assessment could also aid recruitment professionals in the selection of individuals for specialised placement programs, such as medical specialty programs, by identifying those who are likely to thrive and perform consistently in specific, often high-pressure situations, ensuring better alignment between training investments and outcomes. In learning and development contexts, ecological assessment could also allow for better individualisation of training and development programs by facilitating an understanding of an individual’s strengths and weaknesses in different siutations and circumstances. For example, an employee who excels in collaborative social projects but struggles with independent work could receive targeted support to improve their self-management skills.

In educational settings, ecological assessment could similarly be used to identify situations in which students perform relatively better or worse. A student who struggles academically might benefit from an ecological assessment that identifies situational factors impacting their learning. This information could be used to create a tailored learning plan that addresses their unique needs, such as providing additional support in particular subjects or adjusting the learning environment to meet their needs. This approach may also aid in supporting students with learning difficulties or neurodivergent conditions to be fully supported at school or university.

4. Conclusion

This thesis investigated the utility of ecological cognitive assessment for capturing and measuring ecological within-person variation in cognitive performance, with the aim of understanding whether this could improve our conceptualisation and measurement of cognitive ability. Until relatively recently, most dominant theories of intelligence have assumed that abilities differ between individuals but not within individuals, at least not over reasonable periods of time (Neisser et al., 1996). As a result, common psychometric assessments of cognitive ability focus on composite accuracy scores, often taken at a single point in time in a single, standardised context. These scores are then used to predict future between-person differences in cognitive performance in educational, workplace, and other cognitively demanding environments. This approach is problematic because it is not in symmetry with the nature of cognitive performance in these criterion environments: a cumulative pattern of ecological cognitive performances across time and contexts, which inherently encompasses within-individual variability.

Our findings suggested that ecological cognitive assessment has promising utility as a method of cognitive assessment. Cognitive dynamics parameters elicited from EMA data predicted incremental variation in university performance over and above traditional single session cognitive assessments, and this relationship appeared to emerge because ecological within-person variation reflects a construct broader than, or different to, what is measured in context-invariant single-session assessments of cognitive performance. We replicated this finding across three studies and, furthermore, found additional evidence that this may be because cognitive dynamics capture between-person differences in within-person variability over time and contexts that is more symmetrical with criterion outcomes such as university performance. Moreover, this within-person variability reflects an adaptive process that facilitates resilience to contextual changes and is critical for success in cognitively demanding real-world environments.

References

Ackerman, P. L. (1996). A Theory of Adult Intellectual Development: Process, Personality, Interests, and Knowledge. *Intelligence*, *22*(2), 227-257. https://doi.org/10.1016/S0160-2896(96)90016-1

Allaire, J. C., & Marsiske, M. (2005). Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychol Aging*, *20*(3), 390-401. https://doi.org/10.1037/0882-7974.20.3.390

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *J Appl Psychol*, *90*(6), 1054-1068. https://doi.org/10.1037/0021-9010.90.6.1054

Beckmann, J. F. (2010). Taming a beast of burden: On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction*, *20*(3), 250-264. https://doi.org/10.1016/j.learninstruc.2009.02.024

Beckmann, N., Birney, D. P., Beckmann, J. F., Wood, R. E., Sojo, V., & Bowman, D. (2020). Inter-individual differences in intra-individual variability in personality within and across contexts. *Journal of Research in Personality*, *85*, 1-22. https://doi.org/10.1016/j.jrp.2019.103909

Beckmann, N., Birney, D. P., Minbashian, A., & Beckmann, J. F. (2021). Personality dynamics at work: The effects of form, time, and context of variability. *European Journal of Personality*, *35*(4), 421-449. https://doi.org/10.1177/08902070211017341

Birney, D. P., & Beckmann, J. F. (2022). Intelligence IS Cognitive Flexibility: Why Multilevel Models of Within-Individual Processes Are Needed to Realise This. *J Intell*, *10*(3). https://doi.org/10.3390/jintelligence10030049

Birney, D. P., Beckmann, J. F., & Beckmann, N. (2019). Within-individual variability of ability and learning trajectories in complex problems. In D. J. McFarland (Ed.), *General and Specific Mental Abilities* (pp. 253-283). Cambridge Scholars Publishing.

Birney, D. P., Beckmann, J. F., Beckmann, N., & Double, K. S. (2017). Beyond the intellect: Complexity and learning trajectories in Raven's Progressive Matrices depend on self-regulatory processes and conative dispositions. *Intelligence*, *61*, 63-77. https://doi.org/10.1016/j.intell.2017.01.005

Birney, D. P., Beckmann, J. F., Beckmann, N., Double, K. S., & Whittingham, K. (2018). Moderators of learning and performance trajectories in microworld simulations: Too soon to give up on intellect!? *Intelligence*, *68*, 128-140. https://doi.org/10.1016/j.intell.2018.03.008

Birney, D. P., Beckmann, J. F., Beckmann, N., & Stemler, S. E. (2022). Sophisticated Statistics Cannot Compensate for Method Effects If Quantifiable Structure Is Compromised. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.812963

Blasiman, R. N., Larabee, D., & Fabry, D. (2018). Distracted students: A comparison of multiple types of distractions on learning in online lectures. *Scholarship of Teaching and Learning in Psychology*, *4*(4), 222-230. https://doi.org/10.1037/stl0000122

Blatchford, P., Edmonds, S., & Martin, C. (2003). Class size, pupil attentiveness and peer relations. *British Journal of Educational Psychology*, *73*(1), 15-36. https://doi.org/https://doi.org/10.1348/000709903762869897

Bradley, A., Beevers-Cowling, F., Norton, C., Hill, C., Pelopida, B., & Quigley, M. (2021). Falling at the first hurdle: undergraduate students’ readiness to navigate the graduate recruitment process. *Studies in Higher Education*, *46*(9), 1827-1838. https://doi.org/10.1080/03075079.2019.1709164

Brown, J. S. L. (2018). Student mental health: some answers and more questions. *J Ment Health*, *27*(3), 193-196. https://doi.org/10.1080/09638237.2018.1470319

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world witout a heterogeneity revolution. *Nature Human Behavior*, *5*, 980-989. https://doi.org/10.1038/s41562-021-01143-3

Bunce, D., MacDonald, S. W. S., & Hultsch, D. F. (2004). Inconsistency in serial choice decision and motor reaction times dissociate in younger and older adults. *Brain Cogn*, *56*(3), 320-327. https://doi.org/10.1016/j.bandc.2004.08.006

Cacioppo, J., & Petty, R. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116-131. https://doi.org/10.1037/0022-3514.42.1.116

Carless, S. A. (2007). Graduate Recruitment and Selection in Australia. *International Journal of Selection and Assessment*, *15*(2), 153-166. https://doi.org/10.1111/j.1468-2389.2007.00377.x

Carless, S. A. (2009). Psychological testing for selection purposes: a guide to evidence-based practice for human resource professionals. *The International Journal of Human Resource Management*, *20*(12), 2517-2532. https://doi.org/10.1080/09585190903363821

Carroll, L. (2024, 21st April). Record 18,500 students sit selective school test on paper for last time. *The Sydney Morning Herald*. https://www.smh.com.au/national/nsw/record-18-500-students-sit-selective-school-test-on-paper-for-last-time-20240417-p5fkhg.html

Church, T. A., Katigbak, M. S., Ching, C. M., Zhang, H., Shen, J., Arias, R. M., Rincon, B. C., Morio, H., Tanaka-Matsumi, J., Takaoka, S., Mastor, K. A., Roslan, N. A., Ibáñez-Reyes, J., Vargas-Flores, J. d. J., Locke, K. D., Reyes, J. A. S., Wenmei, S., Ortiz, F. A., & Alvarez, J. M. (2013). Within-individual variability in self-concepts and personality states: Applying density distribution and situation-behavior approaches across cultures. *Journal of Research in Personality*, *47*(6), 922-935. https://doi.org/10.1016/j.jrp.2013.09.002

Conner, T. S., Tennen, H., Fleeson, W., & Feldman Barrett, L. (2009). Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. *Social and Personality Psychology Compass*, *3*(3), 292-313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

Cools, R. (2008). Role of dopamine in the motivational and cognitive control of behavior. *Neuroscientist*, *14*(4), 381-395. https://doi.org/10.1177/1073858408317009

Coyle, T. R. (2001). IQ is related to the worst performance rule in a memory task involving children. *Intelligence*, *29*(2), 117-129. https://doi.org/10.1016/S0160-2896(00)00044-1

Coyle, T. R. (2003a). IQ, the worst performance rule, and Spearman's law: a reanalysis and extension. *Intelligence*, *31*(5), 473-489. https://doi.org/10.1016/s0160-2896(02)00175-7

Coyle, T. R. (2003b). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*(6), 567-587. https://doi.org/10.1016/s0160-2896(03)00054-0

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*(11), 671-684. https://doi.org/10.1037/h0043943

Csikszentmihalyi, M., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In *Flow and the Foundations of Positive Psychology* (pp. 35-54). Springer. https://doi.org/doi.org/10.1007/978-94-017-9088-8_3

Danziger, K. (1990). The triumph of the aggregate. In K. Danziger (Ed.), *Constructing the Subject: Historical Origins of Psychological Research* (pp. 68-87). Cambridge University Press.

Demetriou, A., Spanoudis, G., Christou, C., Greiff, S., Makris, N., Vainikainen, M.-P., Golino, H., & Gonida, E. (2023). Cognitive and personality predictors of school performance from preschool to secondary school: An overarching model. *Psychological Review*, *130*(2), 480-512. https://doi.org/10.1037/rev0000399

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880-896. https://doi.org/10.1037/0022-3514.93.5.880

Diallo, T. M., Morin, A. J., & Parker, P. D. (2014). Statistical power of latent growth curve models to detect quadratic growth. *Behav Res Methods*, *46*(2), 357-371. https://doi.org/10.3758/s13428-013-0395-1

Dontre, A. J. (2021). The influence of technology on academic distraction: A review. *Human Behavior and Emerging Technologies*, *3*(3), 379-390. https://doi.org/https://doi.org/10.1002/hbe2.229

Dworak, E. M., Revelle, W., Doebler, P., & Condon, D. M. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, *169*. https://doi.org/10.1016/j.paid.2020.109906

Education, N. D. o. (2024). *400,000 NSW students begin NAPLAN testing*. Retrieved 19th July from https://education.nsw.gov.au/news/latest-news/400-000-nsw-students-begin-naplan-testing

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. https://doi.org/10.3758/BF03193146

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, *33*(7), 865-877. https://doi.org/10.1002/job.1803

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin*, *52*(3), 217-250. https://doi.org/10.1037/h0045276

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distribution of states. *Journal of Personality and Social Psychology*, *80*(6), 1011-1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *J Pers*, *75*(4), 825-861. https://doi.org/10.1111/j.1467-6494.2007.00458.x

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *J Res Pers*, *56*, 82-92. https://doi.org/10.1016/j.jrp.2014.10.009

Geier, C. F., Terwilliger, R., Teslovich, T., Velanova, K., & Luna, B. (2010). Immaturities in reward processing and its influence on inhibitory control in adolescence. *Cereb Cortex*, *20*(7), 1613-1629. https://doi.org/10.1093/cercor/bhp225

Goff, M., & Ackerman, P. L. (1992). Personality–Intelligence Relations: Assessment of Typical Intellectual Engagement. *Journal of Educational Psychology*, *84*(4), 537-552. https://doi.org/10.1037/0022-0663.84.4.537

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence*, *33*(2), 169-186. https://doi.org/10.1016/j.intell.2004.10.002

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, *21*(2), 273-286. https://doi.org/10.1016/j.chb.2004.02.014

Graydon, J., & Eysenck, M. W. (1988). Distraction and cognitive performance. *European Journal of Cognitive Psychology*, *1*(2), 161-179. https://doi.org/10.1080/09541448908403078

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2013). Does Negative Affect Always Narrow and Positive Affect Always Broaden the Mind? Considering the Influence of Motivational Intensity on Cognitive Scope. *Current Directions in Psychological Science*, *22*(4), 301-307. https://doi.org/10.1177/0963721413481353

Hultsch, D. F., Strauss, E., Hunter, M. A., & MacDonald, S. W. S. (2008). Intraindividual variability, cognition, and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 491-556). Psychology Press.

ICAR. (2014). *The International Cognitive Ability Resource*. https://icar-project.com/

Ilies, R., & Judge, T., A. (2002). Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes*, *89*, 1119-1139. https://doi.org/10.1016/S0749-5978(02)00018-3

Jones, A. B., Brown, N. A., Serfass, D. G., & Sherman, R. A. (2017). Personality and density distributions of behavior, emotions, and situations. *Journal of Research in Personality*, *69*, 225-236. https://doi.org/10.1016/j.jrp.2016.10.006

Keis, O., Helbig, H., Streb, J., & Hille, K. (2014). Influence of blue-enriched classroom lighting on students′ cognitive performance. *Trends in Neuroscience and Education*, *3*(3-4), 86-92. https://doi.org/10.1016/j.tine.2014.09.001

Ko, W. H., Schiavon, S., Zhang, H., Graham, L. T., Brager, G., Mauss, I., & Lin, Y.-W. (2020). The impact of a view from a window on thermal comfort, emotion, and cognitive performance. *Building and Environment*, *175*. https://doi.org/10.1016/j.buildenv.2020.106779

Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the Couplings Among Intellectual Abilities and Constituent Cognitive Processes Across the Life Span. *Psychological Science*, *15*(3), 155-163. https://doi.org/10.1111/j.0956-7976.2004.01503003.x

Liu, S., Kuppens, P., & Bringmann, L. (2021). On the Use of Empirical Bayes Estimates as Measures of Individual Traits. *Assessment*, *28*(3), 845-857. https://doi.org/10.1177/1073191119885019

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: behavior, brain activation, and individual differences. *Cogn Affect Behav Neurosci*, *8*(1), 99-112. https://doi.org/10.3758/cabn.8.1.99

Lovden, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-trial variability precedes and predicts cognitive decline in old and very old age: longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*(12), 2827-2838. https://doi.org/10.1016/j.neuropsychologia.2007.05.005

Lüdecke, D. (2023). *sjPlot: Data Visualization for Statistics in Social Science*. In https://CRAN.R-project.org/package=sjPlot

MacDonald, S. W. S., Hultsch, D. F., & Dixon, R. A. (2003). Performance variability is related to change in cognition: evidence from the Victoria Longitudinal Study. *Psychol Aging*, *18*(3), 510-523. https://doi.org/10.1037/0882-7974.18.3.510

MacDonald, S. W. S., Li, S.-C., & Backman, L. (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychol Aging*, *24*(4), 792-808. https://doi.org/10.1037/a0017798

May, K. E., & Elder, A. D. (2018). Efficient, helpful, or distracting? A literature review of media multitasking in relation to academic performance. *International Journal of*

Educational Technology in Higher Education, *15*(1), 13. https://doi.org/10.1186/s41239-018-0096-z

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115-142. https://doi.org/10.1037//0012-1649.38.1.115

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690-707. https://doi.org/10.1037/met0000153

Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *J Appl Psychol*, *95*(5), 793-806. https://doi.org/10.1037/a0020016

Nesselroade, J. R. (1991). The Warp and the Woof of the Developmental Fabric. In R. M. Downs, L. S. Liben, & D. S. Palermo (Eds.), *Visions of Aesthetics, the Environment & Development* (1 ed.). Psychology Press. https://doi.org/10.4324/9780203772652

Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149-155. https://doi.org/10.1016/j.jrp.2016.06.020

Paino, M., Pérez-Albéniz, A., Fonseca-Pedrero, E., Lucas-Molina, B., & Sánchez-García, M. A. (2018). Emotional and behavioral difficulties in adolescence: Relationship with emotional well-being, affect, and academic performance. *Anales de Psicología*, *34*(3), 482-489. https://doi.org/10.6018/analesps.34.3.296631

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., & Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *J Exp Child Psychol*, *108*(3), 621-637. https://doi.org/10.1016/j.jecp.2010.09.008

Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: tools for examining the development of dynamic characteristics and processes. *Psychol Aging*, *24*(4), 778-791. https://doi.org/10.1037/a0017915

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *J Pers Soc Psychol*, *107*(4), 677-718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., Jones, A. B., & Sherman, R. A. (2016). Directionality of Person-Situation Transactions: Are There Spillovers Among and Between Situation Experiences And Personality States? *Pers Soc Psychol Bull*, *42*(7), 893-909. https://doi.org/10.1177/0146167216647360

Rauthmann, J. F., & Sherman, R. A. (2016). Ultra-Brief Measures for the Situational Eight DIAMONDS Domains. *European Journal of Psychological Assessment*, *32*(2), 165-174. https://doi.org/10.1027/1015-5759/a000245

Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality*, *55*, 98-111. https://doi.org/10.1016/j.jrp.2015.02.003

Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. In https://CRAN.R-project.org/package=psych

Richardson, K., & Norgate, S. H. (2015). Does IQ Really Predict Job Performance? *Appl Dev Sci*, *19*(3), 153-169. https://doi.org/10.1080/10888691.2014.983635

Rizopoulos, D. (2006). ltm: An R package for Latent Variable Modelling and Item Response Theory Analyses. *Journal of Statistical Software*, *17*(5), 1-25. https://doi.org/10.18637/jss.v017.i05

Royall, D. R., Palmer, R., Chiodo, L. K., & Polk, M. J. (2012). Depressive symptoms predict longitudinal change in executive control but not memory. *Int J Geriatr Psychiatry*, *27*(1), 89-96. https://doi.org/10.1002/gps.2697

Sackett, P. R., Berry, C. M., Lievens, F., & Zhang, C. (2023). Correcting for range restriction in meta-analysis: A reply to Oh et al. (2023). *J Appl Psychol*, *108*(8), 1311-1315. https://doi.org/10.1037/apl0001116

Sackett, P. R., Demeke, S., Bazian, I. M., Griebie, A. M., Priest, R., & Kuncel, N. R. (2023). A contemporary look at the relationship between general cognitive ability and job performance. *J Appl Psychol*. https://doi.org/10.1037/apl0001159

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *J Appl Psychol*, *107*(11), 2040-2068. https://doi.org/10.1037/apl0000994

Salthouse, T. A. (2007). Implications of within-person variability in cognitive and neuropsychological functioning for the interpretation of change. *Neuropsychology*, *21*(4), 401-411. https://doi.org/10.1037/0894-4105.21.4.401

Salthouse, T. A. (2012). Psychometric properties of within-person across-session variability in accuracy of cognitive performance. *Assessment*, *19*(4), 494-501. https://doi.org/10.1177/1073191112438744

Salthouse, T. A., & Berish, D. E. (2005). Correlates of within-person (across-occasion) variability in reaction time. *Neuropsychology*, *19*(1), 77-87. https://doi.org/10.1037/0894-4105.19.1.77

Salthouse, T. A., Nesselroade, J. R., & Berish, D. E. (2006). Short-Term Variability in Cognitive Performance and the Calibration of Longitudinal Change. *Journal of Gerontology: Psychological Sciences*, *61B*(3), P144-P151. https://doi.org/10.1093/geronb/61.3.p144

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*(2), 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Schmidt, F. L., Oh, I.-S., & Shaffer, J. A. (2016). *The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 100 Years of Research Findings*.

Schmiedek, F., Lovden, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Front Aging Neurosci*, *2*. https://doi.org/10.3389/fnagi.2010.00027

Schmiedek, F., Lovden, M., & Lindenberger, U. (2013). Keeping it steady: older adults perform more consistently on cognitive tasks than younger adults. *Psychol Sci*, *24*(9), 1747-1754. https://doi.org/10.1177/0956797613479611

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *J Pers Soc Psychol*, *109*(5), 872-888. https://doi.org/10.1037/pspp0000036

Spearman, C. (1927). *The abilities of man: Their nature and measurement* (1st AMS ed ed.). AMS Press.

Steinmayr, R., Bipp, T., & Spinath, B. (2011). Goal orientations predict academic performance beyond intelligence and personality. *Learning and Individual Differences*, *21*(2), 196-200. https://doi.org/10.1016/j.lindif.2010.11.026

Sternberg, R. J., Forsythe, G. B., Hedlund, J., Horvath, J. A., Wagner, R. K., Williams, W. M., Snook, S. A., & Grigorenko, E. L. (2000). *Practical Intelligence in Everyday Life*. Cambridge University Press.

Strobel, A., Behnke, A., Gärtner, A., & Strobel, A. (2019). The interplay of intelligence and need for cognition in predicting school grades: A retrospective study. *Personality and Individual Differences*, *144*, 147-152. https://doi.org/10.1016/j.paid.2019.02.041

Team, J. (2023). *JASP (Version 0.14.1)*. In https://jasp-stats.org/

Team, R. (2020). *RStudio: Integrated Development for R*. In (Version 1.4.1103) RStudio, PBC. http://www.rstudio.com

Trull, T. J., Lane, S. P., Koval, P., & Ebner-Priemer, U. W. (2015). Affective Dynamics in Psychopathology. *Emot Rev*, *7*(4), 355-361. https://doi.org/10.1177/1754073915590617

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Dev Psychol*, *45*(4), 1097-1118. https://doi.org/10.1037/a0015864

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The Influences of Emotion on Learning and Memory. *Front Psychol*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

VandeWalle, D. (1997). Development and Validation of a Work Domain Goal Orientation Instrument. *Educational and Psychological Measurement*, *57*(6), 995-1015. https://doi.org/10.1177/0013164497057006009

Vaughan, A. C., & Birney, D. P. (2023). Within-Individual Variation in Cognitive Performance Is Not Noise: Why and How Cognitive Assessments Should Examine Within-Person Performance. *Journal of Intelligence*, *11*(6). https://doi.org/10.3390/jintelligence11060110

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a Unified Framework for the Study of Between-Person and Within-Person Structures: Building a Bridge Between Two Research Paradigms. *Multivariate Behav Res*, *49*(3), 193-213. https://doi.org/10.1080/00273171.2014.889593

von Stumm, S., Chamorro-Premuzic, T., & Ackerman, P. L. (2011). Re-Visiting Intelligence – Personality Associations: Vindicating Intellectual Investment. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (1 ed.). Blackwell Publishing Ltd.

Wendt, L. P., Wright, A. G. C., Pilkonis, P. A., Woods, W. C., Denissen, J. J. A., Kühnel, A., & Zimmermann, J. (2020). Indicators of Affect Dynamics: Structure, Reliability, and Personality Correlates. *European Journal of Personality*, *34*(6), 1060-1072. https://doi.org/10.1002/per.2277

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. In Springer-Verlag New York. https://ggplot2.tidyverse.org

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, *6*, 49-78. https://doi.org/10.1007/BF02209024

Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In *Learning and individual differences: Process, trait, and content determinants.* (pp. 77-108). https://doi.org/10.1037/10315-004

Wood, R. E., Beckmann, N., Birney, D. P., Beckmann, J. F., Minbashian, A., & Chau, R. (2019). Situation contingent units of personality at work. *Personality and Individual Differences*, *136*, 113-121. https://doi.org/10.1016/j.paid.2018.01.026

Wood, R. E., Clarke, T., Beckmann, J. F., & Birney, D. P. (2009). Simulations, learning and real world capabilities. *Education + Training*, *51*(5/6), 491-510. https://doi.org/10.1108/00400910910987273

Yang, H., Yang, S., & Isen, A. M. (2013). Positive affect improves working memory: implications for controlled cognitive processing. *Cogn Emot*, *27*(3), 474-482. https://doi.org/10.1080/02699931.2012.713325

Young, S. R., & Keith, T. Z. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, *38*(8), 1052-1059. https://doi.org/10.1177/0734282920943455

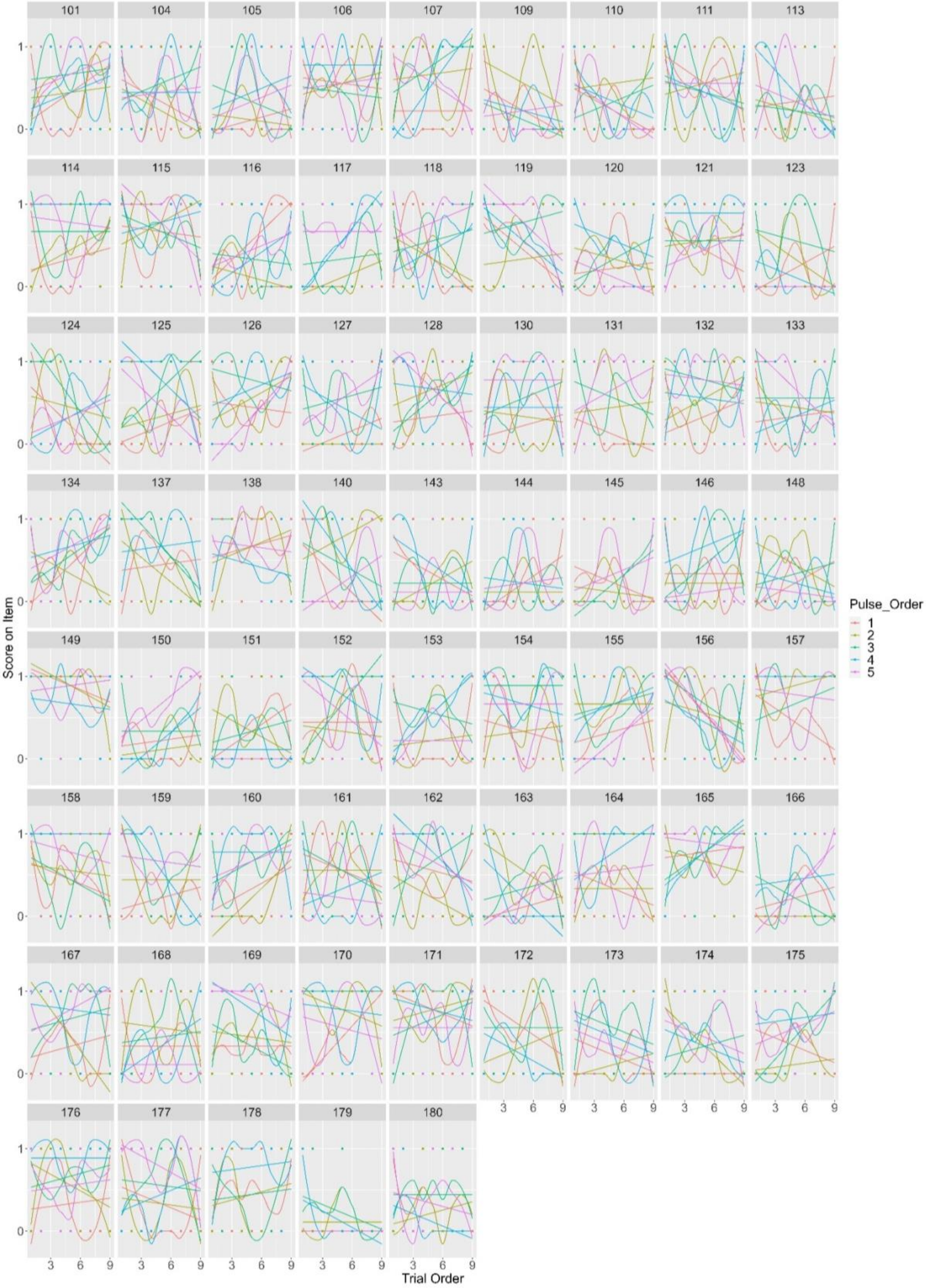
Zaboski, B. A., 2nd, Kranzler, J. H., & Gage, N. A. (2018). Meta-analysis of the relationship between academic achievement and broad abilities of the Cattell-horn-Carroll theory. *J Sch Psychol*, *71*, 42-56. https://doi.org/10.1016/j.jsp.2018.10.001

Zahodne, L. B., Nowinski, C. J., Gershon, R. C., & Manly, J. J. (2014). Which psychosocial factors best predict cognitive performance in older adults? *J Int Neuropsychol Soc*, *20*(5), 487-495. https://doi.org/10.1017/S1355617714000186

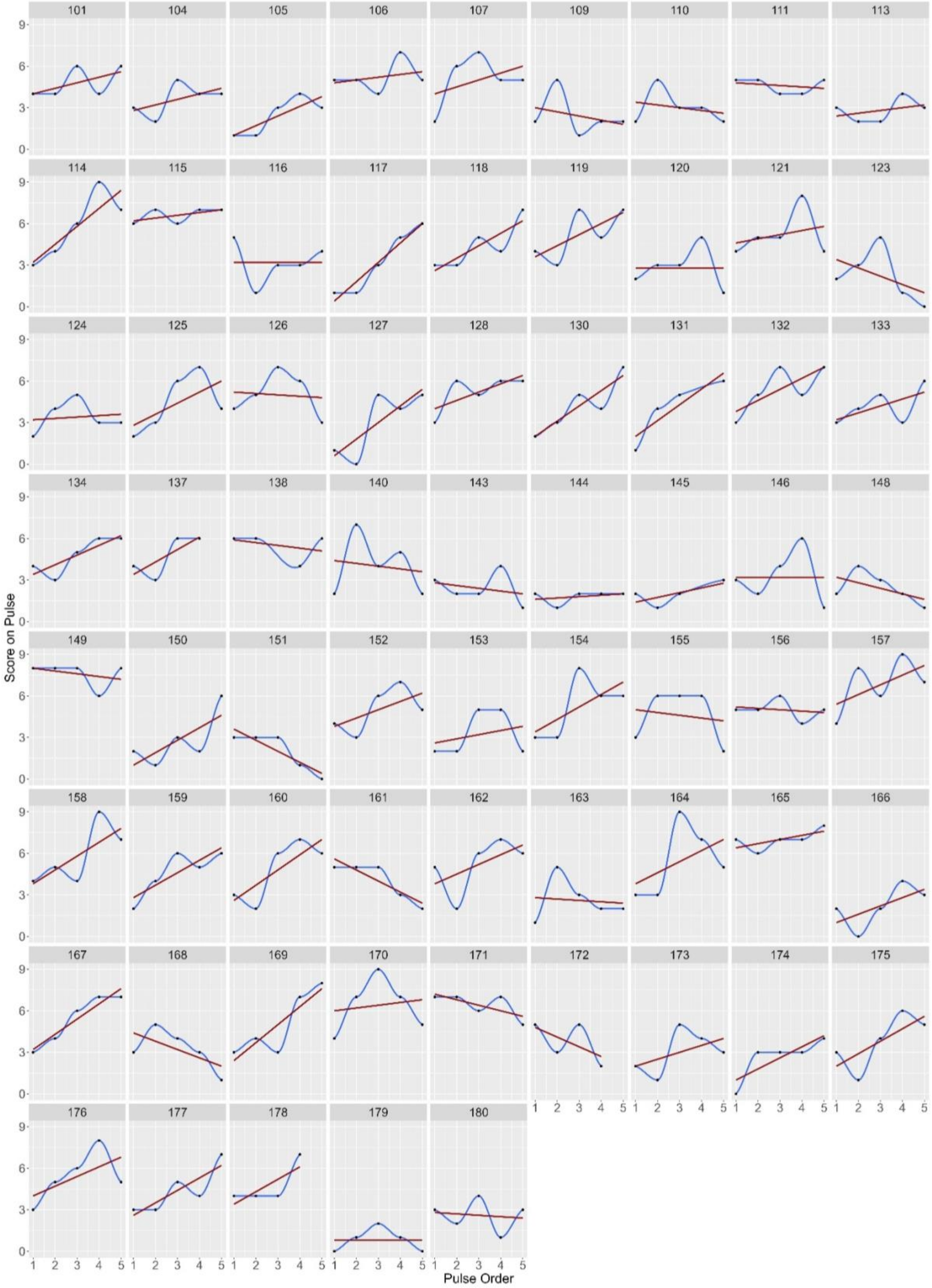
Zhang, F., de Dear, R., & Hancock, P. (2019). Effects of moderate thermal environments on cognitive performance: A multidisciplinary review. *Applied Energy*, *236*, 760-777. https://doi.org/10.1016/j.apenergy.2018.12.005

Supplementary Materials

S1. Raw within-task performance trajectories for all participants (Study 1).



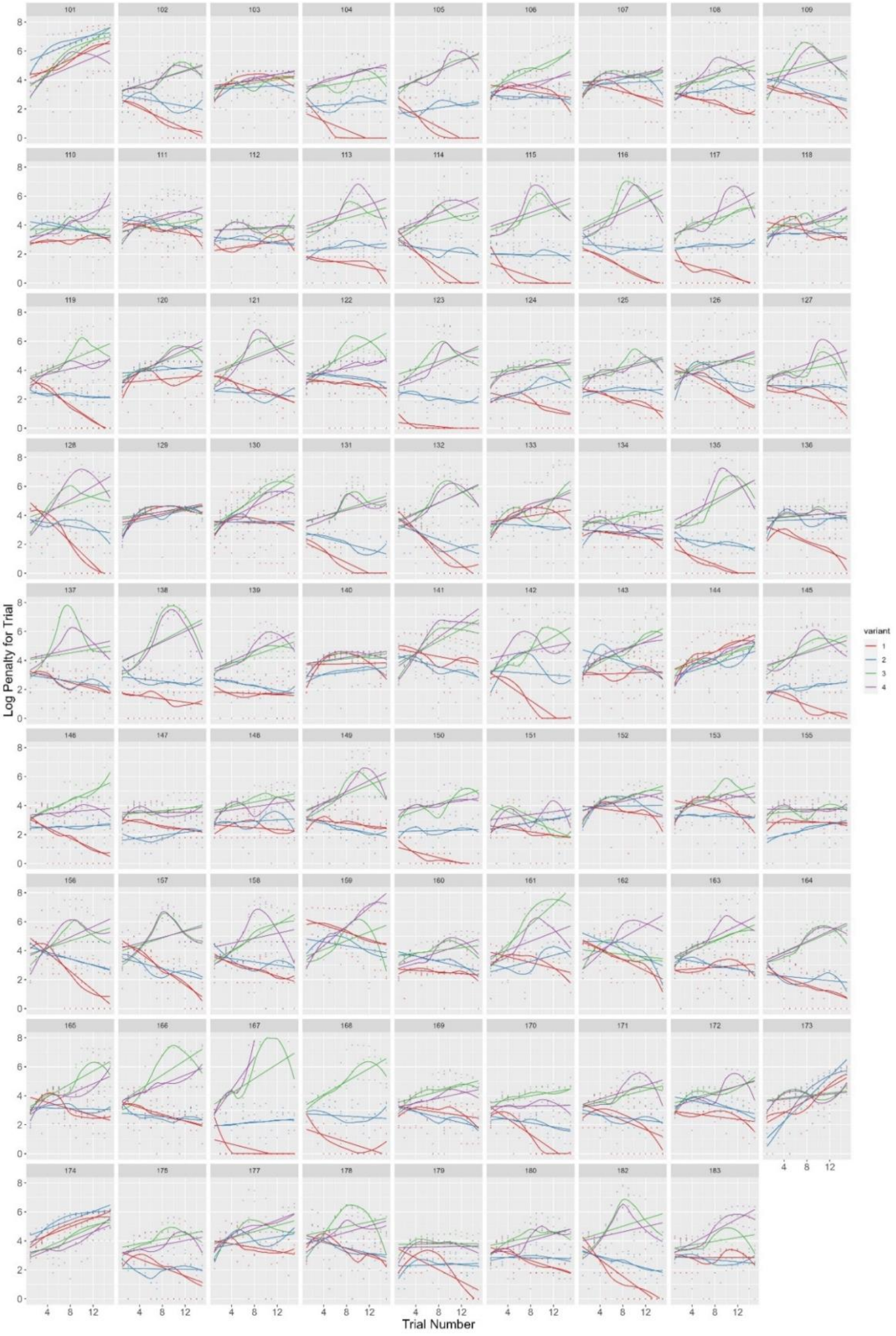
S2. Raw between-task performance trajectories for all participants with loess and linear model lines (Study 1).



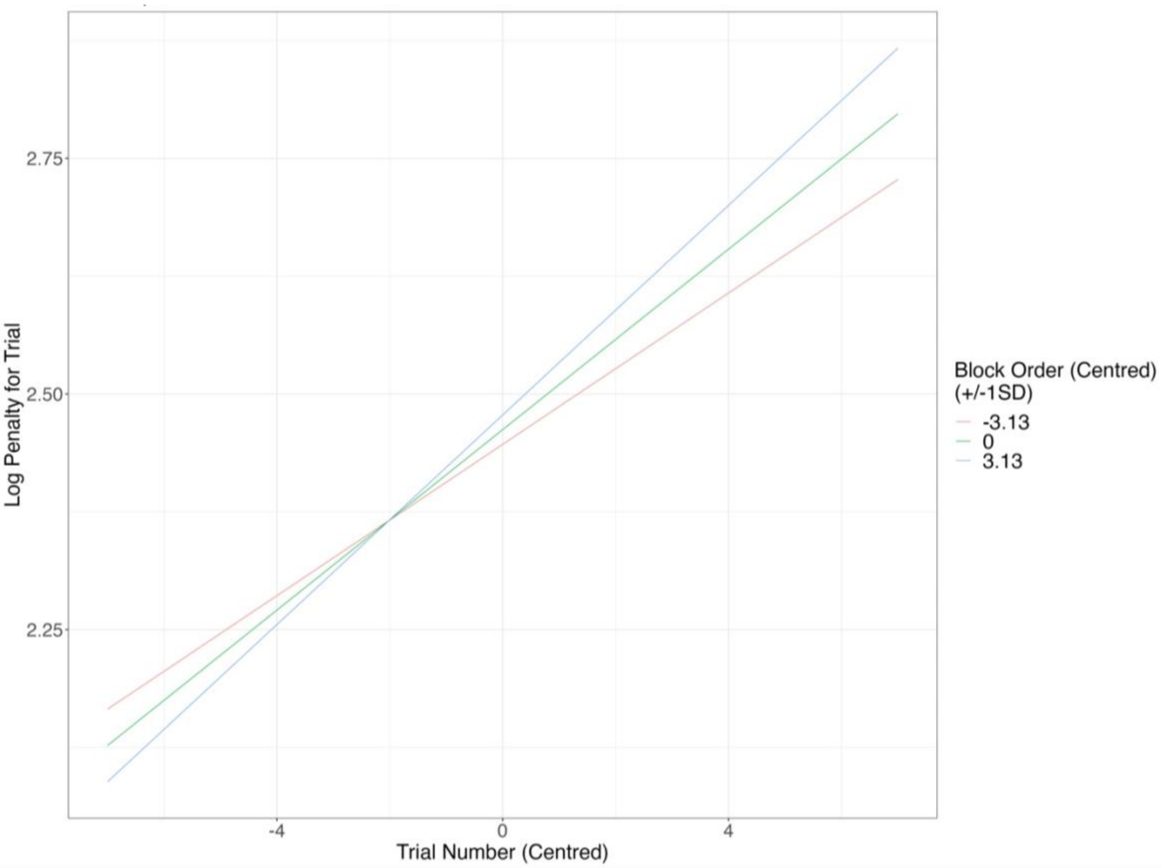
S3. Full intercorrelations between within-person parameters (Study 1).

		Central Tendency		Variance and Distributional				Contingencies		
		Mean	Median	SD	Skewness	Best	Worst	Trial Slope	Block Slope	Difficulty Slope
Central Tendency	Mean									
	Median	.943								
Variance and Distributional	SD	.126	.091							
	Skewness	-.184	-.383	.203						
	Best	.875	.755	.511	.147					
	Worst	.896	.800	-.227	-.024	.695				
Contingencies	Trial Slope	-.144	-.122	.356	-.033	-.015	-.279			
	Block Slope	.299	.334	-.315	-.053	.135	.395	-.433		
	Difficulty Slope	-.888	-.845	-.262	.106	-.840	-.752	-.062	-.312	

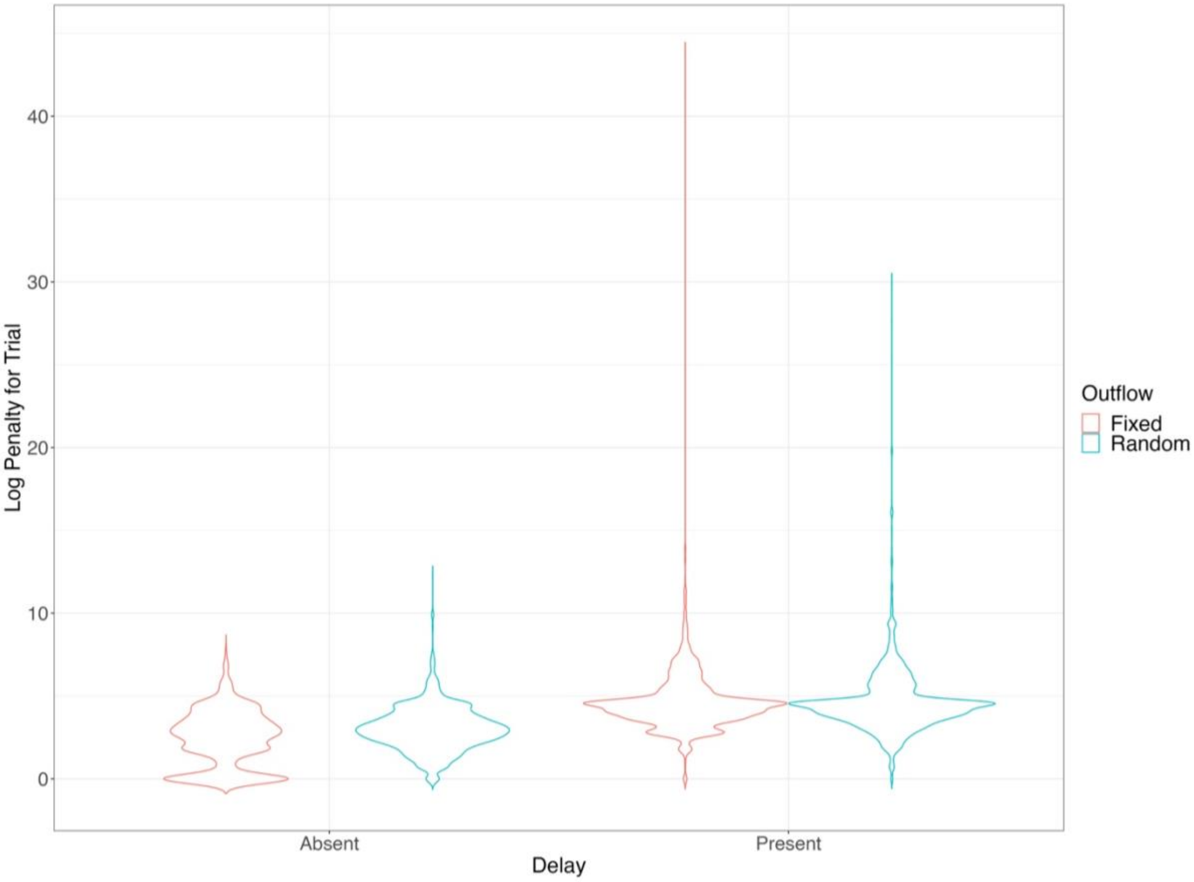
S5. Smoothed within-task performance trajectories for each microworld variant for all participants (Study 2).



S6. Illustration of trial and block interaction (Study 2).



S7. Illustration of delay and outflow interaction (Study 2).



S8. Full intercorrelations between within-person parameters (Study 2).

			Central Tendency										Dispersion																									Contingencies																		
			Mean					Median					SD					MSSD					Skewness					Best Run					Worst Run					Experience		Difficulty		Pre-Change Point Slope					Post-Change Point Slope					Change Point				
			All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4	Trial	Block	Outflow	Delay	All	V1	V2	V3	V4	All	V1	V2	V3	V4	All	V1	V2	V3	V4
Central Tendency	Mean	All																																																				
		V1	.681																																																		
		V2	.784	.731																																																	
		V3	.511	-.116	.040																																															
		V4	.589	-.068	.238	.686																																														
	Median	All	.882	.829	.858	.163	.236																																												
		V1	.633	.971	.666	-.109	-.107	.780																																											
		V2	.758	.757	.971	.019	.173	.849	.702																																										
		V3	.470	.061	.129	.715	.529	.263	.056	.124																																									
		V4	.483	.055	.168	.494	.816	.239	.050	.186	.414																																								
Dispersion	SD	All	.232	-.492	-.120	.735	.731	-.123	-.501	-.196	.338	.429																																						
		V1	.220	.039	.166	.116	.321	.233	.062	.101	.128	.187	.235																																					
		V2	.353	.050	.367	.189	.414	.284	-.009	.225	.099	.136	.361	.419																																				
		V3	.321	-.244	-.079	.784	.493	.007	-.219	-.095	.274	.311	.829	-.022	.083																																			
		V4	.502	-.166	.215	.615	.840	.162	-.205	.111	.334	.504	.872	.296	.445	.655																																		
	MSSD	All	.214	-.263	-.006	.470	.310	.028	-.262	-.078	.059	.039	.767	-.030	.250	.822	.663																																	
		V1	.086	.031	.056	.099	.134	.035	.080	.044	.012	.102	.027	.351	.123	.037	.025	-.082																																
		V2	.154	-.224	.115	.183	.446	.017	-.304	-.040	.089	.112	.498	.347	.820	.138	.553	.349	.074																															
		V3	.153	-.248	-.120	.502	.185	-.017	-.218	-.132	.044	.020	.676	-.146	-.005	.888	.491	.933	-.094	.058																														
		V4	.281	-.241	.156	.382	.462	.089	-.285	.017	.079	.081	.774	.146	.532	.612	.808	.888	-.092	.655	.672																													
	Skewness	All	.361	-.243	.013	.741	.643	-.007	-.264	-.072	.326	.310	.841	.072	.263	.817	.805	.706	.041	.344	.658	.679																														
		V1	-.201	-.714	-.404	.406	.397	-.409	-.752	-.434	.107	.238	.776	.077	.124	.561	.529	.579	-.184	.359	.531	.546	.538																													
		V2	.063	-.195	-.114	.325	.249	-.062	-.176	-.222	.092	.026	.339	.342	.410	.222	.219	.163	.218	.304	.111	.211	.370	.208																												
		V3	.236	-.225	-.107	.640	.433	-.049	-.221	-.153	.145	.225	.663	.043	.153	.723	.554	.560	.100	.134	.582	.460	.801	.439	.340																											
		V4	.353	-.197	.078	.609	.576	.052	-.219	.012	.291	.221	.710	.153	.213	.654	.708	.581	.112	.300	.541	.559	.862	.447	.302	.658																										
	Best Run	All	.622	.907	.672	-.105	-.071	.723	.875	.717	.020	.065	-.487	-.164	.019	-.206	-.183	-.242	.109	-.213	-.212	-.253	-.223	-.702	-.211	-.214	-.177																									
		V1	.609	.923	.648	-.117	-.089	.724	.883	.685	.003	.050	-.492	-.209	.014	-.209	-.194	-.237	.038	-.221	-.207	-.245	-.230	-.688	-.210	-.193	-.205	.973																								
		V2	.722	.728	.922	.034	.125	.812	.674	.957	.158	.140	-.207	.069	.076	-.080	.086	-.070	.019	-.092	-.102	-.008	-.072	-.419	-.274	-.192	.038	.703	.656																							
		V3	.339	-.141	-.025	.827	.605	.056	-.141	-.043	.817	.427	.504	.138	.114	.417	.445	.153	.123	.141	.140	.169	.487	.282	.240	.369	.422	-.132	-.147	-.004																						
		V4	.463	-.140	.102	.691	.898	.154	-.159	.087	.530	.857	.646	.180	.242	.490	.639	.249	.189	.262	.199	.292	.507	.412	.193	.375	.423	-.115	-.133	.046	.643																					
	Worst Run	All	.622	-.063	.269	.781	.852	.269	-.097	.179	.443	.555	.869	.304	.466	.758	.918	.642	.074	.495	.531	.716	.848	.459	.301	.648	.717	-.091	-.097	.150	.515	.682																				
		V1	.665	.950	.717	-.117	-.023	.811	.919	.717	.080	.061	-.422	.234	.124	-.254	-.095	-.236	.026	-.128	-.260	-.168	-.216	-.643	-.143	-.225	-.153	.768	.776	.689	-.144	-.127	-.006																							
		V2	.755	.646	.946	.073	.312	.805	.586	.865	.135	.193	-.025	.255	.552	-.060	.293	.045	.079	.256	-.117	.262	.082	-.346	.050	-.020	.104	.550	.546	.760	.006	.149	.347	.665																						
		V3	.499	-.134	.030	.932	.651	.153	-.125	.009	.520	.481	.798	.086	.169	.903	.665	.623	.066	.175	.672	.487	.795	.469	.307	.730	.627	-.125	-.133	.013	.595	.635	.839	-.130	.063																					
		V4	.603	-.024	.290	.620	.956	.258	-.072	.200	.470	.689	.720	.386	.484	.455	.890	.330	.086	.521	.170	.537	.670	.345	.260	.429	.617	-.045	-.056	.155	.525	.741	.883	.037	.372	.602																				
Contingencies	Experience	Trial Slope	.602	.184	.275	.678	.488	.355	.150	.306	.480	.403	.519	-.171	.080	.666	.503	.524	-.070	.050	.547	.391	.572	.296	-.007	.467	.487	.165	.177	.297	.403	.492	.621	.156	.232	.722	.432																			
		Block Slope	-.342	.025	-.091	-.494	-.494	-.101	.013	-.100	-.181	-.495	-.492	-.065	-.153	-.522	-.430	-.354	.031	-.139	-.370	-.273	-.493	-.321	-.150	-.444	-.389	-.021	.015	-.081	-.256	-.446	-.553	.028	-.103	-.547	-.443	-.447																		
	Difficulty	Outflow Slope	-.053	-.518	.057	.073	.593	-.212	-.567	-.029	.035	.390	.501	.298	.400	.039	.565	.130	.074	.535	-.078	.420	.261	.475	.136	.070	.289	-.473	-.514	-.073	.151	.491	.399	-.438	.149	.067	.595	-.127	-.112																	
		Delay Slope	-.231	-.814	-.588	.597	.550	-.570	-.794	-.620	.317	.389	.803	.072	.066	.594	.538	.425	.049	.296	.418	.388	.601	.789	.314	.516	.487	-.746	-.756	-.611	.527	.593	.513	-.767	-.488	.592	.471	.229	-.354	.501																
	Pre-Change Point Slope	All	.371	.483	.486	-.073	-.081	.453	.490	.474	-.027	-.092	-.219	.010	.189	-.103	-.038	-.051	.126	.042	-.101	.009	-.187	-.368	-.017	-.168	-.154	.510	.498	.467	-.138	-.170	-.015	.410	.438	-.062	-.016	.117	-.025	-.202	-.462															
		V1	.210	.504	.313	-.151	-.352	.369	.538	.306	.021	-.394	-.408	.085	.030	-.228	-.319	-.193	.025	-.150	-.156	-.207	-.265	-.577	.025	-.278	-.171	.417	.426	.300	-.188	-.450	-.216	.496	.293	-.186	-.247	-.043	.111	-.482	-.545	.413														
		V2	.332	.356	.577	-.118	-.029	.454	.346	.557	-.138	-.043	-.139	.028	.353	-.091	.014	.022	.017	.163	-.055	.123	-.119	-.266	-.077	-.037	-.066	.368	.343	.464	-.282	-.083	.042	.332	.577	-.061	.009	.113	-.001	.023	-.408	.528	.365													
		V3	.190	.049	.066	.279	.210	.072	.059	.100	.216	.166	.136	.194	.076	.129	.157	-.015	.230	.085	-.001	-.022	.060	-.006	.135	.017	.072	.126	.066	.121	.318	.215	.183	.048	.026	.206	.191	.047	-.143	.019	.104	.233	-.002	-.045												
		V4	.197	.029	.175	.139	.261	.137	-.014	.108	.105	.153	.217	.246	.295	.074	.311	.138	-.008	.365	.016	.285	.141	.163	.013	-.005	.058	.059	.037	.142	.110	.133	.293	.045	.183	.142	.324	.022	-.072	.246	.067	.307	.049	.183	.240											
	Post-Change Point Slope	All	-.037	-.322	-.241	.339	.021	-.140	-.312	-.237	-.026	-.084	.530	-.223	-.116	.712	.303	.816	-.199	-.039	.892	.547	.530	.527	.014	.501	.437	-.307	-.290	-.215	.054	.090	.332	-.312	-.228	.488	-.018	.441	-.268	-.106	.386	-.433	-.218	-.197	-.168	-.165										
		V1	-.170	-.509	-.276	.219	.386	-.326	-.589	-.271	.064	.424	.453	-.086	.050	.247	.348	.220	-.183	.215	.174	.245	.285	.674	-.052	.240	.196	-.442	-.428	-.261	.264	.478	.255	-.504	-.262	.233	.276	.171	-.160	.473	.557	-.460	-.911	-.366	-.063	-.044	.269									
		V2	-.327	-.362	-.589	.163	-.015	-.434	-.348	-.543	.187	-.028	.131	.001	-.370	.103	-.080	-.240	.009	-.235	.010	-.262	.074	.331	.090	.062	.049	-.398	-.347	-.483	.286	.061	-.101	-.343	-.592	.107	-.056	-.001	.034	-.098	.475	-.599	-.276	-.804	-.071	-.237	.503	.382								
		V3	.014	-.111	-.059	.103	-.174	.040	-.096	-.080	-.153	-.292	.322	-.229	-.061	.519	.196	.773	-.248	-.054	.807	.553	.389	.297	-.060	.378	.312	-.123	-.104	-.064	-.173	-.179	.194	-.114	-.042	.269	-.140	.337	-.177	-.216	.077	-.107	.018	.017	-.359	-.083	.867	.028	-.066							
		V4	-.036	.256	.046	-.241	-.376	.099	.280	.109	-.124	-.190	-.457	-.308	-.259	-.232	-.456	-.395	-.200	-.371	-.175	-.383	-.299	-.349	-.131	-.094	-.243	.224	.261	.106	-.300	-.324	-.339	.211	-.013	-.221	-.383	.049	-.019	-.438	-.381	.028	.088	.077	-.297	-.624	.229	-.096	-.021	.325						
	Change Point	All	-.159	-.481	-.457	.382	.363	-.376	-.466	-.456	.319	.313	.459	.072	-.122	.333	.288	.202	-.030	.069	.226	.142	.409	.453	.105	.351	.334	-.508	-.505	-.432	.423	.443	.282	-.415	-.411	.344	.271	.153	-.141	.227	.634	-.828	-.433	-.524	-.066	-.254	.386	.421	.544	.075	-.182					
		V1	-.084	.031	.094	-.322	-.110	-.027	.005	.078	-.317	-.141	-.215	-.249	.056	-.241	-.084	-.112	.061	.008	-.170	-.023	-.105	-.270	-.221	.003	-.069	.134	.098	.043	-.313	-.144	-.168	-.015	.095	-.267	-.080	-.230	.150	.135	-.171	-.060	-.307	.079	-.137	-.255	-.078	.120	-.078	-.014	.207	-.023				
		V2	-.321	-.182	-.269	-.263	-.186	-.323	-.182	-.257	-.263	-.063	-.121	-.116	-.067	-.219	-.151	-.091	.113	.026	-.193	-.086	-.093	.050	-.018	-.126	-.138	-.100	-.107	-.234	-.131	-.164	-.216	-.196	-.255	-.278	-.173	-.294	.031	.018	.070	-.264	-.202	-.403	.008	-.093	.240	.209	.125	.026	-.156	.115	.014			
		V3	.049	-.290	-.187	.432	.346	-.147	-.274	-.248	.274	.249	.412	-.029	.059	.385	.297	.218	-.062	.115	.216	.212	.447	.289	.158	.408	.355	-.323	-.297	-.250	.350	.329	.357	-.277	-.100	.416	.313	.295	-.235	.174	.481	-.318	-.177	-.225	-.667	-.139	.231	.269	.349	.133	.025	.419	-.126	-.035		
		V4	-.033	-.237	-.189	.213	.265	-.187	-.200	-.177	.156	.222	.274	-.005	-.102	.237	.223	.028	.097	-.032	.108	.007	.270	.131	.080	.246	.322	-.246	-.258	-.209	.254	.315	.150	-.205	-.150	.171	.200	.150	-.109	.174	.366	-.361	-.229	-.257	-.032	-.716	.228	.235	.297	-.174	.082	.569	.164	.111	.319	