Investigating the Elaborated Relational Abilities Index as a Novel Measure of Cognitive Ability

Carina Michelle Kaufmann, B.A. (Hons)

Thesis submitted in fulfilment of a Master of Science (M.Sc.) degree to the Department of Psychology, Maynooth University.

October 2024

Supervised by Prof. Bryan Roche and Dr. Jamie Cummins

Head of Department: Prof. Deirdre Desmond

Table of Contents

Table of Contents ..... **ii**
List of Tables ..... **v**
List of Figures ..... **vi**
Declaration ..... **vii**
Acknowledgements ..... **viii**
Abstract ..... **ix**
Chapter 1 ..... **1**
Literature Review and General Introduction **1**

1.1 Introduction **2**
1.1.1 General Introduction 2
1.1.2 Definition and Conceptualisation of Intelligence 4
1.1.2.1 Spearman 4
1.1.2.2 Horn-Cattel Carroll Theory 6
1.1.2.3 Gardner's Theory of Multiple Intelligences 7
1.1.2.4 Consensus 8

1.2 Overview of Intelligence Testing **9**
1.2.1 Binet-Simon 9
1.2.2 Stanford-Binet 11
1.2.3 US Army Alpha and Beta Tests 12
1.2.4 Performance IQ 12
1.2.5 Wechsler Scales of Intelligence 14
1.2.6 National Adult Reading Test 16
1.2.7 IQ Scoring 17

1.3 Is Intelligence an Inherent Trait or an Acquired Skill? **18**
1.3.1 Nature 18
1.3.2 Nurture 22

1.4 Controversies and Issues Surrounding Intelligence Testing and Theory **28**
1.4.1 The Computation and Meaning of Test Scores 28
1.4.2 Erroneous Conclusions Based on the Positive Manifold 32

1.5 A Relational Frame Theory Perspective on Intelligence **37**
1.5.1 Intelligence in Behaviour Analysis 37
1.5.2 Relational Frame Theory 38
1.5.3 The RFT Account of Human Intelligence 54

1.6 Enhancing Intellectual Performance **60**
1.6.1 Previous Attempts to Enhance IQ 60
1.6.2 Interventions Based on Relational Frame Theory 63
1.6.3 PEAK 64
1.6.4 SMART 65
1.6.4.1 What is SMART? 65
1.6.4.2 Evidence for the SMART Interventions 69

1.7 Measuring Relational Skills: The Relational Abilities Index **74**
1.7.1 First Relational Abilities Index 75
1.7.2 Development of the RAI+ 78
1.7.3 Development of the Elaborated RAI 79

1.8 The Current Thesis **82**
Chapter 2 .... **86**
Assessing the Psychometric Properties of the Elaborated Relational Abilities Index **86**
2.1 Research Questions **87**
1.1.2 RQ1. What is the convergence between the RAI and RSPM? 87
2.1.2 RQ2. How do the variables correlate with each other? 87
2.1.3 RQ3. How does socioeconomic status predict performance on the RAI and RSPM? 87
2.2 Methodology **88**
2.2.1 Participants 88
2.2.2 Settings and Materials 91
2.2.2.1 Demographic Questionnaire 91
2.1.2.2 Elaborated Relational Abilities Index 94
2.1.2.3 Raven's Standard Progressive Matrices 96
2.2.3 Procedure 98
2.2.4 Ethical Considerations 98
2.3 Results **99**
2.3.1 Missing Data and Excluded Cases 99
2.3.2 Descriptive Statistics 100
2.3.3 Reliability 101
2.3.4 Inferential Statistics 102
2.3.4.1 RQ1. Convergence Between the RAI and RSPM 102
2.3.4.2 RQ2. Correlation Matrix 102
2.3.4.3 RQ3. SES as a Predictor of RAI and RSPM Performance 104
2.4 Discussion **104**
2.4.1 Summary 104
2.4.2 Convergence Between the RAI and RSPM 105
2.4.3 SES as a Predictor of RAI and RSPM Performance 108
Chapter 3 .... **115**
An In-Depth Analysis of the Elaborated Relational Abilities Index: Further Relationships Between Test Subscales, Overall Test Scores, Socio-Economic Status, and a Factor-Analysis of Test Structure Using Primary and Secondary Data **115**
3.1 Study 2 .... **116**
3.1.2 Research Questions 117
3.1.2.1 RQ1. How do the individual RAI subscales correlate with RSPM? 117
3.1.2.2 RQ2. How does SES predict performance on the individual RAI subscales? 118
3.1.2.3 RQ3. How does SES predict fluency on the full RAI, RSPM, and the individual RAI subscales? 118
3.1.2.4 RQ4. Does the RAI's current composition confirm to its underlying factor structure? ... 118
3.2 Methodology **118**
3.2.1 Participants 118
3.2.2 Settings and Materials 119
3.2.3 Ethical Considerations 119
3.3 Results **120**
3.3.1 Descriptive Statistics 120
3.3.2 Inferential Statistics 120
3.3.2.1 Correlational Matrix 120
3.3.2.2 RQ1. Correlation Between Individual RAI Subscales and RSPM Performance 123
3.3.2.3 RQ2. SES as a Predictor of Individual RAI Subscale Performance 124

3.3.2.4 RQ3. SES as a Predictor of RAI, RSPM, and Individual RAI Subscale Fluency 127
3.3.2.5 RQ5. RAI Factor Structure 129

3.4 Discussion 135
3.4.1 Summary 135
3.4.2 Convergence Between the RAI Subscales and RSPM 136
3.4.3 SES as a Predictor of RAI Subscale and RSPM Performance and Fluency 143
3.4.3 The RAI's Factor Structure 146

Chapter 4 151
General Discussion 151

4. Discussion 152
4.1 Research Summary and Main Findings 152
4.6 Limitations of the Current Study 156
4.7 Implications for Future Research 158
4.8 Conclusion 159

References 161
Appendices 212

List of Tables

Table 1. Distribution of Participants' Own and Parental Education 89
Table 2. Participants' Own and Parental Income and Education Levels 93
Table 3. Descriptive Statistics for Total RAI, RAI Subscales, and RSPM Accuracy 101
Table 4. Cronbach's Alpha Coefficients for the Full RAI and Its Individual Subscales 101
Table 5. Spearman's rho Correlation Matrix for RAI and RSPM Accuracy Scores and Demographic Variables 103
Table 6. Descriptive Statistics for Full RAI, RAI-M, RSPM, and the RAI Subscales Fluency. 120
Table 7. Spearman's rho Correlation Matrix for RAI and RSPM Fluency Scores and Demographic Variables 122
Table 8. Confirmatory Factor Analysis Model Fit Statistics of the RAI 131
Table 9. Standardised Factor Loadings (Standard Errors) for the Eight-Factor Model of the RAI 133
Table 10. Factor Correlations (Standard Errors) for the Eight-Factor Model of the RAI 134

List of Figures

Figure 1. Familial Network of Established and Derived Relations 41
Figure 2. Distribution of Parental Income 90
Figure 3. Distribution of Own Income 90
Figure 4. RAI Trials From Each Subscale 95
Figure 5. Sample Items From RSPM 97

Declaration

I, the undersigned, hereby certify that this material, which I now submit in fulfilment of an M.Sc. degree, has not been previously submitted as an exercise for a degree at this or any other University, and is, unless otherwise stated, entirely my own work.

Signed: ____________________________

Carina Michelle Kaufmann

Student ID: 23253139

Date: 30 October 2024

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisors, Prof. Bryan Roche and Dr. Jamie Cummins. It has truly been a privilege to work with you and experience your guidance, kindness, patience, and support. Everyone around me knows how incredibly inspiring it is to work with supervisors as knowledgeable and kind as you. Particularly, the time, energy, care, and attention to detail afforded by Prof. Bryan Roche has exceeded all expectations, and for that, I am deeply thankful. I would also like to express my heartfelt gratitude for the invaluable support with PhD applications, and the helpful insights into the world of academia that you took the time to provide me with as well.

Thank you to my friends and siblings. Especially Mia and Renah, who have always been by my side and not only celebrate all of my successes as their own but also support me through every difficulty. I genuinely hope that I am reciprocating that love daily and cannot express enough how much I appreciate having you in my life. Thank you also to Lauren, who has made this year so much fun and also provided helpful insights into many personal and academic issues.

I would like to thank the Psychology Department and Research Postgrads for welcoming me and making this journey less scary. Aisling, we might never solve the issue around the moon and the sun's genders, but hopefully we'll get closer to solving the issues at the centre of our theses. Jennifer, your warmth and positivity undoubtedly brighten everyone's day and make even the most stressful days more enjoyable. I would also like to express my sincere gratitude to my undergraduate supervisor, Dr. Dylan Colbert, who not only has ignited my interest in Relational Frame Theory but whose encouraging words and helpful techniques still get me through times of excessive worry.

Finally, I am incredibly grateful to my parents, who have sacrificed so much and worked tirelessly to give us the opportunities they could never avail of. Thank you also for instilling the right attitudes and values necessary to get me to where I am today. Particularly my mother, who unfortunately never got to finish the education she so desired and deserved; I hope I can continue to make you proud. Since my parents don't speak English, I would like to thank them in our mother tongue: Спасибо большое за всё, Мама и Папа! Без вас это всё не было бы возможно. Я надеюсь что вы сможете так и дальше мной гордиться!

Abstract

The need to reconceptualise intelligence in functional terms becomes apparent when reviewing traditional intelligence theories and testing methodologies. Relational Frame Theory offers a functional understanding of the intelligence concept by proposing a definition of it in terms of a well-understood behavioural process known as derived relational responding. While a growing number of researchers from different fields are coming to similar conclusions, what is needed to complement the Relational Frame Theory account is a functionally understood assessment of derived relational responding proficiency. The current thesis represents an in-depth experimental and psychometric analysis of the most recent version of the Relational Abilities Index (RAI; Cummins et al., 2023). Firstly, it was examined in terms of its convergence with a widely-used proxy measure of intelligence, in what was an entirely exploratory analysis rather than an effort to establish convergent validity. Both the full RAI and all its subscales significantly correlated with performance on Raven's Standard Progressive Matrices. Moreover, the current thesis analysed the impact of four sociodemographic variables on RAI scores in order to quantify the assumed impact of environmental factors on relational responding proficiency. Of the variables examined, only parental education significantly predicted performance on the RAI, a finding that aligns with the literature on the importance of parental education levels on childhood cognitive functioning. Finally, the first-ever factor analysis of the RAI was conducted. The RAI's current eight-factor model was found to be a poor fit, while better than the single-factor model. The implications of these findings for the further development of the RAI are discussed.

Chapter 1

Literature Review and General Introduction

1.1 Introduction

1.1.1 General Introduction

Intelligence testing is among the most widely studied but arguably the most controversial subject in the study of psychology. A brief review of the history of intelligence testing offers a good explanation for criticism and scepticism of such intelligence tests among both the public and academic researchers. The tests were originally developed in France to help identify struggling students in order to provide additional support, rather than to simply deny such children's social opportunities and label them as “sick” (Nicolas et al., 2013). The introduction of intelligence tests in the United States quickly led to their use for various ethically questionable purposes. The use of tests for the purpose of reducing feeble mindedness amongst the US population arose from a US Supreme Court’s legal ruling known as Buck vs Bell. The latter was a eugenicist movement in the US that had sought to use intelligence testing to identify individuals who should be barred from procreation. They used the court ruling in their favour to conduct over 65,000 coerced sterilisations of “intellectually inferior” or “feebleminded” individuals. This activity later directly inspired the eugenicist movement of the Nazis in Germany (Martschenko, 2018). Further, intelligence test results have been used for immigration policies (Knox, 1914b) and to support claims of racial superiority on several occasions throughout time and across the world (Martschenko, 2018). Moreover, the US Army administered intelligence tests to approximately 1.75 million draftees during the First World War. Currently, some districts in the US still screen out individuals above threshold test scores in their police forces.

The foregoing issues notwithstanding, intelligence tests continue to be used for their original purpose to identify learning difficulties or gifted students in order to provide appropriate interventions or even protect them from execution within the US legal system (Martschenko, 2018). Moreover, intelligence tests have helped identify environmental factors (such as structural inequalities or harmful substances) resulting in developmental delays or impairment, such as that from lead exposure (e.g., Bellinger et al., 1992; Rawat et al., 2022). However, controversies regarding intelligence tests extend beyond those relating to their use. Both the general population and experts disagree on what precisely intelligence is, as well as its main determinants, with some arguing it is primarily genetic, whereas others argue it is mainly influenced by environmental factors (Rinderman et al., 2016). The issue of nature versus nurture is as prominent in intelligence research as it is in other areas in psychology. A summarised overview will be provided in section 1.3. The current thesis attempts to articulate with this issue to some extent by examining the viability of a novel means of assessing intelligence conceived quite differently than it has been to this point. In so doing, the thesis will examine the relative impact of social opportunity on scores on this test. By approaching intelligence from a modern behavior-analytic perspective as an acquired skill set which may or may not be limited by genetically determined variables, the current research will redirect the use of intelligence testing to more equalitarian ends and from within a more egalitarian conceptual framework. At the very core of this thesis is therefore the question of whether intelligence should be understood as an inherent trait or viewed as intelligent behaviour in context, where it is understood as an acquired skill. While many of the popular theorists (e.g., Eysenck, 1971; Goleman, 1995; Herrnstein, 1971; Jensen, 1969; Haier, 2014) subscribe to the former view, others (e.g., Dickens & Flynn, 2001; Schlinger, 2003; Sternberg, 2008) have expressed more openness to the later idea. In order to better understand these viewpoints, an overview of the history of intelligence testing seems warranted before both sides are discussed. To provide such context, the following will first outline the issues around the definition and conceptualisations of intelligence, followed by an investigation of traditional intelligence testing and its inadequacies before introducing Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001) as the theoretical basis for the current research. The subsequent section will then outline RFT-based methods to enhance relational skills, with a particular focus on the Strengthening Mental Abilities with Relational Training (SMART) intervention, before an outline of the history of the Relational Abilities Index (RAI); the "intelligence" measure at the focus of the current thesis, is provided. Finally, a brief overview of demographic variables impacting intelligence test performance will be provided before the aims of the current research are outlined in more detail.

1.1.2 Definition and Conceptualisation of Intelligence

1.1.2.1 Spearman

As will become evident in the subsequent section on intelligence testing, the driving force behind the development of early intelligence tests was pragmatism rather than theory (White & Hall, 1980). However, the single most influential theory of intelligence, Spearman's (1904) conceptualisation of a single factor of intelligence, eventually became the impetus behind the development of subsequent intelligence tests and the dominant approach in the academic study of intelligence (discussed in detail in section 1.2).

Spearman's conceptualisation is based on a factor analysis he conducted on the performance of a number of tasks, which he argued measure intellectual performance (Spearman, 1904; 1927). Factor analysis is a mathematical method whereby correlations between large numbers of items are analysed, thereby allowing for the identification of underlying factors serving as an explanation for these correlations. As this analysis revealed no inverse relationships between the different factors, he termed this finding the "positive manifold", providing the basis for Spearman's conclusion of the necessary existence of an underlying factor, which he described as *mental energy*, or, more prominently, *g*, for *general intelligence*. The three constituents of *g*, according to Spearman, are expressed in his three *noegenetic laws*. The first is the apprehension of experience (referring to the ability to process and understand information on a very basic level, which will further guide problem-solving), the eduction of relations (referring to the ability to identify relationships between stimuli), and the eduction of correlates (the ability to derive relationships between these stimuli; Spearman, 1927). To simplify, the apprehension of experience allows for the identification of available information through simple observation, the eduction of relations allows for the identification of explicit pre-existing relations, and the eduction of correlates allows for the deduction of implicit relations. Hence, posited Spearman, intellectual performance is explained by an individual's quantity of *g* (Spearman, 1904; 1927).

Contrasting with Binet and Simon's (1905) approach (outlined in section 1.2.1.1), Spearman argued that intelligence is resistant to improvement or change. This essential assumption directed the focus of intelligence research towards an essentialist and mentalistic approach (Herrnstein & Murray, 1994), proposing there to be a latent factor or faculty determining an individual's performance on all intelligence measures (Colbert, 2015). Moreover, it reduced the construct from a functional account (as approached by Binet & Simon, 1905) to an essentialist, single-factor theory describing a hypothetical construct rather than a repertoire of observable skills (Colbert, 2015). Spearman's theory encountered several criticisms, particularly regarding his factor analysis to arrive at his conclusions, as well as general conceptual issues with *g* as an inherent trait, that will be discussed in greater detail in section 1.2.2.

1.1.2.2 Horn-Cattel Carroll Theory

Despite being met with heavy criticism, Spearman's proposal of the single underlying factor *g* is the foundation of the most widely cited theory of intelligence – the Cattell-Horn theory (Horn & Cattell, 1967). This theory agrees with the assumption of a single factor of intelligence but argues this to be constituted of two components: fluid intelligence (*Gf*) and crystallised intelligence (*Gc*). *Gf* refers to an individual's ability to learn efficiently, problem-solve, and form short-term memories, all of which allow effective adaptation to the environment. These abilities are argued to be independent of education and culture (Horn & Cattell, 1967). Thus, argue Horn and Cattell (1967), *Gf* is inherent and rather independent of education and culture. Consequently, *Gf* is associated with various relatively stable cognitive abilities (Varriale et al., 2018). In contrast, *Gc* reflects an individual's acquired knowledge, thus encompassing vocabulary and related acquired skills (Deary et al., 2007; Horn & Cattell, 1967; Yuan et al., 2018). The theory's terminology utilises a water analogy to reflect the relationship between *Gf* and *Gc*, whereby *Gf* is supposed to crystalise into *Gc* under appropriate conditions (i.e., learning opportunities). This close relationship between crystallised and fluid intelligence is supported by a close statistical correlation (.50) between the two components (Cattell, 1934; Li et al., 2004). It follows that *Gf* peaks and then drops off in early adulthood (around the age of 25), whereas *Gc* keeps rising throughout the lifespan, though the degree declines with age, as the ability and ease of learning reduce while an individual's acquired knowledge constantly increases (Aichele et al., 2015; Alloway, 2012; Tucker-Drob, 2009).

The Horn-Cattell model, too, has been met with criticism. For instance, Brody (1992) and Carroll (1993) criticised the theory for missing a hierarchical conceptualisation with the *g* factor at the top and G*f* and G*c* as components of this higher order. This, argued Carroll (1993), would be required to account for the correlation between *Gf* and *Gc*. As a result of this criticism, McGrew (1997) argued that it was appropriate to synthesise the Cattell-Horn theory with Carroll's (1993) Three-Stratum theory. The Three-Stratum Theory proposes three "layers" that become increasingly narrower in the abilities they relate to. According to McGrew (1997), there is strong correspondence between the two theories, and later factor analytic research indeed provided support for such a synthesis (Caemmerer et al., 2020; Floyd et al., 2010; Jewsbury et al., 2016; Keith & Reynolds, 2010; Salthouse, 2005). In short, this Cattell-Horn Carroll (CHC) theory proposes a three-stratum model of cognitive abilities with *g* at the model's apex in the third stratum, followed by eight broad abilities at the second stratum, which split into further detail at the first stratum, which consists of many narrow abilities. The second stratum abilities include crystallised and fluid intelligence (as introduced in the Cattell-Horn theory), auditory perception, visual perception, retrieval ability, memory and learning, cognitive speediness, and processing speed. This new hybrid model is widely accepted by most contemporary theorists (Flanagan et al., 2000; Flanagan & Dixon, 2014). Indeed, McGrew (2005) and Kaufman et al. (2012) argue that, through the integration of the conceptualisation and operationalisation of cognitive abilities, this newer hybrid model was particularly influential in the development of modern intelligence tests, consequently realising McGrew's (1997) proposal that the CHC model should function as a foundation for future test development.

1.1.2.3 Gardner's Theory of Multiple Intelligences

Contrasting the predominant psychometric approach summarised to this point is Gardner's widely influential and popular (particularly amongst the general population) theory of *Multiple Intelligences*. This argues that there are eight different types of intelligences, each relating to a different modality (Gardner, 1983, 1997). These, according to Gardner, are interpersonal, intrapersonal, logical-mathematical, naturalist, spatial, bodily-kinaesthetic, linguistic, and musical. Such a conceptualisation of intelligence allows for (almost) everyone to be deemed intelligent in some regard, making it easy to understand why the theory gained such popularity amongst the general population. However, many researchers argue it more accurate to refer to those suppositional intelligences as “mere” *systems of abilities* or simply *skills*, as they scarcely, if at all, accurately reflect the traditional definition of intelligence in the sense in which it relates to cognitive performance (White & Bren, 1998). Additionally, the theory was criticised for its lack of data supporting the proposition of such a conceptualisation (Waterhouse, 2006), an issue further culminated in the fact that, in the four decades since its first proposal, it has failed to produce any tests to measure the proposed “intelligences”.

1.1.2.4 Consensus

As demonstrated above, there is significant controversy surrounding the definition, as well as the conceptualisation of intelligence. This controversy probably found its greatest expression in the publication of Herrnstein and Murray’s (1994) *The Bell Curve* and motivated the establishment of an American Psychological Association’s (APA) Board of Scientific Affairs task force appointed to prepare a report on the consensus on the definition, measurement and common theories and issues surrounding intelligence (Neisser et al., 1996). Therefore, the APA definition aimed to capture a consensus in defining intelligence as *the ability to understand complex ideas, effectively adapt to the environment, learn from experience, engage in different forms of reasoning, and overcome obstacles through thought* (Neisser et al.,1996, p. 77). At the same time, the controversy surrounding the definition and conceptualisation of intelligence led most theorists to argue that a theoretical definition of intelligence was not (yet) possible and that, therefore, an operational definition must suffice (Richardson & Norgate, 2015). Specifically, the operational definition states that "intelligence is whatever intelligence tests measure" (Boring, 1961). However, this intentionally pragmatic approach has been met with numerous criticisms, even to the point of arguing that the current measurement of intelligence is inadequate. As the criticisms of intelligence theory and intelligence testing are highly correlated, before discussing the inadequacies of intelligence testing, a precis of how intelligence tests work should be provided. To facilitate this, a chronological summary of the history of intelligence tests is warranted. It is important for the reader to be aware, that intelligence as a construct is to be distinguished from IQ scores as a metric. That is, an IQ score could be calculated for an individual based on any number of theoretical perspectives or test types. While such a distinction is important, this does not take away from the fact that IQ scores appear to be a good indicator of intellectual performance. However, in the following section, the emphasis will be primarily on the method of measurement, although this sometimes inevitably interfaces with the theoretical approach underlying that measurement technique.

1.2 Overview of Intelligence Testing

1.2.1 Binet-Simon

While not the first test of cognitive abilities (Peterson, 1925), the Binet-Simon scale is widely thought of as the first practically useful measure of intelligence. Its development was motivated by the aim to be able to identify children in need of special education (White & Hall, 1980). Tasked with the goal of providing appropriate resources to students who were falling behind on the educational curriculum, Binet and Simon studied individual differences between school children and finally comprised a scale of thirty items measuring a collection of cognitive abilities such as memory, reasoning, and comprehension (Binet, 1904a; Binet, 1904b). The items used in their scale were either already existent, like the Digit Span (explained below), or developed by Binet and Simon themselves (Boake, 2002). Importantly, Binet and Simon (1916) asserted that the individual tests, and consequently performance on the individual tests, are of little to no meaning. What matters is the average performance on all the assessed measures.

Additionally, the Binet-Simon scale, published in 1905, included instructions regarding the test's administration, scoring, and interpretation (White & Hall, 1980). One of the main contributions Binet and Simon made was the idea to conceptualise intelligence in terms of higher psychological processes while disregarding the historical focus on sensory processes (Binet & Henri, 1895; White & Hall, 1980). Moreover, as reflected in his initial motivation to devise such a scale, Binet believed intellectual development to be malleable (within limits) when subjected to the appropriate interventions (Siegler, 1992). A significant change in the scoring of the scale implemented in Binet and Simon's later revisions in 1908 and 1911 was the adaptation to an age scale, as opposed to the previously used point scale (White & Hall, 1980). While the latter views performance relative to the maximum score of the scale, the former interprets performance relative to the average of other individuals in the test-taker's age group. If the average number of tests a child completed successfully was below the general average for that age group, their *mental age* (the metric used to indicate performance) would reflect that of the respectively comparable age group and consequently fall below the child's chronological age. Reversely, if a child performed above average for their age group, their mental age would exceed their physical age.

Many subsequent measures built on the Binet-Simon scales and were used as measures of intelligence for various purposes (outlined in the introduction above). The Binet-Simon scale not only served as a foundation for the form of subsequent tests but also for their content, with many subsequent tests consisting of an array of cognitive tests combined to allow for a composite score

(Boake, 2002). However, Binet himself did not regard his measure as an adequate assessment of intelligence, criticising its failure to assess emotional intelligence and creativity (Martschenko, 2018). Instead, the intention was the identification of students requiring additional educational support, rather than the precise measurement of intelligence as a single faculty (Gregory, 1994; Vial & Hugon, 1998). This motivation might further serve as an explanation for the test's lack of an underlying theory. While Binet himself never postulated a theory of intelligence, his general ideas and interpretations of the concept, namely that intelligence is malleable if exposed to appropriate educational opportunities, are reflected in the aforementioned motivations to devise such a measure.

1.2.2 Stanford-Binet

In 1916, Lewis Terman published a revised, American version of the original Binet-Simon scale (White & Hall, 1980). In this version, he added new tests to the original scale (Boake, 2002). However, more importantly, while keeping the scoring based on age, the Stanford-Binet scale now introduced the unit of *Intelligence Quotient* (IQ) scores. IQ scores were computed by dividing the mental age (as measured by the Binet-Simon scale) by the test taker's chronological age and multiplying this quotient by 100 to get rid of fractions. The additional revision in 1960 further introduced two changes, with the most significant one being the standardisation of scores. This meant that the scoring no longer reflected the grouping of items according to the average age at which children pass these specific items (as in the Binet-Simon version), or as IQ scores as in the first Stanford-Binet version, but instead, IQ scores now reflected a norm-referenced score based on the deviation from the mean (White & Hall, 1980). As such, the mean was set to a score of 100 while the standard deviation was set to 16 points for each age group.

1.2.3 US Army Alpha and Beta Tests

The recruitment strategy of the US Army during the First World War called for an easily and quickly administrable measure of intelligence (White & Hall, 1980). The solution was seen in group tests used for the elimination of unsuitable candidates and sending suitable candidates to training. In 1917, within about eight months, a committee of psychologists, including Terman, devised the Army Alpha. While there were multiple versions of the Army Alpha, their commonality lay in their inclusion of about 200 multiple-choice items primarily focusing on verbal skills like analogies, disarranged sentences, and synonym-antonym combinations. Following the war, group tests of intelligence were so widely accepted that they became part of the school curriculum in many US American schools (White & Hall, 1980).

Though distinct in their form of items, another group assessment of intellectual functioning is the Scholastic Aptitude Test (SAT) developed in 1926 and still widely used in the US. The SAT assesses verbal and quantitative abilities and is used as a predictor of success in a college environment (White & Hall, 1980).

1.2.4 Performance IQ

The desire to assess intellectual function despite possible language barriers due to a migrant background or a lack of education led to criticisms of the Binet-Simon scale (e.g., Healy & Fernland, 1911), and served as the motivation to construct measures that would now fall into the category of *performance IQ* (Boake, 2002). Performance IQ refers to those intellectual abilities that are independent of verbal abilities and are hence sometimes referred to as *nonverbal IQ*. Such tasks were also utilised in the assessment program at Ellis Island to screen arriving immigrants for psychological and physiological disorders (Knox, 1914b). Examples of tests that would fall into the category of Performance IQ are Raven's Progressive Matrices (RPM; Raven, 1940; Raven & Court, 1998) or tests like the Digit Symbol coding Task (Wechsler, 1997).

Performance IQ tests can be interpreted as tests of specific abilities. An example of such a specific ability is visuospatial reasoning, as assessed in the various versions of RPM (Raven, 1941; Raven, 1966; Raven, Court, & Raven, 1994; Raven, Raven, & Court, 1962). Raven's matrices are a visuospatial measure of analogical reasoning where the test-taker is presented with a three-by-three matrix of patterns with the bottom-right pattern missing. The test-taker is then tasked to identify the missing pattern out of the six to eight presented options. Such an identification requires abstract reasoning. Many researchers view RPM as the best and purest predictor of Spearman's *g* (Deary & Smith, 2004; Eysenck, 1998; Holyoak, 2012; Jensen, 1973, 1980, 1987, 1998; Llabre, 1984; Matrinez, 2013; Neisser, 1998; Thorndike, 1986; Vernon, 1947). Indeed, due to the solving of the RPM requiring the eduction of relations and correlates, even Spearman (1946) himself considered them the nearest representation of *g*. However, other researchers argue that, while performance on the RPM is a good predictor of *g*, it is not a pure measure of *g* and that it is wrong to assume so (e.g., Burke, 1958; Gignac, 2015). In fact, Gignac (2015) argued that it is erroneous to equate failure or success in finding correlations between performances on other tests and RPM performance with failure or success in finding correlations between that former test performance and *g* (as done in e.g., Basso et al., 1973; Corben et al., 2006; Day et al., 2005; Schellenberg & Moreno, 2010; Walker, et al., 2013). Moreover, Gignac (2015) states Irvine's argument that due to the lack of a high test-specific variance (Wainer, 2002), it is impossible for RPM to be a pure measure of *g*. Indeed, in Gignac's (2015) own study of four bifactor models, he concluded that not only is the RPM test not a pure measure of *g*, but this test even fails to demonstrate a particularly high *g* loading. Nevertheless, RPM's validity as a proxy measure of *g* and as a useful test to assess abstract reasoning is undisputed (Gignac, 2015).

1.2.5 Wechsler Scales of Intelligence

Among the most widely administered intelligence tests are David Wechsler's intelligence scales. Lezak (1995) stated that a substantial portion of neuropsychological examinations consists of the administration of a Wechsler intelligence scale. Indeed, Wechsler's scales are still the dominant measure for individual intelligence testing (Camara et al., 2000). This is due to a number of factors, which, while not initially his own ideas, were first formally combined by Wechsler into single scales.

The first Wechsler scale was the Wechsler-Bellevue Intelligence Scale published in 1939. Boake (2002) provides a comprehensive and extensive overview of the history of the Wechsler scales and the source of its items. Of fundamental importance is Wechsler's switch from ratio IQ to a deviation score as already in use with the Stanford-Binet scale and his combining of both verbal and non-verbal IQ into a single scale. Despite the name being associated with one of the most popular measures of intelligence to this day, Wechsler did not devise his test from scratch (Boake, 2002). Instead, he borrowed, adjusted, and combined many previously used measures of different cognitive abilities into a singular scale. The practice of combining verbal and non-verbal IQ indices was not Wechsler's idea and, indeed, quite common at the time of the development of his first scale, with many psychologists supplementing the verbal test with performance scales to account for the shortcomings of a purely verbal measure outlined above (Boake, 2002). Still, Wechsler was the first to formally combine such tests into a singular measure and to create scoring instructions for such a measure. Several factor analyses have confirmed the distinction between verbal and non-verbal IQ (e.g., Balinsky, 1941). However, the

Digit Span and Digit Symbol subtests fail to meet the criteria for assignment to either factor (Boake, 2002). Moreover, Wechsler adapted the intelligence test for various ages, ranging from childhood to adulthood.

Wechsler's combination of deviation scores, a single scale assessing verbal and non-verbal IQ, as well as his adult standardisation procedures quickly brought the highest status to his scales, with Boake (2002, p. 401) terming it "almost monopoly status". As such, the Wechsler scales are considered as "gold-standard" intelligence tests. However, it is important to note that they reached this status not by any objective distinction between gold standard and non-gold standard tests but simply by their widespread use and acceptance within the academic community and the general population.

One such scale, the Wechsler Adult Intelligence Scale (WAIS) consists of multiple subindexes and consequently provides multiple scores. At the bottom of this are the verbal comprehension index (VCI) and the Working Memory Index (WMI), which together provide a Verbal IQ (VIQ) score. On the other side, there is the Perceptual Organisation Index (POI) and the Processing Speed Index (PSI), which, combined provide a Performance IQ score. Combining the VIQ and Performance IQ scores allows for the computation of a Full-Scale IQ (FSIQ) score. As a consequence, tests like the Stanford-Binet or Wechsler Scales require test takers to perform a number of tasks, each of which is supposed to tap into a different cognitive skill, thus covering a more extensive array of cognitive skills altogether (Kovacs & Conway, 2019), while still allowing for intricate detail in research and diagnosis of cognitive performances related to more specific abilities.

1.2.6 National Adult Reading Test

While historically the National Adult Reading Test (NART) has not served a specific purpose to the further development of widely used full-scale IQ tests, the NART will briefly be outlined as it will be mentioned as a reference measure for the literature outlined in the following sections. Therefore, an understanding of the NART and its limitations is warranted. Originally developed by Nelson (1982), the NART aims to assess test-taker's intelligence based on the number of phonologically irregular words individuals pronounce correctly (van der Linde & Bright, 2024). According to van der Linde and Bright (2024), such tests are used due to the preservation of single word pronunciation across a wide range of conditions (Crawford, 1992; McGurn et al., 2004; O'Carroll, 1995; Sharpe & O'Carroll, 1991). Moreover, according to Nelson (1982), the correlation between word reading and intelligence is barely moderated by age and social class. While the NART itself only requires the oral pronunciation of 50 words, the score is believed to translate into an accurate estimation of full-scale IQ as measured by the WAIS-IV (Wechsler, 2008). However, despite the ease and speed of the measure's administration and score computation, the NART's standing as a sufficiently accurate measure of intelligence is put into question. Van der Linde and Bright (2024) for instance, found that the measure's (and its various international iterations) fell victim to serious floor and ceiling effects. Their study shows that a significant portion of the population falls outside of the ranges predictable by the NART. More specifically, this ranged from 3.90% in Nelson and Willison's (1991) UK version to 27.55% in Vaskinn and Sundet's (2001) Norwegian version. The 13 versions assessed in the van der Linde and Bright 2024) study averaged a percentage of 13.14 of people ranging outside the measure's predictable scores. More explicitly, having such big ranges fall beyond the measure's prediction limits means an overestimation of below average premorbid functioning and an underestimation of above average premorbid functioning (Thomas et al., 2021).

Such limited prediction of IQ scores can have substantial impacts, not only on research findings, where it can result in the restriction of correlational analyses, but more seriously, in clinical settings, where the true impact of certain conditions might not be estimated accurately, due to individuals still scoring above or below the NART's limits, despite significant drops in their IQs. While the NART is widely recognised as an adequate IQ proxy measure (van der Linde & Bright, 2024), its narrow prediction limits call for caution when it comes to the interpretation of IQ-related conclusions based on the NART alone.

1.2.7 IQ Scoring

To understand some of the criticisms outlined in the following section, it is critical to first understand how IQ test scores are usually computed. Overall test scores for measures like the Wechsler scales are conventionally converted into a scale with a mean of 100 and a standard deviation of 15. Such a score reflects an individual's *relative* performance compared to others in their age group, meaning a score of 100 represents perfectly average performance for the test-taker's age group. While such a score no longer represents an actual quotient, scores on most intelligence tests (including the Wechsler scales) are still referred to as IQ scores (Neisser et al., 1996). The repercussions of this standardised scoring system will be outlined further in section 1.4.1. In the same way that overall quotient scores are normalised across the population and individuals are compared only to their peers, performance on different subscales in a particular test is not always equal in absolute terms but, generally speaking, positively correlated (Rammsayer & Troche, 2016). Put simply, an individual's performance on each subtest of a given cognitive function might not perfectly co-vary with their absolute performance on another subtest, but the difference in performance on both subtests will not vary much in its relative standing to that of other test-takers (e.g., one might perform better on absolute verbal comprehension than working memory, but both of those scores would still be correlated in their relative standing to other test takers).

An assessment of intelligence automatically poses the question of the origins of this capacity. The following section will discuss the nature versus nurture debate and its implications for theory and practice.

1.3 Is Intelligence an Inherent Trait or an Acquired Skill?

Various factors have been found to predict performance on traditional intelligence tests. As with probably all psychological phenomena, the major question is whether nature or nurture plays the sole, or more significant role. The following will provide an overview of this debate with a view to contextualising the environmentalist perspective adopted in the current thesis.

1.3.1 Nature

While the heritability of intelligence is commonly taken as fact by the general population, it is important to investigate the arguments that support this claim. One such argument is that maternal IQ and maternal education have been found to predict a substantial amount of the variance in children's IQ scores (Ceci, 1996). While factors such as maternal IQ and maternal education might suggest the heritability of IQ, this conclusion is premature, as it falls victim to the fallacy of erroneously equating correlation with causation. There is a strong argument to be made that this correlation does not justify conclusions regarding causation, likely related to genetic factors. Specifically, it may in fact be indicative of a rather indirect relationship based on environmental factors related to the higher IQ of the parent. Indeed, Gottfried (1984) and Hart and Risley (1992) found that children's IQ scores correlate with parental use of language and resources at home, which undoubtedly relates to parental IQ but does not in itself point to genetic inheritance. This is further supported by the significant correlation found between parental socioeconomic status (SES) and IQ (r = .33; White, 1982; further supported by Sheshagiri et al., 2016), which will be investigated in more detail below.

Of arguably higher importance to this debate are twin studies. Most twin studies examining the heritability of intelligence cite correlation coefficients between .3 and .8 (e.g., Haworth et al., 2010; Jensen, 1969; Plomin, 1990). These studies, based on the analysis of intelligence quotients for twins raised together and apart (e.g., Petrill et al., 1998; Pike et al., 1996; Plomin, 1997; Plomin & DeFries, 1998), are highly persuasive of the case that there is little room for the influence of environment in the determination of intelligence scores. This is especially true for those studies citing heritability factors of around .8. However, aside from large variance in such coefficients found across studies, the interpretation of these findings has been widely criticised in the literature. These criticisms will be outlined below.

One major criticism of the nurture argument is provided by Harvard Physicist Layzer in his 1972 article “Science or superstition? (A physical scientist looks at the IQ controversy)”. Layzer criticises the methods employed to arrive at conclusions about heritability estimates in intelligence research and deems these “not merely unreliable but meaningless” (p. 266, 1972). More specifically, Layzer argued that it does not adhere to scientific practices to speak of intelligence as if it was a measurable entity when this is not supported by the data. Therefore, any conclusions about heritability are meaningless because it is impossible to determine the heritability of something that cannot be measured in the first place. Nevertheless, Herrnstein treats IQ heritability as if it is a real, observable phenomenon. Indeed, he argued that since heritability coefficients predicted by general genetics (e.g., parental IQ should bear a .5

correlation with their children's IQ, as that is the general covariance that parental genes have with their children's genes) are close to the "actual", observed values, this supports his argument of the heritability of intelligence. However, Layzer warns that even Herrnstein's "actual" values are not the real values but instead, those that indicated the "obtained median correlation". This further makes Herrnstein's conclusions unreliable.

Layzer continues to argue that, even if intelligence met the relevant requirements to be an objectively measurable property, then inferences made about its heritability would still be flawed insofar, as these inferences assume statistical and practical independence between genetic and non-genetic factors with additive contributions. However, this assumption is almost never borne out. Furthermore, argues Layzer, the heritability estimate would still not provide any insight into the impact of genes under different environmental circumstances. The reason for that is, that some people's genetic predispositions still make them respond differently to the same environment. The conclusion that can be drawn from this, he argues, is that the heritability of intelligence does not allow us to make any inferences about how well children might react to environmental changes. The assumption made by both Jensen and Herrnstein, argues Layzer, that low IQ children are practically incapable of acquiring higher cognitive skills, is basically to assume that these skills either cannot be acquired at all or that all children have received the same education in those skills. In line with this argument is the fact that so-called "mock-genetic" environmental factors appear to be of significant importance as well. Factors like cultural values, wealth, social status, and occupational and educational level are all interrelated and significantly correlated with IQ. However, these factors are primarily predicted by parental SES and environmental backgrounds. Hence, argues Layzer, similarities in environment are passed down at least to the same extent as similarities in genes are.

Moreover, as Schlinger (2003) points out, virtually all twin studies rely on the underlying assumption of an equal environment, which, as has been demonstrated by Beckwith (1999) and Joseph (1998) does not hold true in reality. The environments of both identical and fraternal twins are often not equal, or even similar. In the same way that equal environments cannot be assumed for relatives, they also cannot be assumed for people of different races, argues Layzer. Not only is there no genetic difference between the races, but even if there was, it would be meaningless to make any conclusions based on those genetic differences. This is because the environments people are exposed to are so substantially different, it would be premature to attempt to extract a quantifier to assess the impact of heritability on any intellectual differences across individuals. On the other hand, environmental factors can also not be assumed to be independent even in monozygotic twins reared apart. These individuals still share significant factors which substantially impact the role of the environment. Gender, for instance can significantly impact an individual's perception of their own skills, and their access to specific educational opportunities (e.g., maths), thus potentially limiting a twin pair's exposure to such opportunities and consequently their skill development.

The view that intelligence is an inherent trait automatically limits the possible impact of environmental interventions to improve it, with many researchers and theorists even arguing this to be impossible. Jensen (1969), for instance was convinced that no environmental efforts can meaningfully impact cognitive abilities. More specifically, he believed that low IQ children lack the capacity to acquire the cognitive skills that are involved in abstract reasoning and problem-solving. Accordingly, Jensen argues, those children should not receive encouragement to get into occupations demanding higher cognitive skills. Eysenck (1971) and Herrnstein (1971) agree with Jensen's viewpoint about the unmalleable nature of cognitive abilities. Indeed, those who closer concern themselves with Jensen's, Eysenck's, and Herrnstein's conclusions about intelligence's fixed nature on societal structures will quickly recognise racist, classist, and generally discriminative ideologies that appear to aim at limiting the possibilities of those already in minority groups.

Indeed, the failure to properly define a referent for the term intelligence suggests that, perhaps, it is wiser to take a functional approach, viewing intelligent behaviour in context, rather than hypothesising there to be a mental entity that cannot be observed (Schlinger, 2003). Understanding intelligent behaviour as a skill allows researchers to put this skill into context, consequently allowing them to analyse the variables of which this skill is a function. Such an analysis then allows the researchers to alter those variables so that the desired skill can be honed in the ideal environment. Such an approach will be outlined in detail in section 1.6. First, it is helpful for the reader to be familiar with the arguments supporting intelligence to be impacted by an individual's environment, an outline of which will now be provided.

1.3.2 Nurture

Given that the number of potential environmental influences on intelligence is very large compared to biological influences, many of these have been studied within the literature. Deprivation conditions such as impoverishment, racial discrimination, and erratic school attendance, for instance, have been found to negatively correlate with IQ scores, a correlation that becomes increasingly negative with age (Ceci, 1992; Douglas, 1964; Lee, 1951; Neisser et al., 1996). Complementary to this finding, numerous studies have found a positive correlation between early breastfeeding and later performance on tests of cognition (e.g., Angelsen et al., 2001; Mortensen et al., 2002). While Isaac et al. (2010) found a dose-response relationship between breast milk intake and later IQ, which was most significant in males, the impact of breastfeeding on cognitive test performance might also depend on whether the child was born prematurely or not (Anderson et al., 1999). However, the role of breastfeeding can be questioned due to their potential susceptibility to confounding factors, such as the choice to breastfeed being associated with higher socioeconomic status, level of education, and child-rearing attitudes, all of which have been found to impact cognitive development themselves (Isaacs et al., 2010). Similarly, prolonged malnutrition has been found to have a negative impact on IQ scores. For instance, Sheshagiri et al. (2016) found that malnourished and obese children have lower IQ scores than healthily nourished children. Furthermore, B12, folate, iodine, iron, protein, and zinc deficiencies have also been found to be associated with lower IQ scores (Pollitt et al., 1993; Qian et al., 2005; Saloojee & Pettifor, 2001; Schoenthaler et al., 2000; Stein et al., 2005). Additionally, the usage of addictive substances such as alcohol, smoking, and drugs (more specifically cannabis and cocaine) can result in significant cognitive impairment, especially if intake occurs during pregnancy or childhood, as the blood-brain barrier is not yet as effective as in adulthood (Oommen, 2014).

Beyond physiological and nutritional factors, certain activities or environmental conditions can further impact IQ test performance. Sheshagiri et al. (2016) found significant differences in IQ scores based on the purpose of children's media usage, with those using it as a source of knowledge showing higher mean IQ scores than those using it for entertainment purposes. Furthermore, they found higher IQ scores in children who attended extra coaching classes, and Schellenberg (2004) proposed that increased musical skills may transfer to other domains like language and mathematics. Additionally, it was found that children with proactive hobbies, such as drawing, dancing, music, and playing sports, had higher mean IQ scores than those who watched TV in their free time (Sheshagiri et al., 2016). Apart from these factors, over which a child has deliberate control (if it is within their means), studies have repeatedly found bilingualism to positively impact cognitive functions (e.g., Costa & Sebastián-Gallés, 2014). While bilingualism might (initially) result in adverse outcomes, such as lower reaction time and accuracy in lexical tasks (Gollan et al., 2007), speaking more than one language has also been associated with various positive outcomes, such as increased attentional control and inhibition (Bialystok & Senman, 2004) and switching (Costa et al., 2009) in later years. Peal and Lambert (1962) further found higher verbal and nonverbal IQ scores in bilingual children compared to the monolingual control group.

Another factor that has been found to be of significant importance is parental socio-economic status (SES). For instance, SES can substantially impact an individual's access to educational resources. Such access has repeatedly been found to affect cognitive performance in a variety of aspects. Years spent in school and level of education have also been found to significantly predict IQ scores. According to Neisser et al. (1996), the correlation between IQ scores and total years of education is about .55. While this alone does not necessarily suggest causation (and especially not so in any specific direction), other study findings suggest that there might be a causal link. Schellenberg (2004), for example, found that dropping out of school decreases IQ scores by about two points for each year that is not completed. Furthermore, according to Neisser et al. (1996), school attendance and IQ scores show a bidirectional relationship, in which IQ determines how long students stay in school, but school attendance in turn improves mental abilities, such as those measured by IQ tests. Further complementing this finding is the finding that, on average, regularly attending students had higher IQ scores than those who attended school intermittently (Neisser et al., 1996). Similarly, IQ scores were found to drop over the summer vacation, especially in children who spent their vacation engaging in activities that do not resemble the school curriculum. This effect was found to be enhanced among children with a lower SES background (Neisser et al., 1996). All of these findings suggests that environmental influences do have a strong impact on IQ and that, perhaps, the impact of the environment has been underestimated. Nevertheless, strong correlational evidence in the domain of environmental influences does not itself rule out the possibility of the role of genetic inheritance.

Despite evidence for the influence of both genetic and social factors in the determination of IQ scores, it is important to remember that, of course, the nature-nurture divide is a false dichotomy and trying to resolve the question here can lead to unproductive questions. Indeed, while it might appear that some intelligence theorists subscribe to only one view, that is rarely the case. Many theorists agree that intelligence is most likely determined by a complex interplay between nature and nurture. Indeed, recent studies have found support for this viewpoint. One such example is the finding that the heritability of IQ does not appear stable throughout the lifespan. Haworth et al. (2010), for instance, found heritability to increase as the individual gets older. It was proposed that this finding is due to a gradual increase in genotype-environment covariance (Haworth et al., 2010; Johnson et al., 2011). This, in turn, is hypothesised to be due to more intelligent people increasingly putting in more effort to surround themselves with cognitively stimulating environments (Dickens & Flynn, 2001; Haworth et al., 2010; Johnson et al., 2011; Scarr & McCartney, 1983). Indeed, Johnson et al. (2011) concluded that a genotype-environment covariance likely accounts for as much as 30% of the variance in adult IQ.

Brant et al.'s (2013) findings further demonstrate the complex interplay between genetic and environmental influences. Investigating a cross-sectional sample of nearly 11,000 adolescent twin pairs, Brant and colleagues found that the impact between environment and heritability differs across IQ groups. More specifically, they found that the environment played a bigger role in high-IQ individuals, whereas low-IQ individuals appeared to show higher heritability coefficients. This, suggest the authors, supports their proposal of an extended period for intellectual development in individuals with higher IQ. This period is defined by prolonged cortical thickening, which was found to be positively correlated with cognitive performance (Brans et al., 2010; Shaw et al., 2006, 2008). Such a period makes the brain particularly sensitive to environmental influences (Shaw et al., 2006). Brant et al. (2013) propose that this can be traced back to the fact that higher IQ might be related to heightened levels of attention and arousal, which in turn may facilitate plasticity (Knudsen, 2004). Alternatively, higher IQ individuals might just simply be more likely to seek cognitively stimulating environments. As the individual's access to means increases with age, people are more likely to find themselves in an environment that more closely resembles their genetic predisposition regarding cognitive demand and abilities (Plomin et al., 1977; Haworth et al., 2010). To complement these findings, various studies have found a positive correlation between the impact of heritability and IQ, and a simultaneous negative correlation of the impact of environmental influence (Haworth et al., 2010; Bartels et al., 2002; Brant et al., 2009; Lenroot et al., 2009). Similar to Brant et al.'s (2013) finding that gene-environmental interplay varies across IQ groups, Scarr-Salapatek (1971b) found that differences between zygotic and dizygotic twins were larger in advantaged groups. This suggests that, given genetic equivalence, the environment plays a bigger role when the individual has more opportunities rather than when the individual experiences deprivation conditions such as those outlined above.

In conclusion, the false dichotomy between biological and environmental influences in most cases, is created not by asking how environment and genes interact to produce intelligence quotients but rather in asking which of the two has the greater influence. Indeed, Layzer argues that it does not matter which of the two is a better predictor. All that matters is that, if the environment has any meaningful impact at all, efforts should be made to harness the beneficial effects of this environment in interventions by producing what Layzer calls an idealised environment. An idealised environment would provide the best conditions for individuals to thrive, despite genetically predetermined differences in cognitive capacity. Indeed, the idea that intelligence can be enhanced by exposure to idealised environments is more consistent with an environmentally focused behaviour analytic approach to learning and intellectual development. The implications of this possibility that idealised environments can make a real difference to individual intellectual capacity will be discussed in section 1.6.

The issue of the malleability of intelligence and the role of the environment, as well as the view of intelligence as an emerging skill set, might have implications for how intelligence is assessed. For the most part, intelligence tests have emerged from a fairly nurture-focused tradition in which general intelligence was believed to be approximated by IQ scores. In other words, according to classical test theory, the test is merely a proxy for the real entity that lies underneath. This seems to be the dominant psychometric paradigm of intelligence testing. However, if intelligence was to be perceived differently as a malleable skill set that was improvable, high intelligence scores would, therefore, be attainable through appropriate interventions. The current thesis is interested in what such tests would look like that would emerge from such a view. While this will become apparent later, it might be worthwhile to first review some of the controversies and limitations of traditional intelligence testing in more detail to set the stage for the introduction of a more novel approach based on a functional analysis of intelligent behaviour.

1.4 Controversies and Issues Surrounding Intelligence Testing and Theory

1.4.1 The Computation and Meaning of Test Scores

As has been explained in section 1.2, the stability of IQ scores does not refer to constancy in absolute scores across time but to relative scores. In other words, while intelligence is assumed by many researchers to be stable, this is often on the basis of the stability of IQ test scores themselves, which are not, in fact, stable across time. Instead, one's relative IQ score is fairly stable across time insofar as the performance of an individual on an intelligence test tends to remain relatively stable compared to their peers over time, even while intellectual skills are improving. Therefore, the emphasis of stability relates to IQ, not intelligence itself. The intelligence researcher James Flynn observed that population-level IQ has risen substantially since the mid-20th century (Flynn, 1984, 1998, 2007), during which time the average IQ of the population remained steady at a perfect 100 by definition of the scoring method. More specifically, Flynn observed a rise in the lower end of the distribution, which in turn raises the average. These rises are substantial enough to warrant regular adjustments of the scoring system in order to keep the average IQ at a score of 100 and the normal distribution at 15 points. While for the majority of Flynn's observations, the mean IQ has increased, recent evidence suggests this upward trajectory in intellectual ability to have stabilised and even reversed in the last generation (Dutton & Lynn, 2014, 2015; Dutton et al., 2016; Pietschnig & Voracek, 2015; Shayer & Ginsburg, 2009; Shayer et al., 2007; Sundet et al., 2004; Teasdale & Owen, 2008;Woodley & Meisenberg, 2013). Such drastic differences in average IQ scores cannot be (solely) attributed to genetic factors, as suggested by Jensen (1998) and Mingroni (2004, 2007). Hence, various alternative explanations have been put forward, such as environmental factors (Ceci, 1992; Dickens & Flynn, 2001; Flynn, 2007; Lynn, 1990, 2009) or social factors (Blair et al., 2005;

Brand, 1987; Ceci, 1991). While the reason for such fluctuations has not been discovered yet, the Flynn effect not only raises interesting questions regarding the basis of the widely observed stability in IQ scores across the life of an individual and across time for the population as a whole. These questions bear directly on the matter of the influence of environmental factors on general intellectual ability however it is measured.

Another widely assumed property of intelligence test scores is their naturally occurring normal distribution. Indeed, intelligence is probably the first example every Psychology student hears when they learn about the normal distribution. However, the normal distribution of IQ test scores has been called into question. Indeed, it is important to know that IQ test scores are artificially adjusted to fit a normal distribution. McLoughlin (2022) explains how IQ tests are calculated. After administering a set of questions to an age-matched group of people, each person gets assigned a total score, which is then ranked by converting the total score into a z-score. Multiplying the z-score by the pre-set standard deviation chosen to fit the desired spread of the data (usually 15 IQ points), and finally adding 100 results in the final IQ score.

Despite manual intervention into the distribution of intelligence, multiple analyses show that the normal curve still fails to accurately predict the actual distribution of scores. More specifically, Burt (1963) concludes that the normal curve fails to account for the unexpectedly high number of above-average IQ individuals. Even when excluding those cases that would possibly impair the requirements of a normal distribution and excluding cases where environmental factors play a part, the deviation was statistically significant. Indeed, the actual proportion of people with IQ scores above 160 was more than 12 times higher than expected. Moreover, Burt estimated that the proportion of individuals with an IQ score over 175 more realistically exceeds 70 per million, rather than the predicted 3 or 4 per million.

But why does it matter, whether IQ scores follow a normal distribution or not? According to Burt (1963), the implications of such distributions are thought to provide insight into the nature and origin of intelligence. However, he also states that this view has been contested for the mere fact of the distribution's artificial creation, the implications of which are non-existent because this distribution is not occurring in the natural world. This ties in with the general meaning of IQ scores, which will be discussed below.

As aforementioned, IQ test scores do not provide an absolute score but rather a relative indication of where the individual falls compared to their peers. While this might seem trivial to some, the implications of this practice of only having a relative, rather than an absolute index, bear far more weight for theory and meaning than it first appears. Layzer (1972) argues that this alone precludes the assertion that IQ tests measure intelligence. As the test is not an objective measure, but just a rank order, it is not a direct measure of the thing that the tests claim to measure. This argument would remain, even if raw scores were used, as raw scores still do not provide an index of an individual's "possession" of the trait but remain to only hold meaning relative to other scores. In addition to the problem of relative scores being too imprecise, relative scores have been argued to be imprecise as well. Gignac (2015) criticises the incongruency between theory and measurement. He argues that since the positive manifold indicates the correlation between a number of different cognitive tasks, the score, too, should reflect this variety in tasks. His solution is to provide multiple scores, as a single score is meaningless.

In conclusion, the fact that IQ is relational highlights the theoretical issues of defining intelligence, as a proper definition would allow for absolute indices. However, as so far, we only have an approximation of what intelligence is, so too, the tests only provide an approximation of people's "possession" of that trait. In other words, the approximation of scores perfectly resembles the approximation of the concept, instigated by a lack of a proper definition and, indeed, a flawed method to establish a theory.

Finally, as the meaning of scores has been discussed, the question that is left is, what “capacity” the scores actually allude to. This answer seems simple at first, they measure intelligence. However, given that intelligence as a construct is very flawed in itself, as discussed in section 1.1.2, this does not provide any meaningful insight. Indeed, given that the perhaps most widely cited definition of intelligence is that intelligence is “what IQ tests measure”, Eysenck (1973), argues it redundant to ask the question whether IQ measures intelligence. It is a circular argument in which IQ indicates an individual’s intelligence test score, and intelligence test scores only reflect IQ indices. In other words, IQ test scores measure IQ test scores. Eysenck compares this to asking whether thermometers measure temperature. This question is just as meaningless as the very definition of temperature is entirely reliant on the measurement by a thermometer.

While many arguments for the lack of meaning behind IQ test scores have been discussed, these criticisms do not render the idea behind intelligence or IQ tests entirely meritless. Kovacs and Conway (2019) argue that it is more accurate to “interpret IQ as an index of specific cognitive abilities rather than the reflection of an underlying general cognitive ability” (p. 189). Indeed, there is no doubt about the construct’s utility as an index. As demonstrated above, IQ is a good predictor of various desirable life outcomes. However, this does not constitute a sufficient justification for the claim of an underlying trait. Even if we were to disregard the aforementioned issues and instead assumed that IQ test scores provide an index of what *g* is supposed to represent, as the following section will show, investigating the meaning and origin of *g* quickly exposes further issues with traditional intelligence theory.

1.4.2 Erroneous Conclusions Based on the Positive Manifold

The positive manifold is used as proof of a singular underlying trait explaining cognitive performance. However, nowadays, the positive manifold is upheld intentionally through the targeted inclusion or exclusion of new test items, and many of the theoretical conclusions based on the positive manifold turn out to be erroneous or at least lack a sufficient foundation, therefore falling victim to the fallacy of defective induction. To understand why these conclusions need to be scrutinised, the reader will be reminded of the origins of the positive manifold and what it is proposed to represent before the methodology to arrive at *g* theory and resulting conclusions will be discussed.

Spearman's theory of a single underlying factor, *g*, has been met with criticism. The primary focus of this criticism questions what *g* is actually supposed to represent. Some theorists argue that *g* appears to be a more accurate representation of abstract reasoning ability (Gustafsson, 1984) or, more prominently, processing speed (e.g., Grudnik & Kranzler, 2001; Jensen, 2006; Myerson et al., 2003) or working memory (Conway & Kovacks, 2013; Unsworth & Engle, 2005). However, these various accounts have also failed to gain traction. Furthermore, Spearman's single-factor theory is heavily criticised for its methodological issues. One such issue, and arguably the most important one, relates to the use of factor analysis as a sufficient tool to identify and define an underlying factor. While factor analysis and the positive manifold allow an insight into the correlation between performance on different tasks, which hints at the existence of *something* underlying this correlation, it does not allow any insight, and more importantly any *sufficient* insight, into the nature of this underlying capacity (Gottfredson, 1998). Hence, Spearman's theory falls victim to the error of reification, the erroneous utilisation of a label as proof of a construct's existence, when it, in fact, merely provides a description of a pattern of effect, not an explanation for that effect (Bell et al., 2001). This failure of mere correlation to prove the existence of an extent capacity will be further elaborated upon below.

Schlinger (2003) argues that the whole approach of establishing a definition before knowing what is being defined is a flawed method to investigate concepts such as intelligence. Instead, he argues, it is more useful to decide on the behaviour in question, observe it, and then provide a “discovered” (p. 23) definition of such behaviour. The discovery of definitions is a result of experimental analysis of behaviour with the goal of finding some sort of order. Indeed, this latter approach appears to be more closely aligned with the way hard sciences usually approach their research. Therefore, since Psychology fights for its right to be considered one of the sciences, the choice to approach one of the most widely studied phenomena in such an unscientific method is highly questionable. Since that is the conventional approach to intelligence, however, the following section will analyse further issues with the utilisation of *g* as an explanation rather than merely a label for an observed effect.

While many followers of the *g* hypothesis accept *g* as the explanation for the positive manifold, one should practice caution when doing so. Indeed, Covacs and Conway (2019) argue that *g* is nothing more than a statistical abstraction and warn from interpreting *g* as the cause of the positive manifold. While factor analyses confirm the finding that specific clusters of tests show stronger intercorrelations than other clusters (e.g., vocabulary and reading comprehension compared to either of those with mental rotation), these clusters of higher correlation are referred to as *group factors*. *G* theory proposes that group factors are explained by the general factor *g*. So far, this seems like a logical conclusion. However, the interpretation of such a finding varies and Kovacs and Conway (2019) urge for caution, as traditional *g* theory argues for the finding’s justification of an underlying trait. Kovacs and Conway (2019), however, explain that the data only justify the argument for *g* as a statistical construct, where *g* simply labels the common variance found among performance on different tests, rather than a psychological construct, where *g* represents underlying processes or mechanisms causing such correlations. In other words, *g* is the consequence (more specifically, the label), not the cause of those correlations; as Schlinger put it, it is “not a thing being measured” (p. 17, 2003). In line with this argument, they redefine intelligence as an emergent formative construct instead of a reflective latent trait. Formative models argue that the composite variable does not exist outside of measurement (Kovacs & Conway, 2019). In effect, *g* is not inherently “out there” but solely reflects a statistical observation.

Statistically, argue Kovacs and Conway (2019), it is accurate to argue for a latent variable (i.e., *g*) as an explanation for the positive manifold; psychologically, however, doing so would require *g* to represent actual processes and mechanisms. Any argument that lacks this process or mechanism is insufficient in providing an accurate psychological explanation and consequently remains purely statistical. As alluded to above, this makes *g* theory fall victim to the error of reification. Since *g* only functions as a label for the observed variance, it does not offer any further information. Given that *g* is a necessary mathematical consequence of the positive correlations (Krijnen, 2004), Kovacs and Conway (2019) argue that interpreting *g* theory further falls victim to circular reasoning, whereby *g* is used to label the observed positive correlations (i.e., it is described as a consequence) and further used to explain those explanations (i.e., it is described as a cause). Therefore, conclude Kovacs and Conway (2019), it is not erroneous to interpret *g* as a statistical consequence, but only to view it as a psychological cause.

In line with the issue of circular reasoning and *g* as a solely statistical construct is the issue that, whilst *g* is supposed to explain the positive manifold, test development itself is not only founded on but, more importantly, guided by the idea of the positive manifold. This means that the inclusion or exclusion of test items proposed to become part of intelligence tests relies on whether or not they support the positive manifold. If an item shows no or even a negative correlation, it is not included in the next edition of the measure. That is, the positive manifold is used to prove the existence of a singular underlying trait, while simultaneously, anything that would disprove this claim is disregarded as not being linked (or at least not directly) to the concept of intelligence. In effect, the only conclusion allowing confidence is that IQ tests measure some correlation between cognitive abilities. However, the causes or influences of this correlation are yet to be known. Indeed, this error of mistaking the evidence as proof of a psychological, rather than just a statistical construct, leaves room for various alternative interpretations of *g* (Schlinger, 2003), which will be discussed below.

One alternative explanation for the positive manifold is offered by *Process Overlap Theory* (Kovacs & Conway, 2016). The basic argument of Process Overlap Theory is that intelligence is constituted of multiple components which either fall into domain-general (i.e., executive functions) or domain-specific processes. Executive functions, therefore, explain a person's general ability (e.g., tasks related to working memory processes), whereas domain-specific abilities are those related to, for example, verbal or spatial tasks. Kovacs and Conway (2019) argue that Process Overlap Theory further offers an explanation for the frequent finding that *g* appears to be stronger at lower levels of abilities than at higher ones (e.g., Blum & Holling, 2017; Molenaar et al., 2017), a phenomenon known as *ability differentiation*.

Ability differentiation describes the phenomenon whereby cognitive ability gets more specific as IQ increases. Thus, it almost functions like a reverse of the *Anna Karenina principle*, as a deficiency is more generalisable than success. Kovacs and Conway (2019) argue that

Process Overlap Theory's proposal of domain-general and domain-specific processes explains the ability differentiation phenomenon by arguing that a person's basic executive functions must be what fosters domain-specific processes. Put simply, an individual's executive processes determine to which degree their domain-specific processes can develop. This development of domain-specific abilities then gets more refined as abilities increase. Kovacs and Conway (2019) stress that high efficiency in domain-general processes does not guarantee good performance on cognitive tests, but that instead, inefficiency increases the risk of failure. While this might seem somewhat similar to Spearman's theory of general and specific factors, it is important to note that Kovacs and Conway (2019) differentiate the theories by explaining that their theory argues for the domain-general processes to be constituted of multiple executive processes, whereas Spearman's general factor *g* is supposed to be a single, underlying capacity. Moreover, Process Overlap Theory offers an explanation for the *worst-performance* rule, a phenomenon that posits the worst performance on a cognitive task to be a more accurate predictor of *g* than the best performance (Larson & Alderton, 1990; Rammsayer & Troche, 2016). As the worst performance is believed to indicate failures in executive processes (Larson & Alderton, 1990), the theory's focus on executive functions offers a logical explanation for the worst-performance (Kovac & Conway, 2016).

Another alternative explanation for the positive manifold comes in the form of a behaviour analytic theory, namely, RFT (Hayes et al., 2001). RFT completely opposes and contests the assertion that intelligence is an inherent trait and rather argues that a particular set of skills, namely relational responding, underlie all of higher cognition. In doing so, the theory offers an explanation for the positive manifold, as the argument that relational responding defines intelligence implies that the observation of a positive correlation between the performance on various IQ test items attests the proposition that a wide range of tests draw upon the same set of skills. To illustrate this argument, the following section will give an outline of RFT. This particular environmental account is central to the current thesis and lays the foundation for an entire new paradigm within which to understand human intelligence and its measurement. This will be followed by an outline of how this conceptual framework allows an understanding of a wide range of measures of cognitive ability as effective proxies for the assessment of the fluency of derived relational responding repertoires and further explains how specific IQ test items relate to relational responding.

1.5 A Relational Frame Theory Perspective on Intelligence

1.5.1 Intelligence in Behaviour Analysis

With its heavy emphasis on studying *observable* behaviours and reinforcement histories, behaviour analysis has, historically, intentionally neglected the study of individual differences in personality traits and cognitive abilities (e.g., Williams et al., 2008). Given its focus on individual behaviour rather than studying group effects and averages, Williams et al. (2008) argued that such neglect is somewhat surprising. Chomsky (1959) even claimed that it was impossible for operant psychology to explain the complexity and generativity of language and cognition altogether. In cases where such constructs as intelligence were discussed by the behaviour analyst, it was argued that skills in traditional intelligence tests should be viewed as separate behaviours with differing controlling variables (e.g., Staats & Burn, 1981). However, this approach fails to account for the high correlation in performance on all the different measures of intelligence and their subtests (Williams et al., 2008). Despite the behaviour analyst's traditional reservation in studying cognitive processes directly, due to the difficulty in doing so objectively and reliably, recent decades have seen more attempts to analyse cognition through a behavioural lens. The definition of intelligence continued to be approached in terms of intelligent *behaviour* (Goddard, 2018; Schlinger, 2003), thus retaining the focus on observable behaviour. In effect, the fundamental behavioural perspective that hypothesising about intelligence as a mentalistic construct is not beneficial has been retained. However, there has been a renewed emphasis on the importance of analysing behaviours that are often considered to be evidence of intelligence within mainstream psychology. In particular, one behavioural theory has proposed a new approach to studying and understanding cognitive processes in a functional and behaviour-analytically coherent manner. Specifically, RFT (Hayes et al., 2001) is a behaviour analytic theory of human language and cognition that argues that derived relational responding is a learned, generalised operant underlying most, if not all, higher cognition. The following sections will provide a summary of RFT before introducing the RFT perspective on intelligence and complementary research findings.

1.5.2 Relational Frame Theory

RFT is rooted in the principles of functional contextualism (Hayes et al., 2001). Functional contextualism is a philosophy of behavioural science that treats all instances of behaviour as situated within a relevant context. This philosophy puts context at the focus of events and indeed, defines context as the explanatory process. The antithesis to that is the understanding of events in terms of their form or topographical features. This essentially means that any event, no matter how identical its form to other phenomena, can be considered to be functionally different given the different contexts that may give rise to that event. For instance, a hand wave could mean both hello and goodbye depending on the context in which it occurs and to this extent, the event cannot even be conceived fully without making reference to the situational context. Contextual effects are clearly widespread within language in which words with numerous meanings have their effects because of the context in which the word is used. One illustration of this is the German “bitte”, which, in colloquial use, means both “please” and “you’re welcome”. As such, it is the context rather than the form of the word that determines the function of the behavioural event and establishes what might be called meaning.

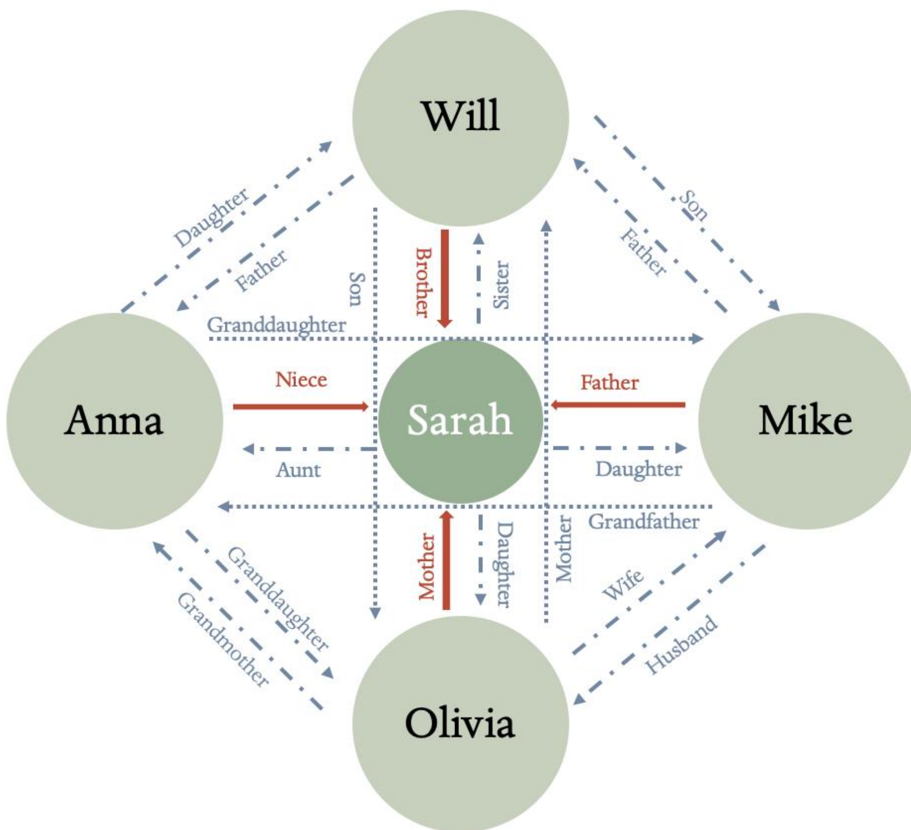
To understand RFT, the reader must first understand what is meant by derived relational responding. Relating simply describes the process whereby an individual responds to one event in terms of another. Where responding to a stimulus involves responding to it in terms of another stimulus only indirectly related to it, the behaviour can be referred to as derived relational responding, which is arbitrarily applicable given the appropriate contextual cues. As a post-Skinnerian account of human language and cognition, RFT arises out of radical behaviourism and Skinner’s analysis of language outlined in Verbal Behaviour (1957). While Skinner focused his attention upon the behaviour of the speaker and on direct training in language acquisition, this approach was met with criticism, due to the unrealistic requirement for every individual word function to be established separately (Chomsky, 1959; see also O’Toole et al., 2009). RFT addresses this shortcoming by conceptualising language proficiency as a generalised operant involving a series of overarching operants that involve a small range of verbal behaviours whilst employing a vast array of word forms produced by the speaker and responded to by the listener. In this account, both speaking and listening are forms of generalised operant responding to common forms involving unique stimuli on each occasion. RFT also takes the view, as suggested by Sidman (1994), that the analysis of human language will have to require an analysis based heavily or entirely on analyses of derived relational responding. While it was Sidman’s work on deriving equivalence relations that provided the impetus for RFT to explain the rapid acquisition of derived relational responding abilities in humans, RFT, in fact, elaborates on Sidman’s account of language, which is focused on equivalence and non-equivalence alone. Specifically, it introduces the analysis of a range of other patterns of relational responding that align with other relational concepts, such as difference and comparison. Importantly, it provides a fundamental account of these phenomena based on fundamental learning principles and without introducing mentalistic concepts.

As illustrated by the theory's name, the process of relating is fundamental to RFT. A simple example of derived relational responding is responding to a series of interrelated statements that together constitute a relational network. For instance, the relations "A is bigger than C" and "B is smaller than C" might be established through merely telling an individual or perhaps across multiple trials in an operant conditioning context. Based on these established relations, an individual can derive that C is smaller than A and that C is bigger than B. However, more importantly, the individual can derive the relationship between A and B, even though no direct relationship between these relata has been trained. This derivation is solely based on responding to these two stimuli in combination with their relations to yet further stimuli. As a result, they can respond to "A is bigger than B" in the laboratory. In practice, these relational networks can become incredibly complex, and the relations can take various forms (Hayes et al., 2001). In fact, it is precisely this complexity that makes the analysis of language in terms of networks of related stimuli under the contextual control of contextual cues (such as the words opposite or more) conceptually appealing.

As a more elaborated example consider the relational network of social relations illustrated in Figure 1. This figure illustrates that by establishing only four familial relations, 16 further relations can be derived consistent with those trained. In other words, the derivation of relations is an excellent explanation of generative language and its asymptotically increasing complexity with the additional establishment of words in the vocabulary. Indeed, establishing as few as eight stimulus relations can allow for the derivation of several thousand (see Hughes & Barnes-Holmes, 2016).

Figure 1

Familial Network of Established and Derived Relations



Note. The red lines indicate the four trained relations, while the blue lines indicate the 16 relations that can be derived from the explicit training of the initial four relations.

According to RFT, it is the acquired skill of derived relational responding which underlies most, if not all of higher human cognition. Indeed, there is a whole host of research supporting the assertion that basic stimulus equivalence fluency facilitates normal language functioning (Barnes et al., 1990; Dugdale & Lowe, 2000), the developmentally normal acquisition of reading skills (de Rose et al., 1992; Farrington-Flint et al., 2007; Mackay, 1985; Sidman, 1971), as well as the development of an appropriate vocabulary (Edwards et al., 2011; McHugh et al., 2004; Nippold & Sullivan, 1987), grammar (Hakes, 1965), and spelling (Brown et al., 2006; Mackay, 1985). However, various studies support the suggestion that relational responding is not special to humans (Harmon et al., 1982). More specifically and complex, animals have been found to engage in transposition, the responding based on relational rather than absolute properties. Examples for this are bees (Giuarfa et al., 2001), fish (Perkins, 1931), monkeys (Harmon et al., 1982), and pigeons (Wright & Delius, 1994). Therefore, RFT identified another essential component relevant to make human cognition stand out from non-human animals, namely, arbitrarily applicable relational responding.

Arbitrarily applicable relational responding refers to relational responding based on arbitrary (i.e., non-physical) stimulus properties rather than properties that can be responded to directly on the presentation of the stimulus. These arbitrary properties are determined conventionally by integration into a culture and in the acquisition of a language. For instance, the threatening functions of a stimulus, such as a gun, are not established by its formal features. Indeed, such a stimulus could not produce fear in an organism that had not participated in the culture in which its function had been established arbitrarily. While this example does not exclude the function's establishment through direct association with aversive consequences, this fails to explain the large number of people who are uncomfortable in the presence of guns.

Instead, for most individuals, the aversive functions of guns as weapons are established arbitrarily through language, making the direct exposure to the threatening attributes of a gun redundant to establish a fear response. The same may be true for a whole host of other stimuli with aversive (e.g., Boyle et al., 2016; Dougher et al., 2007; Leech & Barnes-Holmes, 2020), sexual (e.g., Roche et al., 2000), or other functional properties. In contrast, non-arbitrary properties are those based on physical attributes such as colour, shape, and size. One simple example of a context in which arbitrary and non-arbitrary properties compete against each other in the absence of clear contextual control is a young child's response to the choice of a coin from an array of differently valued coins. More specifically, when presented with two euro coins of different values, such as a 5-cent and a 10-cent coin, and given the choice to take one of these, the child may respond to the non-arbitrary relation of physical size that obtains between these two stimuli by likely choosing the €0.05 coin due to its larger size compared to the €0.10 coin. However, in the context of value, which is not yet controlling the response of the younger child, the invitation to take whichever coin has the greatest value will lead to the choosing of the €0.10 coin by an older child for whom arbitrarily applicable relational responding has been established as a generalised operant. In effect, responding to the €0.10 is based on its arbitrary rather than formal relationship to the €0.05 coin.

It is important to understand in this example that this is the contextual cue relating to value in the speaker's verbal behaviour that controls the appropriate choice. This shift in reliance from formal to arbitrary stimulus relations represents the shift from concrete to abstract reasoning allowing for complex relational responding to emerge across a wide range of contexts in the education of a child (Stewart & McElwee, 2009). Without an understanding of the fact that operants can generalise in this way to involve responding to stimuli regardless of their formal features, the analysis of verbal behaviour would be extremely limited. In effect, it is important for RFT that researchers start to think of stimuli as related not in terms of the class concept but in terms of the concept of relation, which is more fluid and leads to a wider range of response outcomes when it is conceived using the concept of stimulus equivalence classes (Hayes & Barnes-Holmes, 1997).

The derivation of relations entails three defining properties known in RFT as *mutual entailment*, *combinatorial entailment*, and the *transformation of function*. Mutual entailment describes the property whereby a relation specified in a certain direction entails a second relation in its opposing direction, in other words (e.g., if “A > B” is trained, the mutually entailed relation is “B < A”). As such, mutually entailed relations are always bidirectional, though not always in a symmetrical manner. Such specified relations are always entailed at identical levels of precision.

Combinatorial entailment expands upon mutual entailment by combining multiple (derived or trained) relations. If “A > B and B > C” is trained, then “A > C” can be derived. These combinations, however, result in lower levels of precision for the entailed relations compared to the original ones (e.g., if A ≠ B and B ≠ C, it is impossible to derive relations between A and C in either direction as not enough information is presented to do so). The attribute of combinatorial entailment is required to define relational frames, as lower levels of complexity impede upon the emergence of distinct patterns by allowing for derived relations between unrelated stimuli (e.g., instead of deriving B < A when presented A > B, it is now possible to derive C < A, when only A > B and B > C were presented). In short, if the direct relation is specified/trained, the property is referred to as mutual entailment; if the relation is implied/derived, it is combinatorial entailment).

The *transformation of stimulus function* describes the process whereby the function of one stimulus extends that same function to other related stimuli accordingly. To illustrate, if relatum A is taught to be feared and to be related opposite to relatum B, then the individual derives that B should not be feared (Cassidy et al., 2010). For this to occur, the process itself must be under contextual control. This means, that not all functions of a given stimulus transfer to the related stimulus, but instead, only those of functional importance.

While responding to stimuli in terms of the patterns outlined above might be taken for granted by a verbally sophisticated adult human, the role of the scientist is to nevertheless explain how these seemingly everyday skills emerge in the first instance. Given the importance that RFT places on these skills it is even more important to understand how they can be established well and for their acquisition to be accelerated where this is necessary. Moreover, RFT argues that the ability to derive relations appears to be unique to humans, but not necessarily so (see Dymond, Roche & Barnes-Holmes, 2003). Indeed, research supports this suggestion. Specifically, Dugdale and Lowe (2000) found that whilst language-trained chimpanzees were able to engage in relational responding based on symmetry relations, they failed to respond in accordance with derived relations, even in its simplest form of word-object symmetry relations. These word-object relations had to be trained bidirectionally for the chimpanzees to respond to the stimuli bidirectionally. Of course, studies show that various nonhuman animals, including dolphins (Herman & Thompson, 1982), rats (Lashley, 1938), pigeons (Edwards et al., 1982), and monkeys (Nissen, 1951), are able to demonstrate conditional relational responding, but no study to date has found this to result in stimulus equivalence. Devany et al. (1986) argued that, even if nonhuman animals were found to form stimulus equivalence classes, the ease and reliability with which humans acquire such classes is what makes human cognition remarkable and worthy of study.

Contrary to Sidman's conceptualisation of derived relational responding as a basic stimulus function requiring no further or no more fundamental explanation (e.g., Sidman, 2000), RFT conceives of derived relational responding as a learned operant, thus making it malleable to environmental intervention. While the initial acquisition of stimulus relations within a network must be explicitly and bidirectionally trained, the derivation of those relations, combined with an extensive history of multiple exemplar training, is thought to result in the development of the generalised operant skill of derivation. As such, early language acquisition requires the caregiver to engage in guided derivation of a range of types across a large array of stimulus objects in order for this skill to generalise. For instance, they might show a child an object in order to pair it with a spoken word. Variations in the presentation of the word and the object mean that the relation is established in both directions (i.e., word-object and object-word) repeatedly across different occasions. Additionally, the caregiver must reinforce the derivation of appropriate relations across numerous relata and teach this skill with different relation types such as comparison or opposition. With sufficient exemplar training over a sufficient amount of time the various relational operants become generalised. These relational concepts are referred to as relational frames, a concept that is treated carefully and understood as a verb rather than a noun. Various studies support the proposal that the derivation of this operant skill needs to be acquired through external reinforcement agencies and that without such intervention, derived relational responding would likely not emerge for humans.

Support for the assertion that derived relational responding is an acquired operant is widespread. Barnes-Holmes et al. (2004), for instance, found that repertoires of relational responding in accordance with the oppositional frame emerged for children as a generalised operant where this repertoire had previously been found to be absent. In addition, multiple exemplar training was successfully applied to establish derived relational responding fluency at both low and high levels of complexity. Another noteworthy study was conducted by Lipkens et al. (1993). This study aimed to establish derived relational responding in an infant for whom it was absent, in an effort to usher it in prematurely. The infant was 16 to 27 months old during the study. Derived mutual entailment relations and nonverbal exclusion were found as early as 17 months, whereas combinatorial entailment relations and verbal exclusion took longer to emerge. Lipkens et al. (1993) conclude that derived relational responding is not reliant on sophisticated inherent verbal abilities but can develop earlier if sufficient training is provided. In other studies, young children between the ages of three and six years old were found to fail to derive such relations and require explicit training to form such concepts (e.g., Barnes et al., 1995). These findings are complemented by Devany et al.'s (1986) study, which found that only language-abled children were able to form equivalence classes in a brief laboratory training procedure, whereas nonverbal children failed to do so. To the behaviour analyst, this suggests that the skill of derived relational responding is an acquired operant; a proposition that differs from other perspectives, which describe relational responding as a mental mechanism (see McLoughlin et al., 2020). Conceiving of derived relational responding as an acquired operant, by extension, implies this skill to be malleable, subject to environmental interventions (Barnes-Holmes et al., 2001, 2004a, 2004b; Berens & Hayes, 2007; Cassidy et al., 2010; Dixon et al., 2021a, 2021b, Weil et al., 2011).

The notion that relational responding functions as a generalised operant is widely recognised in the RFT literature and among its researchers. While the mentioned studies support this perspective, it is important not to disregard recent critiques that question the breadth and depth of the RFT literature. Specifically, a recent paper that raises concerns about the underlying assumptions of Acceptance and Commitment Therapy (ACT) also presents broader criticisms regarding RFT. Despite being RFT researchers (with a partially vested interest) themselves, McLoughlin and Roche (2023) do not shy away from addressing common issues within the RFT literature. They discuss that while studies like O'Connor et al.'s (2017) report that the 521 papers from 2009-2016 make RFT appear to be a well-supported and frequently cited theory, a closer examination reveals that only 55% of these studies are empirical. This suggests an overrepresentation of theory-based and conceptual work. Furthermore, of the existing empirical studies, more than half do not classify as "RFT," and nearly a third of those that remain focus specifically on implicit bias. McLoughlin and Roche (2023) critique that such studies fail to provide substantial theoretical support for RFT. This raises the significant question of whether the empirical evidence is adequate to persuade non-behaviourists of RFT's value in explaining human language and cognition. Although RFT is a *promising* theory for understanding language and cognition, the authors also highlight the issue of small sample sizes prevalent in many relational responding studies (Dymond & Barnes, 1995; May et al., 2017; McHugh et al., 2004; McLoughlin & Stewart, 2017; McLoughlin et al., 2018; Steele & Hayes, 1991; Villatte et al., 2010). The problem with underpowered studies lies in their findings' generalisability, which is further complicated by inconsistencies in procedures and outcome measures. Therefore, while studies with limited sample sizes can provide valuable insights into potential underlying conditions, the authors call for adequately powered studies and standardised procedures to facilitate more generalisable conclusions.

Responding in accordance with relational frames is controlled by two types of contextual cues according to the literature and RFT. One type of queue specifies the relationship according to which one stimulus would be responded to in terms of another. This type of cue is called a relational queue and is denoted as $C_{rel.}$ Examples of such queues include the words same, opposite and so on. Their presence guides responding towards the stimulus in particular relational terms. However, other contextual features control which stimulus functions will be responded to and transformed in the course of relational responding. These types of cues are referred to as functional cues and are denoted as C_{func}. Examples of functional cues are words like taste, sound, or appearance. Functional contextual cues specify the psychological properties that are produced by the stimulus when it is responded to in terms of another stimulus. For example, an individual could be instructed to imagine something that tastes like a lemon (i.e., bitter) or, in contrast, to imagine something that tastes the opposite of a lemon (i.e., sweet). In the first case, the phrase “like a" functions to specify an equivalence relation between the imagined stimulus and the lemon, whereas in the second case, the contextual cue specifies the relationship of opposition. Importantly, however, not all of the functions of a lemon or something opposite to a lemon are produced by these statements. Rather the psychological properties of the stimulus represented by the relevant statements, is constrained to the domain of taste by the functional contextual cue “tastes”.

While the number of frames that characterise human language is potentially infinite, the primary frames that likely control the majority of most everyday complex language activity are likely: coordination (Steele & Hayes, 1991), distinction (Roche & Barnes, 1997), opposition (Barnes-Holmes et al., 2004a, 2004b; Steele & Hayes, 1991), comparison (Dymond & Barnes, 1995), temporality (O'Hora et al., 2008), analogy (Stewart et al., 2004), hierarchy (Slattery & Stewart, 2014), and deixis (McHugh et al., 2004). These will briefly be outlined below.

The most fundamental frame is that of coordination (sometimes referred to as sameness). The frame of coordination can be considered akin to equivalence relations but with the proviso that it is not considered a stimulus class and is subject to other processes, such as the transformation of stimulus functions. Importantly, coordination functions as the basis for all word-object relations and, therefore, plays an invaluable role in language acquisition. Once the fundamental proficiency to engage in coordination relations has been established sufficiently, the individual has the basic skills necessary to acquire subsequent patterns of relational framing such as opposition. That has not yet been proven experimentally but the reasoning goes that combinatorically entailed relations in opposition frames inherently include frames of coordination and therefore that this form of relational responding is highly related to basic coordination responding. Because it follows that an individual must first acquire the concept of similarity before they can acquire the concept of opposition, it stands to reason that the frame of coordination is acquired first, although this sequence of events is not necessary from this point of view. One example of an oppositional relational frame is the phrase "big is opposite to small". The frame of opposition always specifies the continuum along which the relata are related, but they are always at extreme ends of a precise continuum whether it be formal or abstract. In other words, the relata always share some common properties that under the appropriate contextual control can render the relata identical or opposite. For instance, the terms hot and cold are both temperature labels and in that sense are coordinate. However, insofar as they refer to two extreme temperature labels they are opposite.

The frame of distinction is similar to that of opposition in that it refers to a difference between relata. However, there are two ways in which responding in accordance with a frame of difference can come about. The first is when there is no obvious continuum according to which the relevant stimuli are related. In other words, the stimuli may share no common properties at all and not participate in any relations in which they could be coordinate. An example of this might be to say that hot is different to large. In this case, the relevant stimuli are neither coordinate nor opposite. However, one stimulus can also be responded to as different to another when in fact the relata do share a property and are related to each other along a specifiable continuum, but where the relata are not at extreme ends of that continuum. For instance, the statement that warm is different to hot is meaningful in terms of a frame of difference, whereas the statement "warm is opposite to hot" is not meaningful. The latter form of difference responding does not allow the individual to make any inferences about the degree to which the relata differ. For instance, saying "this water is different to warm" does not provide enough information for the individual to derive whether the water is freezing or boiling. By extension, all opposition relations are specific types of distinction relations under a more restrictive form of contextual control.

The frame of comparison describes a pattern of relational responding in which relata are related along a continuum, but the contextual cue specifies the valence in the difference between one relata and the other. The continuum can be physical or can be itself a relational abstraction. For instance, a statement can specify that "John is half the age of Anna". Or it can be less specific in saying that John is younger than Anna. In both cases the continuum is quantitative. However, it is also possible to specify a qualitative relation between the two relata, such as saying that John is a nicer person than Anna. Therefore, comparative frames are particularly specific in controlling responding to one stimulus in terms of another in relation to precise distances from each other along some continuum.

The frame of hierarchy specifies relations in which one relatum is contained "within" another (e.g., "an apple is a type of fruit"). Due to their lack of focus on mere quality and quantity, hierarchical frames provide more precise insight into combinatorial relations regardless of the presence of quantification. To illustrate, stating that "dogs are animals" and "dogs have four legs" allows the individual to derive that (at least) some animals have four legs. This allows for specificity in hierarchical relations that is not provided in comparison relations. For instance, saying that "Paul is Tim's and Daniel's father", allows the individual to derive that Tim and Daniel are siblings. However, saying that Daniel is taller than Tim and Paul (a comparative relation) does not allow one to derive whether Tim or Paul is the taller of the two. It is reasonable to assume that this form of relational responding depends upon the development of coordination, distinction, opposition and comparison relational framing.

The frame of temporality specifies before/after relations and thus focuses on differences between relata based on time. One such example is "morning comes before afternoon". This frame is unique in the way in which its arbitrary dimension (i.e., time as a concept) and non-arbitrary dimension (i.e., time as it only refers to unidirectional change) interact. The concept of time is fundamentally verbal as it requires an abstraction of the physical dimension along which temporal or causal comparatives (e.g., yesterday, today, tomorrow) are ordered. In other words, it is impossible to physically represent the idea of "yesterday".

One form of higher order relational responding that likely builds upon those discussed to this point is that of analogy. This form of derived relational responding refers not to relations between stimuli but on relations between relations. More specifically, it refers to the specific way in which one set of relations relates to another (e.g., “foot is to sock, as hand is to glove”). Due to the increased complexity of this frame, it is believed that analogical relations in particular are of fundamental importance to complex cognition (discussed in further detail in section 1.5.3).

Finally, deictic relations are those that entail perspective-taking (e.g., “I am here and you are there”). They are the only relations listed here which are always to be interpreted from the speaker’s point of view and thus define the difference between deictic and non-deictic relations. If the relational term (e.g., front and back) relates to the speaker’s point of view (e.g., “I am behind the lady in the blue shirt”) the relational frame at question is one of deixis. However, if the relational term does not apply from the speaker’s point of view, but rather describes a general attribute (e.g., “the toolbox is by the backdoor”), the frame in question is not deictic. Deictic relations further stand out from other frames due to their special attribute of missing any nonarbitrary or formal equivalents. These frames are unique in that it is impossible to trace them back to environmental dimensions; it is solely the relationship between the individual and the other relata that functions as the dimensional basis for this frame. Questions like, “What are *you* doing *now*?” or “What am *I* doing *here*?” never require an interaction with the environment in terms of relational properties. Therefore, the acquisition of this frame is thought to be abstracted from the individual’s exposure to someone else’s point of view. For instance, acquiring the ability to respond meaningfully to the question “what did *you* do when you got *there*?” relies upon the listener’s acquisition of a generalised operant response repertoire to the question enquiring about their own (and not someone else’s) perspective. “You” always refers to different people; it is not a given name (e.g., Alexander) of which there are usually no more than perhaps two or three (if not just one, as in most contexts) in any given room.

RFT's proposition is that the proficiency to engage in different forms of relational framing does not develop simultaneously, but that the mastering of one is required for the acquisition of another (Hayes et al., 2001). Therefore, some frames might facilitate other frames to emerge better than others. This is a crucial question from a theoretical point of view and is only just beginning to be explored (e.g., Cummins et al., 2023). RFT further proposes that the "language explosion" and a sufficient repertoire of relational networks align developmentally (Stewart & Roche, 2013). This makes sense, as an extensive repertoire of relational networks facilitates vocabulary, or, to be more specific, is what defines vocabulary. It also follows that an increase in the number of frames a concept is engaged in consequently increases the detail and precision of its definition. While the above outlines *how* relational responding underlies complex cognition, the following section will outline empirical evidence for the close correlation between the two.

1.5.3 The RFT Account of Human Intelligence

Conceiving of higher cognition as generalised operant relational framing immediately allows for the scientific analysis of higher-level cognitive processes precisely because these are highly contextually controlled and nuanced. Cognitive activities involving creativity and problem-solving appear not to be controlled by direct contingencies, an issue that according to RFT, can now be understood in functional terms. That is, these types of activities appear to involve combinations of complex forms of derived relational responding for which there is no obvious history of reinforcement (Hayes et al., 2001). It is no wonder, therefore, that intelligence was conceived as a trait because the functional analysis of the emergence of apparently intelligent behaviour in the repertoire of an individual has not hitherto been possible (see Cassidy et al., 2010; Schlinger, 2003). Importantly, mainstream intelligence theorists have identified flexibility and complexity in behaviour as constituting an important component of what is called intelligent activity (Cattell, 1971; Kyllonen et al., 1984). This aligns well with an emerging RFT view of intelligence, as will be discussed below. Both psychometricians and behavioural analysts alike concur that cognitive inflexibility is a hallmark of low intellectual acumen and is associated with several psychological disorders (Lovecky, 2004; Turner, 1999; Wulfert et al., 1994).

Another similarity between traditional theories and RFT can be seen in the focus on relational responding. Spearman (1904; 1927) argued that the best reflection of *g* comes in the form of the eduction of relations and correlates, part of Spearman's noegenetic laws (described in more detail in section 1.1.2.1). As Jensen (1998) explained, this refers to the ability to understand relationships, inductive and deductive reasoning, the inference of rules, generalisation, the recognition of similarities and differences and the decontextualisation of problems. The parallels to an RFT account of intelligence are easily apparent. Indeed, traditional IQ tests, which are heavily built on Spearman's idea of intelligence, align quite well with the types of content that might be inserted on those tests by a relational frame theorist. Specifically, Cassidy et al. (2010) argued that even traditional IQ tests can be conceived of as measures of derived relational responding proficiency. Indeed, scrutinising intelligence test items easily allows anyone familiar with RFT to see how many of the items depend on the test-takers ability to identify relations of coordination. One common task is to identify word-word or word-object coordination relations, such as being asked to pick the picture of the bird when presented with a number of visual stimuli. These tasks are included in a variety of subtests such as WAIS Vocabulary, WISC Vocabulary, WASI Vocabulary, Stanford-Binet Vocabulary, K-BIT Verbal Knowledge, WJ Verbal Comprehension, WJ Rapid Picture Naming, WJ Visual Auditory Learning, WJ Picture Vocabulary, DAS Naming Vocabulary, WPPSI Receptive Vocabulary and WPPSI Picture

Naming (Colbert, 2015). Beyond word-word or word-object relations, a common task is to identify non-verbal coordination relations relating to physical characteristics (Colbert, 2015). Examples for this type of task include the Processing Speed section (Visual Matching, Decision Speed & Cross-out) of the Woodcock-Johnson Test of Cognitive Abilities, as well as WISC Cancellation. While this is only a small selection, examples of relational responding tasks in traditional IQ tests are plenty and span across virtually all commonly used test types and items (Colbert, 2015).

Within the behavioural field, several studies have identified response variability and relational responding as providing an appropriate basis for a wide array of higher-level complex behaviours (Barnes, 1994; Barnes et al. 1995; Catania, 1996, 1998; Hayes & Barnes, 1997; Hayes, 1991; Healy et al., 1998; Horne & Lowe, 1996; Lowekron, 1998; Roche & Barnes, 1997; Smeets et al., 1997). More specifically, a wide range of studies support the idea that relational responding underlies general intellectual function as defined and measured by standardised IQ tests (e.g., O'Hora et al., 2005, 2008; Gore et al., 2010). Furthermore, studies conceived from a cognitive perspective, support the proposal that a better proficiency of abstract reasoning is positively correlated with intellectual functioning (Andrews & Halford, 1998; Cattell, 1971; Gentner & Loewenstein, 2002). Indeed, recent studies have found substantial evidence for the strong correlation between relational responding and intelligence. More specifically, Colbert et al. (2017) found relational responding proficiency, as assessed by the RAI, to significantly predict performance on various measures of cognitive ability. Their first study found significant moderate correlations between RAI scores and performance on the NART (Nelson, 1982) and the Trail Making Test (Lezak, 1995). Moreover, a significant, strong, positive correlation was found between RAI scores and performance on the Rey Auditory Verbal Learning Test (RAVLT;

Rey, 1985; English version: Taylor, 1959). Given that both the NART and the RAVLT heavily rely on verbal processes, their correlation with the linguistically based RAI might be unsurprising. However, in their second study, Colbert et al. (2017) found moderate-to-strong significant correlations between RAI scores ($M = 48.2$) and Full Scale ($M = 117.2$, $r = .74$, $p < .001$), Verbal ($r = .78$, $p < .001$), and Performance IQ ($r = .55$, $p < .01$), as assessed by the *Wechsler Adult Intelligence Scale III* (WAIS-III; Wechsler, 1997). RAI performance showed significant moderate-to-strong correlations with the three main indices of IQ as well as four further subindices and 10 out of 13 subtests, thus suggesting a high degree of correlation between relational ability and IQ. More specifically, the moderate-to-strong significant correlations between RAI scores and each of the WAIS' Verbal subtests indicate the importance of relational responding, as assessed by the RAI, to linguistic performance (Colbert et al., 2017). This further supports the RFT proposal that relational responding underlies language processes. Based on Colbert et al.'s (2017) study findings, it can be cautiously concluded that relational skills appear to most strongly underlie intelligence in the verbal domain. The finding that RAI performance only significantly correlated with three out of six of the Performance IQ subtests might indicate that the RAI more closely resembles verbal tasks rather than performance tasks in terms of intellectual skills repertoires, argue Colbert et al. (2017).

These findings are strongly complemented by Colbert et al.'s (2020) study, which found significant correlations between the RAI+ (a reiterated version of the RAI) and FSIQ (.52), Verbal IQ (.42), and Performance IQ (.48), as assessed by the WASI IQ (Wechsler, 1999). Furthermore, it found significant correlations between performance on the RAI+ and the WASI IQ's subtests, namely Vocabulary (.37), Similarities (.37), Block Design (.42), and Matrix Reasoning (.48). Additionally, RAI+ performance significantly correlated with WAIS-III

Arithmetic scores (.43) but not WAIT-T scores, suggesting relational abilities to be critical to numeracy, but failing to find a correlation with scholastic ability, as assessed by the WAIT-T. These findings suggest the importance of relational abilities to intelligence test performance and numeracy, further highlighting their relevance to higher cognition. Colbert et al. (2017) further found that out of the 13 WAIS subtests assessed in their study, those with the highest correlation with the RAI were the same ones previously found to have the highest *g* loadings (Gignac, 2006). These correlations are of great significance to the functional analysis of behaviour insofar as they raised the exciting possibility that fluency in derived relational responding either underlies complex intellectual activity or, in fact, even defines it.

Moreover, the frame of coordination is likely the most central relational frame pertaining to the acquisition of vocabulary (Stewart et al., 2013). A strong relationship between the extent of an individual's vocabulary and their general intelligence has been widely noted and appears to be bidirectional (Smith et al., 2005; Marchman & Fernald, 2008; Vetterli & Furedy, 1997). That is, vocabulary acquisition is predicted by intelligence but simultaneously positively affects intellectual development. Therefore, it might not be surprising that behaviour analysts have observed strong correlations between relational responding proficiency as assessed in a variety of ways and cognitive performance (Barnes-Holmes et al., 2005a, 2005b; Cassidy et al., 2011; Moran et al., 2010; O'Connor et al., 2009; O'Hora et al., 2008; O'Toole & Barnes-Holmes, 2009). In the same vein, comparison relations might be considered fairly fundamental to the acquisition of mathematics (Stewart et al., 2013), as they describe relations reliant on the responding to one event amongst a specified quantitative or qualitative dimension with another event (Hayes, Fox, et al., 2001). Premises such as "given X, show that Y" (Marr, 2015) perfectly exemplify the relational nature of mathematics. Moreover, O'Hora et al. (2005, 2008) and

O'Toole & Barnes-Holmes (2009) further argued that temporality relations are strongly related to intelligence, as they significantly positively correlated with performance on FSIQ as assessed by the WAIS-III and the K-BIT.

Interestingly, behaviour analysts have also zoned in on analogy as critical to higher cognitive ability and several studies have examined the acquisition of analogies as a highly complex form of human verbal behaviour (Stewart et al., 2001, 2013). The contention that analogical reasoning can be understood in terms of relational responding is supported by findings involving the analysis of brain activation during both standardised and logical reasoning tasks and standard equivalence reasoning tasks in the laboratory (Barnes-Holmes et al., 2005a, 2005b). Moreover, the studies found analogical reasoning to be more complex than simple comparison relations, due to their involvement of multiple frames. This is not only reflected in the RFT literature, but also in measures of the event-related potentials of these two response patterns (Barnes-Holmes et al., 2005a). Indeed, analogical relations are perhaps the most complex, but simultaneously the most widely recognised to be of fundamental importance to intelligence out of all the relational frames (Esher et al., 1942; Sternberg, 1977).

An RFT view of intelligence (e.g., Cassidy et al., 2010, McLoughlin et al., 2020) involves approaching various types of complex behaviour indexed on many widely used intelligence assessments, and illustrating or investigating how those tasks might be understood as forms of highly contextually controlled and complex derived relational responding. Importantly, the idea here is not to try to preserve the concept of intelligence as a psychometric trait based and mentalistic concept, but rather to reduce it to more fundamental functional analytically understood terms. In effect intelligence would therefore not be viewed as a trait or as a construct but simply as the description of the properties of a behavioural repertoire. As a result, it would be inappropriate to equate derived relational responding with *g*, for example. More correctly, the apparent stability of *g* might be understood in terms of the relative stability in the fluency of a relational skills repertoire not attributable directly to the individual but to their historical and current context.

Beyond the RFT approach suggesting the crucial importance of relational responding proficiency for intelligence, alternative psychological perspectives and, indeed, even other disciplines have recently all started to converge on the idea that relational responding underlies complex cognition. More specifically, these accounts come from cognitivist theories (e.g., Halford et al., 2010), computational models (Doumas et al., 2018), and even artificial intelligence research (Hammer, 2022). The fact that so many approaches and disciplines have independently arrived at the same conclusion highlights the plausibility and the strong correlation between intelligence and relational responding abilities. However, relative to RFT, these other accounts are still infantile. In the decades since the theory was proposed, RFT has produced a robust research programme encompassing not only a detailed investigation of the derived relational responding-intelligence connection but, perhaps more impressively, a theoretically based intervention (aimed to increase intelligence by training relational responding) and an assessment of relational responding repertoires. The following two sections will outline the intervention and the assessment that have been produced by this extensive research programme to date.

1.6 Enhancing Intellectual Performance

1.6.1 Previous Attempts to Enhance IQ

As explained in section 1.2.1, Binet and Simon's motivation behind the creation of the first measure of intellectual function was to assess an individual's current abilities in order to identify any deficits, which could then be targeted through environmental interventions. However, this idea of the malleability of intelligence was quickly replaced by the, nowadays common, view that intelligence is unmalleable and individuals either have it since birth or they fall behind their peers across their entire lifespan. Nonetheless, attempts to increase intelligence are numerous. A brief review of proposed and tested interventions quickly makes apparent that the attempts to increase intelligence that were taken in the past appear rather unsystematic and "blunderbuss". Such interventions included the use of chess (e.g., Aciego et al., 2012; Kazemi et al., 2012; Sala & Gobet, 2016), music instruction (e.g., Benz et al., 2016; Bergman Nutley et al., 2014; Franklin et al., 2008; Miendlarzewska & Trost, 2014), as well as video games (e.g., Ballesteros et al., 2014, 2015; Dye et al., 2009; Green & Bavelier, 2012; Toril, Reales, & Ballesteros, 2014; Toril et al., 2016; Wang et al., 2016). While some success has been found, a closer investigation of methodological and statistical practices calls for caution in the interpretation of the findings (Haier, 2014; Melby-Lervag & Hulme, 2013; Owen et al., 2010; Sala & Gobet, 2017; Simons et al., 2016).

Despite the reservations expressed by many traditional psychometricians regarding the malleability of intelligence (e.g., Haier 2014), one specific intervention provided a challenge for sceptics to reconsider their standpoint. More specifically, several studies produced by Susan Jaeggi et al. (e.g., Jaeggi et al., 2008; Jaeggi et al., 2010; Rudebeck et al., 2012) appeared to suggest that intensive working memory training could raise IQ as measured by Raven's Matrices and other measures of fluid intelligence. The results initially led leading theorists such as Robert Sternberg (2008) to consider the possibility that fluid intelligence can be improved to a significant and meaningful extent. However, the excitement was relatively short lived as shortcomings of the early studies began to come to light. For example, even the most successful working memory training interventions only increased intelligence by about 4 IQ points. However, more importantly, perhaps, are the methodological issues that were raised in the literature. Perhaps the most major issue is the inconsistency of measures used to assess improvements in intelligence scores from baseline to follow-up which remarkably were sometimes different within the same study (Moody, 2009). In addition, the original 2008 study had no baseline measures at all and made estimates of gains compared to a control group based entirely on population norms. Moody (2009) also criticised the reduced time limit in the control group's intelligence assessment, allowing only 10 minutes rather than the recommended 45 minutes, which might have considerably impacted performance.

Comparing working memory training interventions to alternative approaches highlights some theoretical advantages of this endeavour. Working memory has been found to contribute to a variety of cognitive functions, such as arithmetic (Geary et al., 1999; Swanson & Sachse-Lee, 2001), fluid intelligence (Ackerman et al., 2005), learning (Cain et al., 2004; Cowan & Alloway, 2008), reading comprehension (Cain et al., 2004; Carretti et al., 2009; Daneman & Carpenter, 2004), reasoning (Kyllonen & Christal, 1990), scholastic aptitude (Alloway & Alloway, 2010), and writing (McCutchen, 1996). Therefore, targeting working memory to improve intellectual function was theoretically coherent. Nonetheless, the specific mechanism through which working memory might affect intelligence is yet to be explained. This criticism of a lack of an understanding of the mechanisms further translates to other previously proposed interventions. To that point, Schubert et al. (2014) and Simons et al. (2016) raised the concern that studies reporting on interventions aiming to increase intelligence rarely, if ever, provide an explanation for why the training protocol should be efficacious. Providing a strong theoretical, and ideally empirical, case for the development of such interventions is therefore essential for the practice of good science. As it happens, this is precisely what an RFT approach brings to the table.

1.6.2 Interventions Based on Relational Frame Theory

The review provided in the previous section makes it apparent that, so far, previous attempts have failed to increase intelligence by a meaningful degree. Indeed, such a lack of empirical evidence for the possibility of raising intelligence allows one to understand the critical position that most traditional intelligence theorists take. From a behaviour analytic view, it might be the case that the failure to raise intelligence most likely has to do with the absence of sound functional analytic models specifying precisely which behaviours require enhancing in fluency. The correlations outlined in section 1.5.3, however, provide sufficient grounds to believe that it is the relatively well-understood behavioural phenomenon of derived relational responding, which underlies intellectual activity rather than vice versa. Understanding intellectual abilities in functional terms as an acquired generalised operant immediately raises the possibility of enhancing the fluency of a wide range of behaviours considered germane to performance on traditional assessments of intelligence (i.e., vocabulary, numeracy). Indeed, two independent research strands within behaviour analysis seem to suggest exactly this; that intellectual ability, as measured by gold standard IQ tests, can be enhanced by training relational responding skills. Unsurprisingly, both interventions are based on RFT. The first intervention is known as PEAK (Promoting the Emergence of Advanced Knowledge; Dixon, 2014a), while the second one is known as SMART (Cassidy et al., 2016). Each will be briefly outlined as part of the case that intelligence can be conceived functionally in terms of derived relational responding.

1.6.3 PEAK

The PEAK training system relies on the idea that training relational responding consequently increases intellectual function. It consists of four modules, the PEAK-Direct Training (PEAK-DT; Dixon, 2014a), which teaches foundational learning skills via contingency-based learning, the PEAK-Generalisation (PEAK-G; Dixon, 2014b) which uses the same approach of contingency-based learning but applies this to an alteration of training and testing stages aimed to assess generalisation and responding, the PEAK-Equivalence (PEAK-E; Dixon, 2015), which trains stimulus equivalence and measures derived relational responding proficiency, and finally, PEAK-Transformation (PEAK-T; Dixon, 2016), which aims to improve derived relational responding skills within the frames of opposition, comparison, distinction, hierarchy, and deixis. While PEAK-DT and PEAK-G focus on the acquisition of basic verbal skills, PEAK-E and PEAK-T focus on the acquisition of generalisable arbitrarily applicable relational responding proficiency. As detailed in Dixon et al.'s 2017 and 2022 reviews, the PEAK programme appears successful in the significant improvement of multiple language deficits in people with disabilities, proficiency to engage in deictic framing, rule-following, derived categorical responding, analogical reasoning, and cross-modal transfer across different stimulus modalities. While this suggests the value that relational responding training interventions provide towards the successful improvement of various beneficial skills relevant to cognitive performance, a different, RFT-based intervention known as SMART provides further support for the argument that derived relational responding interventions can directly enhance performance on IQ tests and educational attainment. Indeed, due to the SMART intervention's more direct focus on intelligence and its direct relevance to the measure which is the focus of the current thesis, the following section will review the SMART intervention in closer detail.

1.6.4 SMART

1.6.4.1 What is SMART?

While SMART takes the same theoretical approach as PEAK (i.e., training relational responding to improve intellectual abilities), it differs in the sense that SMART assumes an existing baseline relational responding fluency for users of the system, and also focuses on different core relations. As established in section 1.6.1, it appears that previous interventions' failure to significantly enhance IQ may be a result of the attempts' lack of aligning the intervention with existing theories of intelligence. Therefore, it seems as though these attempts almost adopt a trial-and-error approach in which concepts vaguely linked to intelligence theory through correlational analyses are used as the basis for training interventions. However, this approach fails to explain the link established by correlational analyses and disregards the common warning that correlation does not equal causation. To a behaviour analyst, the failure of such an approach is unsurprising. The lack of an established functional understanding makes it practically ironic to expect meaningful increases in intellectual performance. Cassidy et al. (2010) argued that RFT offers the required theoretical background to develop a successful intervention that generates significant and meaningful increases in IQ. As derived relational responding is conceived of as a generalised operant class, it is, by definition, subject to modification through environmental interventions. As such, Cassidy et al. (2016) invented the SMART program. With its foundations in RFT, the program, in effect, is based on the idea that behaviours considered "intelligent" are simply the application of derived relational responding or relational framing to given tasks (Cassidy et al., 2010). The intervention draws on the idea that multiple exemplar training in derived relational responding should facilitate the rapid expansion of relational responding repertoires generalised across a wide array of relata. Consequently, this should result in vocabulary expansion and improved creative problem-solving in novel contexts that align in regard to the problem-solving strategy but not in the formal nature of the tasks with the training contexts.

SMART utilises a gamified online format designed to train relational operants in the form of multiple exemplar training. Based on RFT's assumptions (laid out in section 1.5), this would result in increased intellectual performance, measurable through increased IQ. Using a multiple exemplar training approach aims to allow for the generalisation of the trained skill, so that participants gain far-transfer effects allowing for a wide application, instead of solely training specific skills relevant only to a limited number of tasks.

The program consists of 55 progressively more complex multitask training levels with two blocks, each of which trains relations within the frames of coordination, opposition and comparison. Provided with relational cues in the premises, which allow the deduction of the implicit relation, participants are required to solve a variety of items enquiring about the relationship between two stimuli in the form of nonsense words. All presented relata are nonsense words guaranteed not to be repeated throughout the whole training programme for any individual participant. To aid pronounceability and, consequently, fluency in reading the task items, nonsense words are presented in the forms consonant (C) – vowel (V) – C, CVVC, CVCC, or CCVC. Relational cues are English words. One such example is "CUG is opposite to BEH, BEH is opposite to VEK. Is CUG opposite to VEG?". Participants are given a 30 second time limit per trial.

All 55 stages are individually trained through multiple exemplar training, allowing for a potentially infinite number of trials and duration, based on the participants' performance (i.e., as outlined below, trial numbers were determined based on correct responses and therefore varied across participants). The first 29 stages focus on the frame of coordination and opposition, and the second block focuses on comparative relations. Corrective feedback is utilised to maximise learning outcomes. Progression to the next stage requires the participant to consecutively pass 16 items by giving correct responses. If a trial is not answered correctly, it will be assessed again in the same form but with different relata. Once the participant has successfully completed the training stage, they will enter the testing stage, in which they must consecutively pass all 16 novel tasks, matched for the relational complexity of the preceding training stage but with the exclusion of corrective feedback. If this is passed, they move on to the next training stage.

Task complexity is increased gradually and controlled for by increasing the number of presented premises from one to four, changing the order of premises from sequential to random, and varying the directionality of the relational question as well as the number of specified relation types across premises. Directionality of the relational question refers to whether the relational question follows the linear direction of presented premises (e.g., if the premise is “A is more than B, B is more than C, and C is more than D”, then a question congruent in direction would ask whether “A [is] more than D”. In contrast, a question incongruent in direction would ask whether “D [is] more than A”. Incongruency in directionality requires a more advanced skill repertoire). Additionally, the congruency of relational cues presented in the relational premises and questions (i.e., whether the relational term was enquired about in the question presented in any of the premises) varies across trials. Moreover, the nodal distance between relata increases with subsequent trials. To illustrate the varying nodal distances, given the premises “A is the same as B, B is the same as C, and C is the same as D”, a number of relations with varying nodal distances can be derived. A zero-node derived relation probe would enquire about the relation between A and B, whereas a one-node derived relation probe would focus on A and C, for instance. A two-node derived relations probe would enquire about the relation between A and D. Therefore, the number of relata laying between the to-be-derived relata in the premises determines the number of nodes. According to RFT, there is a positive correlation between the number of nodes and the complexity of the relational task.

Aligning with behaviour-analytic techniques, the intervention makes use of reinforcement, whereby every correct response is positively reinforced (through verbal and auditory feedback and points on a fixed-ratio schedule of FR1). To further enhance engagement, the intervention awards level badges and utilises traditional gamification methods like time-constrained responding, clearly defined performance targets throughout the whole intervention and finally, moderate mental "challenge" (Linehan et al., 2015). The success of the SMART intervention complements the argument that relational responding repertoires underlie intellectual performance. While interventions to increase intelligence are interesting themselves, it is important for the reader to remember that they are not the focus of the current thesis. Instead, the current thesis is concerned with the measurement of relational responding skills. One such measure, the RAI is understood to have excellent face validity for the relational skills on SMART. That is because it is, in fact, composed of sample tasks from the training protocol. It is important here for the reader to be aware that any correlations between RAI performance and SMART training should not be automatically interpreted as evidence of far-transfer effects from the SMART intervention to IQ, as measured by traditional intelligence tests. Since the RAI assesses the skills developed during the SMART protocol, this constitutes a training-to-test situation, making it impossible to demonstrate far-transfer effects between SMART training and RAI performance. However, exploring how increases in RAI scores might translate into increases in IQ scores would be a valuable direction for future studies. Beck et al. (2023) further suggest investigating the predictive power of RAI scores regarding improvements in standardised IQ scores, IQ subtests, and other measures of intelligence. Indeed, the RAI is the behavioural standard measure of relational responding fluency, regardless of its correlations with standard intelligence tests and whether or not any interventions lead to changes on those intelligence tests. The primary aim of the RAI is to function as an assessment of relational responding fluency. Nonetheless, a brief review of additional SMART intervention studies is warranted, as it highlights the support for the assertion that conceiving intelligence as an acquired skill rather than an inherent trait is more appropriate.

1.6.4.2 Evidence for the SMART Interventions

As evinced above, studies on relational responding interventions are successful in their aim to increase intellectual performance. While these provide the participants with a heightened proficiency in skills critical and applicable in various domains of daily life, they are somewhat limited in their ability to offer extensive insight into the close relationship between relational responding and intellectual performance. Developing and investigating the assessment of such skills can help resolve this issue and consequently not only further enhance the understanding of significant correlation between intellectual performance and relational skills, but further aid in the development of actionable interventions designed to enhance these skills. The following section will outline the importance of assessing relational skills, explain how doing so can improve upon the shortcomings of traditional IQ tests, and ultimately introduce one such measure, the RAI.

A number of studies have found that the SMART intervention successfully increased intelligence. In one of the earliest of such studies, Cassidy et al.'s (2011) first study (N = 8, 6 female, Age = 8-12) found significant increases in FSIQ as assessed through 12 of the 13 subtests of the Wechsler Intelligence Scale for Children (WISC-III^{UK}; Wechsler, 1992) for the experimental but not the control group. Similar results were found in their second study (*N* = 8, five female, Age = 11-12), which utilised a revised training protocol and observed significant FSIQ rises post-intervention, reaching at least one standard deviation in seven of the eight cases. The development and administration of a RAI confirmed these increases in relational skill repertoires to result from the employed multiple exemplar training intervention. Aiming to replicate Cassidy et al.'s (2011) findings, Cassidy et al.'s (2016) first study (*N* = 15, nine female, Age = 10-12) found significant increases in FSIQ, as assessed by the Wechsler Intelligence Scale for Children IV^{UK} (Wechsler, 2004) with a mean IQ of 97 points pre intervention and 120 points post intervention. More specifically, without exception, the increase in FSIQ was about one standard deviation, with the lowest increase reaching 14 IQ points (i.e., 1 point below the standard deviation), and the highest three increases reaching over 28 IQ points), consequently resulting in the participants' changes in percentile rank and IQ category (from average to high average). Study 2 (*N* = 30, 15 female, Age = 15-17) of Cassidy et al. (2016) used the Differential Aptitude Test (5th edition; Bennett et al., 1990) as a group-administered test of cognitive ability, which further allowed the computation of a composite score for educational aptitude and was administered at baseline and at follow-up. Increases in verbal and numerical reasoning were found to be significant in the majority of participants, consequently suggesting the SMART intervention to further be efficacious in the enhancement of educational aptitude.

Further success was found in studies with active control groups and possible far-transfer effects were also observed. Hayes and Stewart (2016; *N* = 28, 13 female, Age = 10-11), for instance, used an active control group which conducted Scratch training (a computer coding training program aiming to teach children how to program their own computer game). Results showed a significant improvement of intellectual ability and academic attainment in the experimental group only, consequently resulting in significant improvement differences between the two conditions favouring the experimental group in overall intellectual performance, digit span and letter/number sequencing as well as academic attainment (i.e., spelling, reading, and numerical operations). Moreover, significant increases were found in studies that utilised a blind testing methodology. Using the WAIS-III (Wechsler, 1997) to assess IQ, Colbert et al. (2015; *N* = 34, 19 female, Age = 18-44) found significant mean FSIQ, Verbal, and Performance IQ increases in the experimental group with FSIQ increasing by 18.8 points, thus exceeding one full SD and consequently increasing the mean group IQ from the average to the high average classification band, with no significant increases of FSIQ in the control group. However, it must be noted that significant between-group differences for increased IQ scores were only found for full-scale and verbal IQ. Furthermore, neither baseline FSIQ, nor the number of completed training levels significantly predicted the IQ increases at follow-up. Possible far-transfer effects were found in McLoughlin et al. (2021; *N* = 49, 23 female, Age = 6-10), who found that, after observing a significant increase in nonverbal IQ for the experimental group only, contrary to the control group, the experimental group's follow-up nonverbal IQ was a better predictor of reading test scores than baseline nonverbal IQ, suggesting a possible far transfer effect towards educational attainment (spelling, reading, and numerical operations). Moreover, a 2022 study by McLoughlin et al. (*N* = 125, Age = 12-15) tested the efficacy of SMART training in improving nonverbal IQ and processing speed. The results showed that, in contrast to the active control group's nonverbal IQ, the experimental group's nonverbal IQ increased ($M_{Increase\ in\ Nonverbal\ IQ} = 5.98$) following the SMART intervention, suggesting the intervention was successful. The study further found a strong, negative correlation between baseline ability and nonverbal IQ improvements but no effect of baseline nonverbal IQ in the control group.

While the research on the SMART intervention is growing and seems promising, it would be scientifically remiss to disregard some of the methodological shortcomings and simply present the SMART intervention as the magic solution to reliably increase IQ regardless of external conditions. While the above-presented studies found significant differences in FSIQ increases between the experimental and control groups, favouring the former, three studies (Colbert, 2015; Roche et al., 2023; McLoughlin et al., 2018) did not find any significant difference in FSIQ increases. Despite the non-significant differences in FSIQ rises, Colbert (2015) found a significant difference in the increase of the Verbal Comprehension score, favouring the experimental group. Additional methodological shortcomings are insufficient research on far-transfer effects (Beck et al., 2023), as well as low completion rates and sample sizes (Cummins, 2023).

While some might question the intervention's validity due to the study's methodological issues, a very recent study by Stricker et al. (2024; $N = 59$, 21 female, Age = US 10th and 11th grade) aimed to improve upon some of those methodological issues by employing a single-blind design with a matched control group. Assessing the intervention's positive effects beyond IQ, the study investigated changes in educational performance, specifically reading and mathematics, as measured by the Pennsylvania State Department of Education classroom diagnostic tool (CDT). Groups were matched on baseline CDT scores, and, where possible, ethnicity, gender, and socioeconomic factors. While completion rates were an area of concern in previous SMART studies, the current study reached a completion rate of 100% for all participants. Results found significant differences in mathematics ($M_{\text{Treatment}} = 9.24$; $M_{\text{Control}} = 2.30$) and reading ($M_{\text{Treatment}} =$

11.24 ; M_{Control} = 1.77) improvements. This effect remained when controlling for baseline scores. Such a finding highlights the far-transfer effects and real-world applications of the SMART intervention, as well as the intervention's success, whilst methodological concerns are being decreased. While one should still be careful when interpreting such findings, due to the statistical lack of power, the findings highlight the irrelevance of selection bias, nonetheless.

In conclusion, the SMART program appears to be successful in its aim to raise IQ. However, its success might be within limits, as various boundary conditions have been found, which require further investigation and attention in the future. At the current moment, it is unclear which of these boundary conditions are actual limitations of the intervention, and which limitations are simply due to methodological differences. One of the major questions is the intervention's enhanced efficacy and generalisability if it was expanded with further relational frames. Nonetheless, given the extent of its success with regard to the number of increased IQ points and the implemented real-world effects, this is undoubtedly a worthwhile endeavour.

Despite the support for the success of SMART studies, it is important to acknowledge some shortcomings in the existing literature. In their systematic review and meta-analysis, May et al. (2022) summarise several methodological limitations in the SMART studies they investigated. One major concern is that many SMART studies lack an (active) control group. The inclusion of both an active and a passive control group is desirable (Gobet and Campitelli, 2006), as it enhances the certainty that the findings are attributable to the intervention rather than to known or unknown confounding variables, such as placebo effects (Colbert et al., 2018; McLoughlin et al., 2021). In the same vein, May et al. (2022) highlight that double blind procedures would reduce the risk of expectancy effects and intention-to-treat analyses would help identify potentially relevant differences between participants who completed the training and those who did not. Longitudinal studies are also necessary to analyse the long-term effects of the SMART intervention and studies that assess educational outcomes as a result of the training, such as Stricker et al.'s (2024), allow for conclusions about far-transfer effects.

Adequately powered studies would further enhance the quality of the SMART literature by providing more accurate effect size estimates (Button et al., 2013; Gelman & Carlin, 2014; Ioannidis, 2008). Finally, May et al. (2022) point out that many SMART studies are influenced by vested interests or theoretical biases, and they call for large-scale independent clinical trials "with no vested interest whatsoever" (p. 1256). Despite these limitations, which are crucial to acknowledge, May et al. (2022) conclude that "nonetheless, SMART is a promising, theoretically plausible, and empirically grounded approach to increasing general cognitive ability" (Dymond & Roche, 2013; McLoughlin et al. 2020, p. 1258). Therefore, this thesis will consider the SMART literature in this light, while reminding readers to remain aware of these shortcomings going forward.

The success in increasing IQ test scores by training relational responding highly supports RFT's assertion that relational responding underlies complex cognition and that it is more appropriate to view intelligence as a malleable skill rather than an inherent trait. Studying skills requires a reliable measure of such skills in order to allow for reliable conclusions. Such a measure is provided through the RAI and is based on principles proposed by RFT.

1.7 Measuring Relational Skills: The Relational Abilities Index

Studying the importance of relational responding repertoires calls for the ability to measure such repertoires. Not only does the measurement of relational responding proficiency allow researchers to quantify the repertoire fluency with a usable quantum, but it further provides a benchmark for assessing the effects of any intervention to improve the fluency of that repertoire.

While the outcomes that are provided by any such assessment is not to be mistaken as the index of an extended construct, it nevertheless can function as such in a pragmatic manner in day-to-day research. Or, the purpose of such an index is not to provide normalised scores for the purpose of population-level estimates but merely to provide a systematic way to assess behavioural fluency that is relevant to intellectual development.

While not conceived as a psychometric test in the usual sense, a RAI has the advantage that due to its functional approach fluency at a particular level of relational complexity can be assessed with a variety of alternate forms of any given assessment, so long as they are identical in terms of the generalised offerings under assessment. This means that in effect, a well-produced RAI can circumvent practice effects and should have excellent test-retest reliability. To more fully understand the advantages of the RAI over traditional IQ tests, a brief overview of the development of the RAI is warranted.

1.7.1 First Relational Abilities Index

As mentioned in section 1.6.4, the RAI was initially developed by Cassidy et al. (2010) and later expanded upon in Cassidy et al. (2016). The initial RAI was preliminary, consisting of 55 relational tasks with the primary function of assessing increases in relational responding proficiency within the frames of opposition (29 trials, e.g., “HEH is opposite to WAK, WAK is the same as XIX. Is HEH the same as XIX?”) and quantity (26 trials, e.g., “NON is less than NIG, PIL is less than NON. Is NIG less than NON?”) by comparing RAI scores pre- and post-intervention. The relational tasks of the RAI consist of syllogistic puzzles mirroring those in the SMART program, with the exclusion of corrective feedback in order to facilitate the appropriate assessment of the skill repertoire independent of external input. Colbert et al. (2017) carried out an initial in-depth analysis of the RAI’s validity and reliability. As outlined in section 1.5.3, significant, moderate correlations between the RAI and various measures of cognitive function (NART, Nelson, 1982; RAVLT, Rey, 1958; English version: Taylor, 1959; TMT, Lezak, 1995; WAIS-III, Wechsler, 1997) were found. One should be careful not to interpret these significant correlations as proof that the RAI is a sufficient measure of intellectual ability as traditionally conceptualised, especially since out of these, only the WAIS-III is an accepted gold standard measure, while the NART is the only recognised proxy measure. The Cognitive Failures Questionnaire (Broadbent et al., 1982) did not significantly correlate with RAI scores. Based on this convergent and divergent validity, the RAI was concluded to be a sufficient measure of intellectual ability (Colbert et al., 2017). Indeed, the effect size of the RAI's correlation with FSIQ approached that found between FSIQ and other established proxy measures of intelligence, such as that between the WAIS-III and RSPM (0.64 in Wechsler, 1997), the Wechsler Individual Achievement Test (.53-.81 in Silva, 2008), the Woodcock-Johnson Test of Cognitive Ability (.72 in Cheramie et al., 2012 and 0.82 in Metz, 2005), the General Ability Measure for Adults (.8 in Martin et al.,2000 and .75 in Naglieri & Bardos, 1997), WASI Verbal IQ (.75 in Axelrod, 2002), and the Oklahoma Premorbid Intelligence Estimate (.69 in Spinks et al., 2009). Moreover, analyses showed a high test-retest reliability of .809.

Colbert et al.'s (2017) findings further highlight the crucial role that relational abilities have on intellectual performance. Similarly, the moderate-to-strong significant correlations between RAI scores and each of the WAIS' Verbal subtests indicate the importance of relational responding, as assessed by the RAI, to linguistic performance (Colbert et al., 2017). This further supports RFT's proposal that relational responding underlies language processes. Based on Colbert et al.'s (2017) study findings, it can be cautiously concluded that relational skills appear to most strongly underlie intelligence in the verbal domain. The finding that RAI performance only significantly correlated with three out of six of the Performance IQ subtests might indicate that the RAI more closely resembles verbal tasks rather than performance tasks in terms of intellectual skills repertoires, argue Colbert et al. (2017). Furthermore, and perhaps more importantly, this highlights the inappropriateness of using the RAI as a proxy measure or estimate of IQ (Colbert et al., 2017). Nonetheless, they found that out of the 13 WAIS subtests assessed in their study, those with the highest correlation with the RAI were the same ones previously found to have the highest *g* loadings (Gignac, 2006). Importantly, this does not bear any weight to the behaviour analyst, but regardless, it is a noteworthy finding.

Further, it is important to note the studies' limitations, calling for a cautious interpretation of the findings. One of the most significant limitations is the limited sample size, potentially allowing for the interpretation that it is high intelligence which causes good performance on the RAI, instead of relational reasoning proficiency that causes high IQ test scores (Colbert et al., 2017). Indeed, it might not even be a direct causal relationship, but a relationship affected by an unknown confounder variable. This, argue the authors, is not necessarily negated by the finding that relational skills interventions successfully raise IQ scores. McLoughlin et al. (2020), however, argue that such high correlations between traditional IQ tests and the RAI further suggest the synonymity of relational reasoning and intellectual ability, where a higher proficiency in relational reasoning translates to higher levels of cognitive ability. This view aligns with the behaviour analyst's conceptualisation of intelligence.

Methodologically, the RAI not only has the advantage of precluding practice-effects due to its function as a measure of a generalised operant and the ease with which this can be implicated (i.e., simply programming the RAI stimuli to vary across trials), but its automation in both administration and scoring prevent any subjective interference, which can skew results in traditional IQ tests.

1.7.2 Development of the RAI+

Notably, the RAI only assesses relational responding proficiency on two relational frames (distinction and comparison). Colbert et al. (2017) noted that this limitation might serve as the explanation for the measure's limited correlation with certain WAIS subtests. Consequently, they proposed that future versions of the RAI should assess a wider array of relational frames in order to cover the entire repertoire assessed by traditional IQ tests. This could also benefit the reduction of the observed ceiling effect (Colbert et al., 2017; Gore et al., 2010). Indeed, Colbert et al. (2020) extended the RAI by adding three modules (namely difference (e.g., "SAJ is different to LIR, LIR is the same as VUS. Is VUS the same as SAJ?"), temporal (e.g., LOF is after FEH, WUC is after LOF, LON is after WUC. Is FEH before LON?"), and analogy (e.g., QUD is less than KON, JOL is more than JIT. Is KON to QUD the same as JIT to JOL?"), consequently providing insight into proficiency in a broader array of relational frames. This revised measure, the RAI+, was not developed as a superior alternative to traditional IQ tests but instead as a functional assessment of intellectual performance. Through the analyses of its correlation with tasks traditionally perceived to indicate intellectual ability and scholastic performance and the measure's individual subtest's correlation with those traditional tests, the study aimed to validate the RAI+ as a functional proxy measure of intelligence and scholastic ability. Additionally, it aimed to investigate the individual frames' correlation with the various skills usually assessed in traditional intelligence tests.

The RAI+ is constituted of 67 logical syllogistic relational puzzles, following the same format as that of the RAI, across five modules (15 opposition trials, 14 distinction trials, 13 comparison trials, 13 temporal trials, and 12 analogy trials). Again, task complexity is controlled for through the same mechanisms as in the SMART program. Apart from the analogy module, each module's first trial is restricted to a single premise followed by a relational question to assess the participant's proficiency in deriving mutually entailed relations by either changing the premise's direction or inverting the relational frame in the relational question. Modules progressed to 10 trials consisting of two premises and three relata in which every derived relation allowed for within this network was enquired.

The study found significant correlations between the RAI+ and FSIQ (.54), Verbal IQ (.42), and Performance IQ (.48), as assessed by the WASI IQ. Furthermore, it found significant correlations between performance on the RAI+ and the WASI IQ's subtests, namely Vocabulary (.37), Similarities (.37), Block Design (.42), and Matrix Reasoning (.48). Additionally, RAI+ performance significantly correlated with WAIS-III Arithmetic scores (.43) but not WAIT-T scores, suggesting relational abilities to be critical to numeracy, but failing to find a correlation with scholastic ability, as assessed by the WAIT-T. These findings suggest the importance of relational abilities to intelligence test performance and numeracy, further highlighting their relevance to higher cognition.

1.7.3 Development of the Elaborated RAI

The latest version of the RAI, the Elaborated RAI, was first introduced by Cummins et al. (2023). Methodologically, the Elaborated RAI mirrors the RAI+ with the important difference being the extension of the number of frames that are assessed. The Elaborated RAI extends upon the RAI+ by adding deictic (e.g., "TIY is here, QIH is there. If here is there and there is here, is TIY there?"), mathematical (e.g., "YAV + MIY is the same as TEM + GIX. YAV is less than TEM, is GIX less than MIY?"), and containment (QED contains VOP, KIB is within VOP, KIB contains CUG. Is CUG within QED?") relations. Cummins (2023) assessed the Elaborated RAI's psychometric properties regarding its efficacy as a measure of individual-level abilities. As highlighted by Cummins (2023), if the requirement of sufficient inter-participant discriminability is not met, a given participant's score will be a poor indicator of their abilities, as no meaningful distinguishments to other participants can be made. In other words, participants' performances, according to their scores, are so similar, that it is basically impossible to tell participants apart based on their performance, thus rendering the measure redundant, as it fails to provide any valuable information. To analyse the degree to which the elaborated RAI allows inter-participant discriminability, Cummins (2023) computed a "discriminability score" for each participant at each number of trials for each subscale, which was modelled using a linear regression model (discriminability score $\sim 1 + n_{trials} + subscale + n_{trials}: subscale$). It was found that the individual-level precision of the RAI's subscales is lower than desired. On average, participants were only discriminable from about a third of other participants (with the lowest discriminability even reaching only 7% of other participants). Cummins (2023) highlights that this finding can critically affect future research using the RAI, but that the degree to which one would like to be able to discriminate between participants depends on the specific purposes of the studies. The level of detail provided at higher levels of performance was found to be quite limited, showing that the RAI is more efficient in discriminating between low-ability participants, in effect making the test vulnerable to ceiling effects.

Cummins (2023) also investigated the RAI subscales' split-half reliabilities. While the difference, temporal, and containment subscales showed rather good split-half reliabilities, the analogy and opposition subscales performed particularly poorly. This highlights the need for specific subscales to be improved. However, it should be noted that the overall split-half reliability for the test as a whole was respectable. A coefficient of .90 for the overall RAI was found in Cummins et al. (2023) whereas Cummins (2023) found the individual subscales to fall between .35 and .81, though it should be noted that split-half reliabilities are naturally weaker as the number of items decrease. Finally, Cummins (2023) investigated the RAI's adherence to its supposed Guttman structure, which is highly coherent with the assumptions laid out in RFT (Hayes et al., 2001). Interestingly, it was found that the data showed a violation of the Guttman structure. More specifically, rather than showing a steady increase in trial difficulty within each trial, Cummins (2023) found that initial trials within a subscale appeared to be rather difficult in yielding fewer correct responses, before difficulty drastically decreased for the middle trials. Some increases in difficulty towards the final few trials were found, however not in all cases. This pattern was observed for all subscales except the analogy and mathematical subscales.

While those findings definitely highlight the RAI's need for improvement, especially where it is used as a measure of improvements in relational fluency across time or across individuals, it is important to note that Cummins' (2023) overall conclusion was that the RAI is "broadly promising" and that the required inter-participant discriminability will vary based on the needs of the studies utilising the RAI. Compared to the previous versions, the Elaborated RAI has shown substantial improvements in scope. Now, however, while continuing to question and consider the test's psychometric properties going forward, we are in a position to assess the test in its more elaborated form as one might the advent of any assessment for use in the educational or psychometric context. That is, it seems timely now to try to understand some of the contextual features that influence responding on the test. We are also in a position to revisit the issue of the convergence of scores on this type of assessment with measures of interest in the educational and psychometric context as has been conducted on previous iterations of the test.

1.8 The Current Thesis

The elaborated RAI has shown great promise as an improvement upon previous versions of the test in terms of its scope and internal psychometric properties, therefore making it a more reliable and better-understood measure of the repertoire of interest. Nevertheless, several questions remain regarding its utility, many of which relate to its interpretability in relation to real-world scholastic performances or in terms of how performance on a RAI is indicative of performance on other tests of cognitive ability. Such questions will be undertaken in the current thesis in an entirely exploratory manner. While previous versions of the RAI have been examined in relation to standardised tests of several types, this has yet to occur for the elaborated RAI. In addition, it is important to start considering sociodemographic variables that are often considered to function as confounds for standard intelligence tests, such as parental education and income (see section 1.3). More specifically, the current thesis investigates own and parental education and income, as these have been found to correlate with traditional measures of intelligence, but more importantly, it is believed that these variables provide insight into an individual's exposure to educational experiences, either through direct learning (own and parental education) or the financial means to avail of e.g., tutoring. This exposure to educational opportunities, in turn, is suggested to facilitate the acquisition of relational responding proficiency. That is, given that RFT proposes relational responding to be an acquired skill, which, more specifically, is acquired through educational opportunity, it would be reasonable to assume that exposure to educational opportunities positively correlates with relational responding proficiency. The reasoning behind the inclusion of one's own education level as a covariate of relational responding fluency is quite apparent; individuals exposed to more years of formal education are more likely to acquire and improve in the relevant skills. Parental education was included here, not as a proxy index of genetic influence, like it has been in other studies (e.g., Sheshagiri et al., 2016), but instead as an index of the quality of the educational interactions that individuals have experienced in their upbringing at home. This was primarily motivated by the landmark findings of Hart and Risley (1995), which pointed to substantial differences in the educational quality of parent-child interactions based on the parent's socioeconomic background (this will be discussed in further detail below). In effect, parental income was assessed here as an index of the quality of education the individual would likely have received at home and in school. Previous studies found that these two variables are positively related (e.g., Hotz et al., 2023). Moreover, parents with a higher income are likely to be in a better position to facilitate educational extracurricular activities for their children and less likely to subject their children to neglecting circumstances (e.g., malnourishment), as has been discussed in section 1.3.2. Finally, own income was assessed as an index of which educational opportunities the individual might be able to access for themselves in early adulthood.

The current study aims to examine several such factors. This work is important because from a behavioural point of view, sociodemographic factors fractures are not confounds that interfere with the accurate measurement of a construct, as might be assumed within traditional test theory. Instead, they are interpreted as contextual factors that help to determine that score. In other words, it would be surprising if a relational repertoire assumed to emerge as a generalised operant within the social community was not itself sensitive to the dynamics within that social community (e.g., level of interaction with that community, the quality of educational instruction received, access to high-quality educational opportunities that require financial resources, etc.). For this reason, it might be expected that various forms of social opportunities, such as the level of education of caregivers and financial status, could impact on the relational repertoire fluency of an individual. These explorations will occur here without prejudice as to the moral rights or wrongs of such social influence on RAI performance.

In terms of the convergence analysis, the elaborated RAI will be examined here in terms of how its scores correlate with those of RSPM. The RSPM is a widely used proxy measure of intelligence and arguably the best correlate of *g* (Deary and Smith, 2004, Eysenck, 1998, Jensen, 1998, Llabre, 1984, Neisser, 1998, Thorndike, 1986, Vernon, 1947). While other tests could be chosen for this convergence analysis, they present the challenge of requiring considerable time to administer. In any case, the current research is just the beginnings of such a series of explorations attempting to understand the relationship between proficient relational responding repertoires and standard measures of fluid or crystallised intelligence. This exploration in no way suggests that the degree of convergence or divergence between the two test results reflects on either measure, as neither will be standardised against the other. Instead, the investigation is entirely exploratory in the sense that it will provide information about where different repertoires overlap and where they diverge, purely for the purpose of scholarly inquiry. It is important to understand, therefore, that the attempt to quantify the RAI's convergence with the RSPM is not in any way intended to establish the validity or invalidity of the RAI. Put simply, the degree of divergence or convergence it displays with other widely used measures may be as likely to indicate a lack of validity on the part of either measure. In addition to addressing this question, further psychometric integrity analyses will be undertaken to complement the work of Cummins et al. (2023). Specifically, inter-subscale correlations will be analysed for the purpose of replicating the work of Cummins et al. (2023). This analysis will provide important insights into the function of relationships between various types of relational responding thereby allowing a better understanding of which are more pivotal amongst the entire relation responding repertoire and also which correlate most strongly with external measures of fluid intelligence (*g*). However, the current thesis will report on an additional in-depth secondary analysis of the data for the purposes of conducting the first factor analysis of the RAI. McLoughlin (2022) explained that, as a consequence of offering valuable insight into the existing factor structure of the RAI, a factor analysis would allow additional insight into how relational responding proficiency as an underlying skill of intelligence performs compared to single underlying factor that is proposed by *g* theory. Additionally, a factor analysis can offer valuable information which will help guide the future development of the RAI. Overall, the current thesis will add to the literature by conducting a deep and widely ranging assessment of the RAI's psychometric properties and thus contribute to its development.

Chapter 2

Assessing the Psychometric Properties of the Elaborated Relational Abilities Index

2.1 Research Questions

1.1.2 RQ1. What is the convergence between the RAI and RSPM?

This research question aimed to analyse the convergence between RAI and RSPM performance. The question will be addressed using Spearman's rho correlation analyses to allow insight into the degree of correlation between relational responding as assessed by the RAI and RSPM performance. A strong, significant correlation was expected due to both measures arguably assessing the same skills (see section 1.5.3), but not necessarily conceptually desired, as explained in section 1.8.

2.1.2 RQ2. How do the variables correlate with each other?

To get a general picture of how parental income, own income, parental education, own education, age, as well as performance on the total RAI, RAI-M performance, individual RAI subscales performances and RSPM performance intercorrelate, a Spearman's rho correlation matrix will be computed. This will offer insight into general relationships and highlight any surprising correlations which might not have been considered before. As this analysis was purely exploratory, no predictions about the underlying correlations were made, though it was expected that RAI and RSPM scores would intercorrelate, as well as significantly correlate with the individual SES variables.

2.1.3 RQ3. How does socioeconomic status predict performance on the RAI and RSPM?

This research question focuses on the impact that socioeconomic status has on RAI, RAI-M, and RSPM performance (all assessed individually). This impact will be analysed using multiple regression analyses, which modelled the four SES variables as predictor variables and RAI, RAI-M, and RSPM performance as criterion variables. While multiple regression models generally assume a causal relationship, it is important to remember that they only provide insight into correlational impacts and thus, results fail to offer sufficient insight into any causal inferences. Nonetheless, correlational inferences offer insight into possible differences in performance based on differences in socioeconomic backgrounds. It was expected that SES would significantly predict performance on all four measures.

2.2 Methodology

2.2.1 Participants

An a-priori power analysis was conducted using G-POWER (Faul et al., 2007, 2009) and indicated that to detect an effect size of .15 at an alpha level of .05 with 80% power in a multiple regression analysis, a sample size of 127 would be required. For the Spearman's rho correlation coefficients the power analysis was conducted using software for estimating power of a Pearson's correlation, as the two are computationally identical (Faul et al., 2008). The power analysis indicated that to detect a medium effect size of .3 for a two-tailed test at an alpha level of .05 with 80% power, a sample size of 82 would be required. A total of 198 participants took part in the study. Inclusion criteria were being between 20 and 24 years old and having grown up in Ireland or the UK. Following the application of necessary data exclusion methods (see Results), data from 156 participants were retained for analysis (82 female [52.6%]; 72 male [46.2%)], two non-binary [1.3%], M_{Age} = 21.29 [SD = 1.03]). 135 (86.54%) of these participants were recruited in the form of self-selecting sampling via Prolific Academic (a participant recruitment service), whereas the remaining 21 participants (13.46%) were undergraduate students recruited via an internet posting campaign across MU on Moodle. Prolific participants were renumerated at €10.20 per hour, while student participants were undergraduate psychology students in their second year of study and received course credit for participating in the study. Table 1 below shows the distribution of own and parental education levels for all participants.

Table 1

Distribution of Participants' Own and Parental Education

Education Level	Parental Education		Own Education	
	N	%	N	%
No formal education	4	2.6	0	0
Some secondary school	12	7.7	1	.6
Junior Certificate /GCSE Examinations	22	14.1	7	4.5
Leaving Cert / A Levels	29	18.6	77	49.4
Certificate of Higher Education / Diploma	13	8.3	9	5.8
College / University Degree (Undergraduate)	47	30.1	52	33.3
Master's Degree (Postgraduate)	27	17.3	9	5.8
Ph.D.	2	1.3	1	.6

As can be seen in the table above, the most common parental education is an undergraduate university degree followed by a leaving certificate, which is closely followed by a Master's degree. Own education shows a similar pattern, though the leaving certificate is the most common education, followed by an undergraduate degree. This can very likely be attributed to the young age of participants. PhDs and education below the leaving certificate were relatively uncommon in general, though more common in parental than own education.

Figure 2 and 3 show the distribution of parental and own income respectively. The parental income distribution ranges across the whole range, though the majority collect around the 40th percentile income range. Comparatively, the distribution of own income was much more focused around the lower percentiles, though again, the spread covered the whole range. Again, this can likely be attributed to the young age of participants.

Figure 2

Distribution of Parental Income

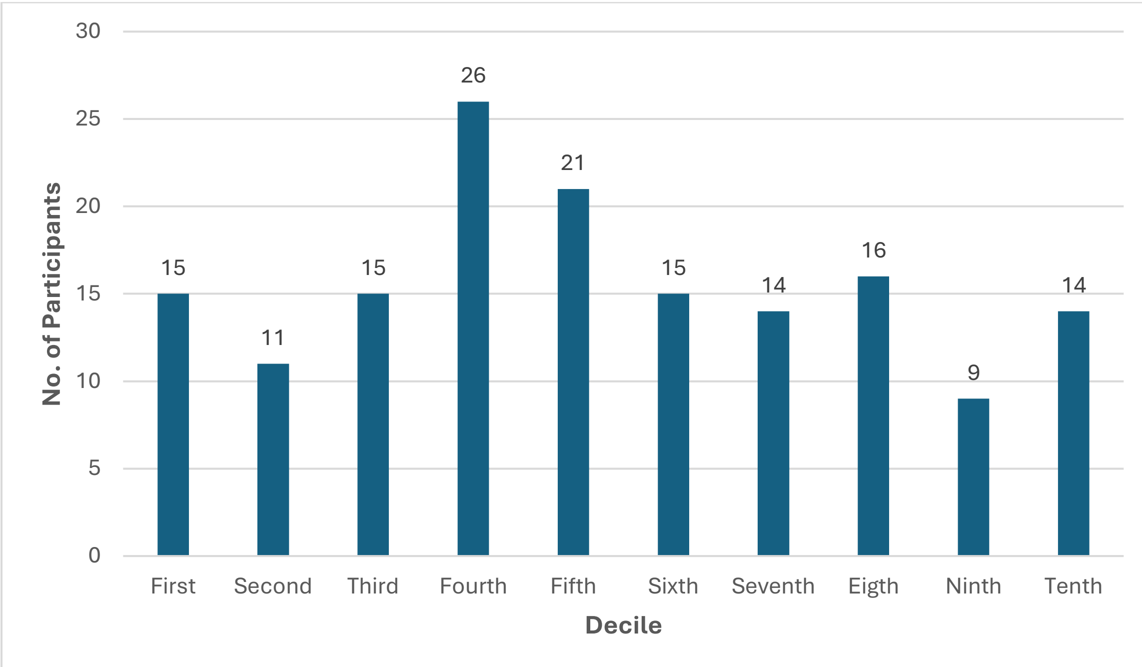
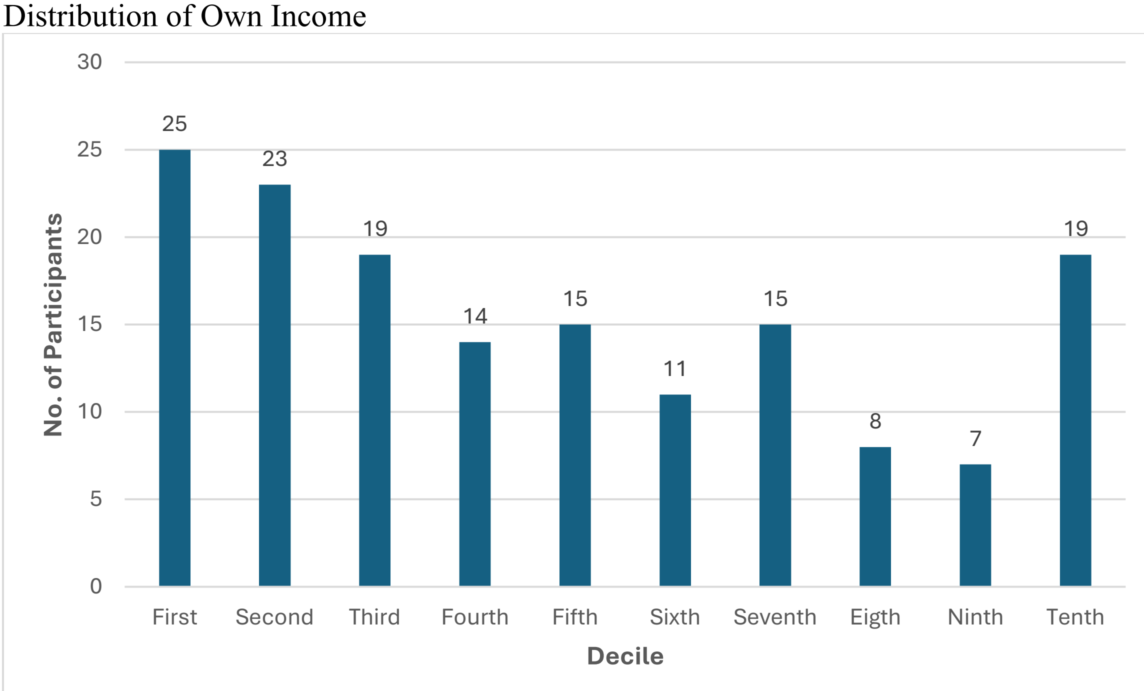


Figure 3

Distribution of Own Income



2.2.2 Settings and Materials

The study was conducted entirely online using lab.js software, which allows computer-controlled experiments to be delivered remotely via a web browser. Except for 25 student participants from Maynooth University in Ireland, the study was conducted on participants' own devices. Participants ($N = 4$) who did not complete the study on a laptop or a computer were excluded from analysis. Student participants from MU completed the experiment in the MU Psychology Departments' research cubicles. The research cubicles (approx. 1.8m x 1.8m) contain two desks facing the wall, two chairs, and a computer. The researcher left the research cubicle after briefing the participant on the study and ensured a quiet environment to allow the participants to focus on the tasks. Apart from location, and being chaperoned to the research cubicle, all participants were exposed to the same experimental conditions. Responses were exclusively made up of keyboard button presses or mouse clicks and recorded as either number of correct responses or, in the case of demographic questions, as Likert-scale answers. Moreover, the software programme recorded response times for every single response from the beginning to the end of the experiment.

2.2.2.1 Demographic Questionnaire

A demographic questionnaire (Appendix A) was administered to assess participants' age, gender, English language proficiency, own and parental education level, and own and parental income (the latter four in combination are referred to as socioeconomic status [SES]). Participants' distribution in terms of their sociodemographic background can be found in Table 1 above. Income ranges for own income for Irish participants were based on average weekly household income by net disposable household income deciles in 2022 (Central Statistics Office, 2023), as this was the most recent income range for Ireland easily accessible online. Own income ranges for UK participants were based on average equalised disposable income deciles in 2021/2022 (Office for National Statistics, 2023). Parental income ranges were based on averages from 2010, as this appeared to be an appropriate average income during important developmental stages of the current participant sample, which would have been between the ages of seven and eleven years old at this point. Irish parental income ranges were based on average weekly disposable income by gross household income deciles from 2009-2010 (Central Statistics Office, 2015), whereas UK parental income ranges were based on data from average disposable income deciles from 2010-2012 (Office for National Statistics, 2013). Both parental income and education were measured as ordinal-level data, with specific scores specified in Table 2.

Table 2

Participants' Own and Parental Income and Education Levels

Question	Score
Average disposable income growing up	
≤€9,800 (Ireland) OR *≤£11,400 (UK)*	1
€9,801 – €15,700 *(Ireland)* OR *£11,401 – £15,800 (UK)*	2
€15,701 – €22,400 *(Ireland)* OR *£15,801 – £20,400 (UK)*	3
€22,401 – €28,600 *(Ireland)* OR *£20,401 – £25,900 (UK)*	4
€28,601 – €34,800 *(Ireland)* OR *£25,901 – £32,100 (UK)*	5
€34,801 – €41,700 *(Ireland)* OR *£32,101 – £39,300 (UK)*	6
€41,701 – €50,600 *(Ireland)* OR *£39,301 – £48,100 (UK)*	7
€50,601 – €61,600 *(Ireland)* OR *£48,101 – £60,100 (UK)*	8
€61,601 – €76,600 *(Ireland)* OR *£60,101 – £80,700 (UK)*	9
≥€76,600 *(Ireland)* OR *≥£87,700 (UK)*	10
Present annual disposable income	
≤€12,700 *(Ireland)* OR ≤£12,300 *(UK)*	1
€12,701 – €19,900 *(Ireland)* OR £12,301 – £21,400 *(UK)*	2
€19,901 – €28,300 *(Ireland)* OR £21,401 – £27,100 *(UK)*	3
€28,301 – €35,600 *(Ireland)* OR £27,101 – £32,400 *(UK)*	4
€35,601 – €42,900 *(Ireland)* OR £32,401 – £37,000 *(UK)*	5
€42,901 – €51,500 *(Ireland)* OR £37,001 – £43,400 *(UK)*	6
€51,501 – €61,300 *(Ireland)* OR £43,401 – £49,600 *(UK)*	7
€61,301 – €72,900 *(Ireland)* OR £49,601 – £58,600 *(UK)*	8
€72,901 – €89,100 *(Ireland)* OR £58,601 – £70,500 *(UK)*	9
≥€89,100 *(Ireland)* OR >£70,500 *(UK)*	10
Primary caregiver's education	
No formal education	0
Some secondary school	1
Junior Certificate / GCSE Examinations	2
Leaving Certificate / A Levels	3
Certificate of Higher Education/Diploma	4
College/ University Degree (Undergraduate)	5
Master's Degree (Postgraduate)	6
Ph.D	7
Own Education	
No formal education	0
Some secondary school	1
Junior Certificate / GCSE Examinations	2
Leaving Certificate / A Levels	3
Certificate of Higher Education/Diploma	4
College/ University Degree (Undergraduate)	5
Master's Degree (Postgraduate)	6
Ph.D	7

2.1.2.2 Elaborated Relational Abilities Index

The elaborated RAI (Cummins et al., 2023, original version introduced in Colbert et al., 2015) is a measure of relational reasoning conceptualised in terms of the fluency of one's relational skills. The test involves the presentation of a series of questions that can be answered by deducting relationships based on the presentation of relevant premises. Questions are answered by choosing a yes or no option on the computer screen by clicking on it with the computer mouse. All stimuli in the logical premises are nonsense syllables, consequently rendering the problems entirely abstract. Answers to questions are time-limited to 30 seconds. This RAI version consists of eight relational subscales, each relating to a different relational frame (opposition, difference, quantity, temporal, containment, analogy, deictic, and mathematical) and consisting of 16 items (frames, presentation of premises and response options are illustrated in Figure 4). The RAI includes a total of 128 trials equally distributed between the subscales, with 16 logical reasoning tasks for each frame. As with previous versions of the RAI, task complexity is controlled for by varying the number (two to four) and order (sequential/random) of premises, directionality of the relational question, number of relation types per trials, and finally the presence/absence of relational cues presented in the relational premise(s) and question. Data were recorded on accuracy (binary format) and response time (milliseconds). Accuracy scores were computed as the mean of correct responses.

Figure 4

RAI Trials From Each Subscale

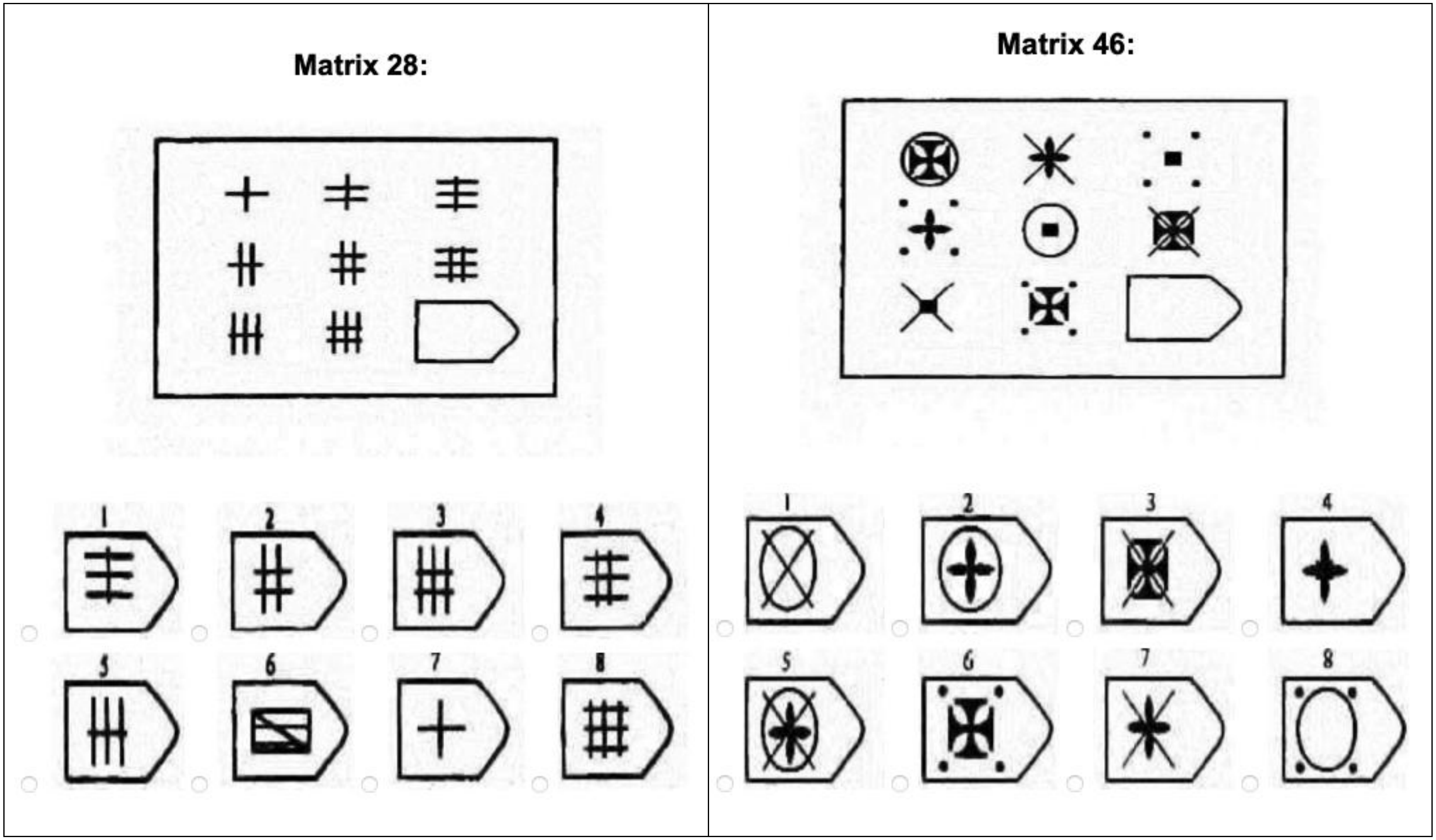
Opposition	**Difference**
Trial 1 out of 128 GOG is the same as VUL VUL is the same as HAV *Is GOG the same as HAV?*	Trial 17 out of 128 JIP is the same as JUC JUC is the same as BER *Is BER the same as JIP?*
Quantity	**Temporal**
Trial 33 out of 128 XUS is more than ROX ROX is more than TUV *Is XUS more than TUV?*	Trial 49 out of 128 DIL is before VUC VUC is before YUG *Is YUG before DIL?*
Containment	**Analogy**
Trial 65 out of 128 GIR contains TAH *Does TAH contain GIR?*	Trial 81 out of 128 MOH is the same as SUW WUN is the same as YUS *Is MOH to SUW the same as WUN to YUS?*
Deictic	**Mathematical**
Trial 97 out of 128 TIY is here QIH is there If here is there and there is here *Is TIY there ?*	Trial 113 out of 128 QUP plus NOC is the same as YEV plus HIL QUP is more than YEV *Is NOC more than HIL?*

2.1.2.3 Raven's Standard Progressive Matrices

Participants were administered the RSPM in its revised 1956 order to allow the analysis of the RAI's convergence with a widely accepted proxy measure of fluid intelligence. RSPM consist of 60 items divided into five blocks of 12 items each for which participants are given a total time limit of 40 minutes. The tasks require the identification of the missing piece required to complete a presented pattern in a 3x3 matrix. Ergo, there are eight elements presented in the matrix and participants are asked to identify which of the 8 options presented beneath the matrix completes the pattern. All items and response options are presented as visual stimuli where participants are instructed to press the key corresponding to their chosen response option (1-8, see Figure 5 for a sample items). Data were recorded on accuracy and response time and performance is computed as the mean number of correct responses (ranging from 0-60). Frequency scores were computed using the same adjusted equation as for the RAI frequency scores.

Figure 5

Sample Items From RSPM



2.2.3 Procedure

All participants were provided with an on-screen information sheet (Appendix B), as well as consent form (Appendix C) at the outset. Data collection was entirely anonymous with no identifying data being collected and began with the set of sociodemographic questions. The next phase of the experiment was the administration of the elaborated RAI followed by the final measure, the RSPM. Lastly, participants were provided a debrief sheet (Appendix D). Participants were not provided any feedback on their performance at any point of the study or afterwards. The entire procedure (from consent form to submission of data) took participants approximately 45 minutes to complete.

2.2.4 Ethical Considerations

Ethical compliance was ensured through the receipt of ethical approval by the Maynooth University Research Ethics Committee (BSRESC-2023-37306), the provision of detailed information and debrief sheets, the requirement of informed consent, and anonymous, confidential, and password protected gathering and handling of all data. The ethical guidelines complied with were those laid out by the Psychological Society of Ireland's Code of Professional Ethics (2019) and the Maynooth University Research Ethics Committee. As explained in the information sheet, participation did not bear any known risks to physical or mental health or of distress but participants were directed to contact the supervisor or the primary researcher should they have any concerns about the study. No personally identifiable information was collected and apart from gender, age, and own and parental income no individual level information was inquired about. Since the RAI and RSPM are not clinical measures, no diagnostic information was provided by the collected data.

2.3 Results

RAI scores indicate the sum of correct responses on the total RAI. Correspondingly, individual subscale scores indicate the sum of correct scores on the reflective subscales. After collection of the data, it was realised that the RAI's mathematical subscale differed somewhat from the other subscales in the sense that it was not a conventionally used subscale of relational reasoning as typically defined within RFT. Therefore, an additional score, named *RAI-M*, was computed, which consisted only of the "classical" subscales (i.e., a total RAI score which excluded the mathematical subscale.

As the RAI's mathematical subscale was included originally in the elaborated RAI for purely exploratory reasons, a second overall RAI score was calculated with the exclusion of the mathematical subscale score. This score will be referred to as the *RAI-M* and all analyses including the classical RAI score were repeated with the *RAI-M* score. After assessing the RAI's convergence with RSPM via a Spearman's rho correlation analysis (RQ1), a correlation matrix was computed to assess overall correlation between the assessed variables (RQ2). Finally, the impact of SES on RAI and RSPM performance was analysed using Multiple Regression Analyses (RQ3).

2.3.1 Missing Data and Excluded Cases

While excluding cases and missing data is often done based on e.g., the two-standard-deviation method (Dunn, 2021), it is important that the decisions regarding the exclusion of outliers is always arbitrary. Therefore, the current analysis aimed to only exclude data which was viewed as originating from a different data-generating process than the one of interest (e.g., distracted or random responding) in order to exclude fewer cases. As a result, the current analysis only excluded cases where the RSPM were not completed (N = 3), or the RAI reaction time was less than a minute for the entire RAI (N = 39). While some might argue that accuracy scores below chance level (i.e., 50%) should be excluded as well, as this could point to random responding (e.g., Colbert, 2015), the current analysis did not exclude those cases, as it was assumed that below chance level responding would only occur in cases of intentional responses. This is because the distribution of correct and incorrect responses is even, therefore, clicking the same response every time would result in 50% accuracy. Therefore, accuracy levels below chance level were not excluded from analysis. The final sample size used in this analysis was 156. However, before the criteria around exclusion of cases were finalised, analyses were run with case exclusions based on a two-standard-deviation method applied to reaction times. While those results slightly differed in effect sizes, the statistical significance generally did not change, therefore, the analyses can be found in Appendix E, but the canonical analyses are those reported in the study here. It is important for the reader to understand that the decision on how to exclude observations was not planned in advance and was decided upon after data were collected but was not conditional on the results of the original analyses. Instead, it was based on the conceptual argument that was only figured out after data collection.

2.3.2 Descriptive Statistics

As can be seen in Table 3, the mean accuracy for the full RAI was 71%. Individual subscale accuracies reached similar numbers, with the lowest being the mathematical and the opposition subscales. The highest mean accuracy was achieved for the difference subscale.

Table 3

Descriptive Statistics for Total RAI, RAI Subscales, and RSPM Accuracy

Variable	Mean	Variance	SD	Minimum	Maximum	Range	Skewness	Kurtosis	CV
RAI	.71	.01	.12	.43	.92	.50	-.24	.39	16.90
RAI-M	.72	.01	.12	.43	.91	.48	-.41	-.56	16.67
RSPM	44.70	64.71	8.04	22	58	36	-.35	-.46	17.99
Opposition	.64	.02	.14	.18	.88	.71	-.78	.33	21.88
Difference	.87	.03	.18	.13	1.00	.88	-1.69	2.26	20.69
Quantity	.76	.03	.17	.31	1.00	.69	-.45	-.72	22.37
Temporal	.73	.04	.19	.25	1.00	.75	-.39	.72	26.03
Containment	.72	.04	.20	.13	1.00	.88	-.31	-.70	27.78
Analogy	.66	.01	.12	.31	.88	.56	-.98	.33	18.20
Deictic	.65	.04	.20	.19	1.00	.81	.00	-.86	30.77
Mathematical	.64	.04	.19	.25	1.00	.75	.37	-.77	29.69

Note. $N = 156$

2.3.3 Reliability

Reliability analyses were run for the full RAI, each individual subscale, and between-subscale correlations. Specific scores are shown in Table 4. The highest reliability was found for the full RAI and the RAI-M, whereas the opposition and analogy subscales showed low reliability.

Table 4

Cronbach's Alpha Coefficients for the Full RAI and Its Individual Subscales

Scale	Cronbach's Alpha
Full RAI	.91
RAI-M	.90
Opposition	.39
Difference	.85
Quantity	.66
Temporal	.71
Containment	.74
Analogy	.37
Deictic	.70
Mathematical	.70
Between subscales	.83

2.3.4 Inferential Statistics

2.3.4.1 RQ1. Convergence Between the RAI and RSPM

A Spearman's rho correlation found that there was a strong significant relationship between RAI and RSPM performance (rho(155) = .71, $p < .001$, CI (95%) .61, .78).

2.3.4.2 RQ2. Correlation Matrix

As data were not distributed normally, a Spearman's rho correlational matrix was produced to investigate the correlations between age, own and parental incomes and education, as well as performance on the full RAI, its individual subscales, and RSPM. Specific correlations are shown in Table 5. Significant correlations were found between parental income and RAI performance, RAI-M performance, RSPM performance, and performance on the Difference, Quantity, and Temporal subscales. Own income significantly correlated with parental education, as well as performance on the full RAI, RAI-M, and the Quantity and Mathematical subscales. Parental education was significantly correlated with own education, as well as performance on the full RAI, RAI-M, RSPM, and all individual RAI subscales apart from the Analogy and Deictic subscales. Own education did not significantly correlate with any RAI subscales or RSPM. Performance on the full RAI and RAI-M was significantly correlated with Performance on RSPM. Apart from the opposition and analogy subscale performance correlation, performance on all individual subscales was significantly correlated with performance on all other subscales, reaching correlation coefficients between .17 (opposition and deictic subscales) and .73 (quantity and temporal).

Table 5

Spearman's rho Correlation Matrix for RAI and RSPM Accuracy Scores and Demographic Variables

Variables	Age (CI [95%])	1 (CI [95%])	2 (CI [95%])	3 (CI [95%])	4 (CI [95%])	5 (CI [95%])	6 (CI [95%])	7 (CI [95%])	8 (CI [95%])	9 (CI [95%])	10 (CI [95%])	11 (CI [95%])	12 (CI [95%])	13 (CI [95%])	14 (CI [95%])
1. Parental Income	-.02 (-.18, .15)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2. Own Income	.04 (-.12, .20)	.63** (.52, .72)	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Parental Education	.01 (-.54, .17)	.33** (.18, .47)	.22** (.06, .37)	-	-	-	-	-	-	-	-	-	-	-	-
4. Own Education	.39** (.24, .51)	-.09 (-.24, .08)	.03 (-.13, .19)	.20* (.04, .35)	-	-	-	-	-	-	-	-	-	-	-
5. RAI	.03 (-.14, .19)	.19* (.03, .34)	.18* (.02, .33)	.27** (.12, .42)	.07 (-.10, .23)	-	-	-	-	-	-	-	-	-	-
6. RAI-M	.03 (-.14, .19)	.18* (.02, .33)	.16* (-00, .31)	.27** (.11, .41)	.06 (-.10, .22)	.99** (.98, .99)	-	-	-	-	-	-	-	-	-
7. RSPM	-.05 (-.21, .11)	.17* (.01, .33)	.15 (-.01, .31)	.15* (-.01, .31)	-.08 (-.24, .08)	.71** (.62, .78)	.68** (.58, .76)	-	-	-	-	-	-	-	-
8. Opposition	-.03 (-.19, .14)	.05 (-.21, .11)	.02 (-.14, .19)	.16* (.00, .31)	.03 (-.13, .19)	.49** (.36, .61)	.52** (.39, .63)	.30** (.15, .44)	-	-	-	-	-	-	-
9. Difference	.03 (-.13, -.19)	.19* (.03, .34)	.14 (-.02, .29)	.19* (.03, .34)	.03 (-.13, .19)	.73** (.64, .80)	.74** (.66, .81)	.52** (.38, .63)	.43** (.29, .55)	-	-	-	-	-	-
10. Quantity	.05 (-.11, .21)	.16* (.00, .32)	.19* (.03, .34)	.28** (.12, .42)	.00 (-.16, .17)	.81** (.75, .86)	.81** (.74, .86)	.61** (.50, .70)	.32** (.17, .46)	.54** (.42, .65)	-	-	-	-	-
11. Temporal	.04 (-.12, .20)	.18* (.02, .33)	.13 (-.04, .28)	.23** (.07, .37)	-.01 (-.17, .15)	.82** (.76, .87)	.82** (.76, .87)	.66** (.55, .74)	.32** (.17, .46)	.61** (.49, 70)	.73** (.65, .80)	-	-	-	-
12. Containment	-.01 (-.17, .15)	.14 (-.03, .29)	.10 (-.06, .26)	.26** (.10, .40)	.07 (-.10, .23)	.74** (.66, .81)	.74** (.66, .80)	.50** (.36, .61)	.26** (.10, .40)	.44** (.30, .56)	.60** (.48, .69)	.53** (.40, .64)	-	-	-
13. Analogy	.07 (-.10, .23)	.03 (-.13, .19)	.10 (-.07, .25)	.00 (-.16, .17)	.06 (-.10, .22)	.45** (.31, .57)	.45** (.31, .57)	.27** (.11, .41)	.13 (-.04, .28)	.32* (.17, .46)	.27* (.11, .41)	.30** (.15, .44)	.31** (.15, .45)	-	-
14. Deictic	-.01 (-.17, .15)	.06 (-.10, .22)	.07 (-.10, .23)	.10 (-.06, .26)	.12 (-.05, .27)	.56** (.44, .67)	.59** (.48, .69)	.32** (.16, .46)	.17* (.01, .33)	.31** (.15, .45)	.35** (.20, .48)	.35** (.20, .49)	.30** (.14, .44)	.14 (-.02, .30)	-
15. Mathematical	.06 (-.11, .22)	.17 (.01, .32)	.20* (.04, .35)	.19* (.03, .34)	.05 (-.11, .21)	.76** (.69, .82)	.66** (.56, .74)	.63** (.52, .72)	.24** (.08, .38)	.47** (.34, .59)	.61** (.50, .70)	.62** (.51, .71)	.54** (.42, .65)	.28** (.13, .43)	.31** (.15, .45)

**. Correlation is significant at the .01 level (2-tailed); *. Correlation is significant at the .05 level (2-tailed).

2.3.4.3 RQ3. SES as a Predictor of RAI and RSPM Performance

Coefficients for insignificant multiple regression models can be found in Appendix F.

Multiple regression was used to test whether own income, parental income, own education, and parental education were predictors of RAI performance. The results of the regression indicated that the overall model explained 6% of the variance (R^2 = .06, $F(4, 151)$ = 3.56, p = .008). It was found that parental education significantly predicted RAI performance (β= .02, p = .008, *95% CI* = .00, .03) while the other variables did not.

Multiple regression was used to test whether SES predicted RSPM performance. The results of a multiple regression analysis indicated that the overall model did not significantly explain the variance in RSPM performance (R^2 = .05, $F(4, 150)$ = 1.98, p = .100).

Multiple regression was used to test whether SES predicted RAI-M performance. A multiple regression analysis indicated that the overall model explained 5.5% of the variance (R^2 = .06, $F(4, 151)$ = 3.27, p = .013) in RAI-M scores. It was found that parental education significantly predicted RAI-M accuracy (β= .02, p = .010, *95% CI* = .00, .03) while the other variables did not.

2.4 Discussion

2.4.1 Summary

Research question 1 investigated the correlation between RAI and RSPM performance and found a strong, significant relationship. The implications of this will be discussed in section 2.4.2. Research question 2 investigated general correlations between the variables in the dataset and found significant correlations between most variables. Most notably however, were the significant correlations between parental education and RAI performance and the non-significant correlations between own education and RAI performance. The final, third, research question investigated SES as a predictor of RAI and RSPM performance and found the four SES variables to significantly predict RAI and RAI-M but not RSPM performance. In the case of significant correlations, parental education stood out as the sole significant predictor. This will be discussed in detail in section 2.4.3.

2.4.2 Convergence Between the RAI and RSPM

Overall, the RAI had a strong, significant correlation with RSPM. While this suggests good convergent validity, it is important to remember that the aim of the RAI is not to correlate with traditional measures of intelligence in order to create a new measure of "intelligence". Instead, the aim was to investigate further the various ways in which relational skills may converge with other intellectual performances, such as those on nonverbal tests of intelligence (e.g., RSPM). Thus, it is important to understand that a strong correlation between the two measures no more validates the RAI than it validates the RSPM. That is, the current thesis is merely mapping out the statistical relationships between these measures as part of a broader effort beyond this thesis to understand functional relationships between relational skills and a wide range of other activities and test performances. The strong correlation found between RAI performance and RSPM performance highlights, at the very least, the high relevance of relational skills to performance on one widely used and respected measure of nonverbal intelligence and a respected index of *g*, supporting various theoretical positions in the behaviour analytical field (Andres & Halford, 1998; Barnes-Holmes et al., 2010; Cassidy et al., 2010; Cattell, 1971; Dixon et al., 2014; Gentner & Loewenstein, 2002; Gore et al., 2010; Moran et al., 2010; O'Hora et al., 2005, 2008; O'Toole & Barnes-Holmes, 2009, Stewart et al., 2013).

Bearing in mind that the RAI was not devised as a proxy measure for IQ but as a potential substitute for and improvement upon traditional intelligence measures, the level of correlation between RAI and RSPM performance is rather telling. That is, the level of correlation closely resembles that previously found between other IQ tests in convergent validity studies. For example, while the RAI correlated with RSPM with the coefficient of .71 in the current study, the correlation between RSPM performance and performance on the WAIS-III has been found to be similar but slightly lower in most studies (.65, Fletcher & Hattie, 2011; .64, Wechsler, 1997; .49 to .79, Silva, 2008). Correlations between other measures of intelligence were at comparable levels. For instance, one of the most widely used measures of intelligence, the WAIS-III, correlated with other measures at .53 to .81 (Wechsler Individual Achievement Test; Silva, 2008), .71 or .82 (Woodcock-Johnson Test of Cognitive Ability; Cherame et al., 2012; Metz, 2005, respectively), .75 or .80 (General Ability Measure for Adults, Naglieiri & Bardo's, 1997 and Martin et al., 2000, respectively). Additionally, correlations between WASI Verbal IQ and WAIS Full IQ were at the .75 level (Axelrod, 2002), whereas the correlation between the WAIS and Stanford-Binet Intelligence Scale (Fourth Edition; Thorndike et al., 1986) fell within the .77 to .89 levels (Groth-Marnat, 2003; Kamphaus, 2005; Silva, 2008; Zimmerman & Woo-Sam, 1973). These correlation coefficients highlight the utility of RAI scores as predictors of IQ as assessed by several gold standard tests.

Interestingly, Colbert (2015) explained that correlation coefficients between tests of intelligence are often affected by low variability within the test score ranges. More specifically, low levels of variance can substantially decrease the level of correlation between two measures. Indeed, the current study found variance in RAI scores was very low (for both overall scores as well as individual subscale scores, ranging from 16.90 to 30.77, thus only covering a range of 13.87). As correlational analyses are easily impacted by the variability of data (Abrami et al., 2001; Aron & Aron, 2003; Bates et al., 1996; Crocker & Algine, 1986; Glenberg, 1996; Hopkins, 1998), lower variability decreases correlation coefficients (Goodwin & Leech, 2006). Therefore, it is reasonable to assume that, were RAI scores distributed more evenly, the correlation between RAI and RSPM performance would have likely been even stronger. Nonetheless, RAI performance predicted RSPM performance to a meaningful extent, comparable to other IQ test correlations found in previous studies.

While the high correlation coefficient highlights the RAI´s utility as a possible proxy measure of intelligence, it also and equally could be interpreted as evidence that traditional IQ tests can be conceived of as measures of relational responding (Cassidy et al., 2010). The fact that the correlation between RAI and RSPM performance is not perfect, however, is explicable in several different ways. For instance, it might be the case that performance on RSPM is subject to external variables, which are not directly related to cognitive abilities but nevertheless impact performance (see below). Alternatively, it might be the case that there are other, undiscovered (or perhaps simply not yet integrated into the measure) relational frames that would allow RAI performance to fully account for RSPM performance. The first possible explanation will be explored below, while the latter possible explanation will be explored in Chapter 3.

There is valid reason to assume that RSPM performance is not solely determined by intellectual ability. Indeed, nonverbal measures of IQ, such as RSPM, have been found to be particularly susceptible to non-cognitive variables such as attention, mood, and task persistence (Kaufman, 1990; Njiokiktjien & Verschoor, 1998; Sackheim et al., 1992). Similarly, non-cognitive factors might impact performance on both measures simultaneously, though perhaps to different degrees. Stankov et al. (2014) provides an overview indicating that various non-cognitive variables were found to significantly predict intelligence test performance to meaningful degrees. Scores on rationality, openness to experience, and self-concept measures, for instance, were found to predict IQ at correlation coefficients of up to .35. However, he argued, that the best non-cognitive predictor of cognitive performance are scores on measures of self-confidence, with correlation coefficients of up to .45. In terms of non-cognitive predictors of RSPM performance in particular, Birney et al. (2017) found neuroticism to have a significant impact on RSPM performance. Therefore, it might be the case that the skills assessed in the RAI fully predict the cognitive component of RSPM test performance, with the remaining variance attributed to non-cognitive variables. Of course, this interpretation is highly speculative but is outlined here purely for illustrative purposes in terms of addressing the issue of the meaning of divergences between measures. That is divergences are not always based on measurement error or differences in the constructs assessed by those measures.

2.4.3 SES as a Predictor of RAI and RSPM Performance

Another research question investigated in the current thesis related to the impact that socioeconomic variables, more specifically, own and parental education, as well as own and parental income, have on RAI performance, relative to the impact of these variables on RSPM performance. The current thesis found that these four SES variables predicted accuracy on the RAI but not on RSPM. It is important for the reader to remember why this outcome may be conceptually important.

More specifically, as outlined in section 1.8, the current thesis investigated own and parental education and income, as these have not only been found to correlate with traditional measures of intelligence, but more importantly, it is believed that these variables provide insight into an individual's exposure to educational experiences, which, in turn, is suggested to facilitate the acquisition of relational responding proficiency.

Having reminded the reader of why these socioeconomic variables are of importance, some methodological shortcomings of these arguments will be laid out before the findings will be discussed in more detail. Firstly, the range of own education levels represented in the current dataset was compromised. This is perhaps not surprising, given that participants were all in their early twenties, and culturally coherent as a group. Thus, they shared many commonalities with regard to education and routes to education. Specifically, in the Republic of Ireland, the majority of school leavers go on to third level education and a significant number of these are entitled to government financial supports in a concerted effort to harmonise access to higher education (Department of Further and Higher Education, Research, Innovation, and Science, 2022). Ireland ranks third in the OECD nations in terms of attendance at third level. The differences in educational opportunity within the current participant sample were constrained even further by the fact that some level of recruitment occurred within a University setting, thus further enhancing the homogeneity of the participant sample. Indeed, the participants' income distribution was skewed towards the lower income ranges, supporting the idea that many of these individuals were likely students in third level education. Parental education and income were more evenly distributed. However, an argument can be made that they are not sufficient to serve as indices for educational opportunity. That is, as Hart and Risley (1995) pointed out, measuring SES does not offer a complete picture of someone's access to educational opportunity. While parental education was found to predict the quality of educational interactions at home in their study, it is premature to assume that parental education necessarily and directly translates into those interactions. That is, a parent's education level does not necessarily predict that they will spend sufficient time with their children for this variable to have a direct effect on the intellectual capacity of their child. Indeed, some educated parents may spend very little time with their children at all. This is true for both low and high-income parents (Council of Economic Advisers, 1999). While the current study aimed to address this issue by recording education levels for the *primary* caregiver, enquiring about frequency of interactions more specifically may have provided data that was more predictive of educational advantage.

Similarly, while parental income might be broadly indicative of educational opportunity either directly provided by the parent or through paid-for services, it is at best, an indirect measure. Various factors, such as the number of siblings that the parents need to support, debt, and perhaps even health-related expenses, can significantly impact the amount of disposable income that a parent has to spend on their child's education. Thus, it may have been more informative to record details about access to extracurricular activities, private tutoring and the quality of schooling received. Given all of these considerations, the following interpretations of the current findings need to be made with caution. With that said, a summary of the findings in relation to how various aspects of SES impact performance on the RAI and RSPM will now be provided.

Own education was not found to be a significant predictor of RAI or RSPM performances. From an RFT perspective, this is a surprising finding, as more years of formal education would be expected to result in more extensive training and rehearsal of complex forms of derived relational responding. Such an enhanced repertoire attained over a long number of years should in turn lead to advanced academic skills and more crystallised intellectual skills in terms of knowledge content. Perhaps, however, a failure to observe a relationship between one's own level of education and their proficiency at relational responding and RSPM makes sense in the context of the caveats outlined above. That is, the current study may have failed to secure a sufficiently wide range of individual participants with sufficiently varying educational histories, as well as a sufficiently wide range of RAI performance. At the same time, however, it is important to remember that this was not an experimental study and was never intended to establish functional relationships between these two variables. The approach here was entirely exploratory and correlational. Even if strong correlations had been observed, it would have supported a RFT position, but in no way proven it. What one is left with, however, is uncertainty regarding the absence of such a correlation but with sufficient methodological concern that no firm conclusions can be drawn. Thus, it would be unreasonable to dismiss the contention that educational opportunity is related to relational responding proficiency. Indeed, previous studies suggests that this relationship is somewhat robust (e.g., Strenze, 2007). However, some researchers may be cautious even where correlations between years of education and intellectual skill proficiency are robust and unambiguous. That is, it could be reasonably argued that levels of education correlate with levels of measured intelligence only because more intelligent people tend to pursue more years of education compared to lower IQ individuals (e.g., Deary et al., 2007; Deary & Johnson, 2010, Roth et al., 2015). Alternatively, some have suggested that indeed it is education that results in higher IQ scores rather than vice versa (e.g., Ceci, 1991).

In an effort to address the issue of causality in the relationship between education and intelligence, Brinch and Galloway (2012) analysed the effect of a new compulsory schooling reform introduced in Norway from 1955 to 1970. More specifically, the reform increased the number of compulsory years of schooling from seven to nine, established a new unified type of middle school for grades seven to nine, and standardised the baseline academic curriculum at middle-schools. Indeed, Brinch and Galloway (2012) found that following the changes based on this school reform, substantial raises in average IQ scores of 19-year-old military recruits (male) were observed. Studies like this allow for more confidence regarding the causal link of the education-IQ correlation and allow certainty about the direction of this link. Indeed, Ritchie and Tucker-Drob (2018) conducted a meta-analysis of 48 datasets ($N > 600,000$) based on various study types which would allow more insight into the causation and its direction of the education-IQ correlation. More specifically, the three quasi-experimental designs aiming to minimise endogeneity confounds based on selection processes were; (1) controlling for baseline IQ, (2) investigating the impact of policy changes, and (3) investigating the impact of school-age cutoff. Their findings were in line with the current argument. That is, it appears to be the case that education predicts IQ through a causal relationship, likely due to the fact that education establishes the relevant intellectual skills. More specifically, each year of education was found to raise intelligence by one to five standardised IQ points. This effect was found not to be limited to certain levels of cognitive ability and likely to persist across the lifespan. Still, the question remains whether this positive impact is consistent regardless of baseline intellectual ability prior to extra or intensive education. While the *Matthew effect* (Stanovich, 1986) argues that individuals with higher baseline cognitive abilities benefit more from education than lower baseline ability individuals, Downey et al. (2004) argued that education allows lower baseline ability individuals to "catch up" to their higher baseline ability peers. Indeed, the latter was further supported by Hansen et al. (2004). Regardless of who benefits the most, these findings further support the argument made in the Introduction, that intelligence (or at least observable intellectual skills) is more accurately conceptualised as an acquired skill malleable through environmental interventions rather than a fixed, inherited trait.

Another possibility for why the current study found no significant correlation between RAI performance and own education is that the extra education received by those with the higher levels of education as reported in the current study, was not particularly facilitative of increased complex relational responding. That is, arbitrarily applicable relational responding would form a considerable part of the content of a mathematics or a philosophy degree but perhaps not a degree in a manual task or some forms of knowledge acquisition that, for the most part, simply involve memory of information, or repetitive intellectual tasks lacking in complexity. Undertaking a study that functionally linked education type to relational reasoning improvements would be highly ambitious but is certainly needed within the literature at this time as part of the effort to understand the functional relationship between various forms of relational reasoning training and general or specific intelligences.

In all correlational analyses conducted in the current thesis, parental education was identified as either the strongest predictor (Spearman's rho correlations) or the sole significant predictor (multiple regression analyses) of RAI performance. Given the posited link between education and intelligence, it could be suggested that in this case parental education functioned as a proxy for the intelligence levels of parents, inviting an easy interpretation of the current effects in terms of genetic transmission from parent to child. Indeed, it is the consistent observation that the intelligence levels of children correlate with those of their parents and siblings that has led to this sort of low-hanging fruit explanation over the past century. While there is undoubtedly a genetic component to the development of intellectual skills, this is not typically used as explanatory within the experimental analysis of behaviour and is an argument that is flawed in many ways in terms of its ability to coherently identify causality. Firstly, as discussed above, educational attainment cannot easily be explained away by trace-level intelligence because there is much evidence to suggest that the causal relationship at least sometimes or perhaps often occurs in reverse. Moreover, it is not only premature but conceptually naïve to separate out the effects of environment and genetics (see Roche & Barnes,

1996) on any form of complex behaviour. These conceptual issues are beyond the scope of the current thesis, but the important point is that the literature abounds with evidence of the cultural transmission of intelligence through social interaction. Perhaps in particular, the work of Hart and Risley's (1995) suggests that better-educated parents provide their children with higher-quality interactions, resulting in significant differences in language acquisition, which in turn has important implications for intellectual ability. It is important to note that this does not preclude the possibility of genetic influences completely. Indeed, Hart et al. (1997) argued that "the most plausible candidate for such a fundamental cause is parental intellectual ability, operating both genetically and environmentally upon the child" (p. 3). In effect, these authors develop a nuanced and conceptually coherent approach to the transmission of intellectual skills that does not lazily rely on a simple biological explanation.

Chapter 3

An In-Depth Analysis of the Elaborated Relational Abilities Index: Further Relationships Between Test Subscales, Overall Test Scores, Socio-Economic Status, and a Factor-Analysis of Test Structure Using Primary and Secondary Data

3.1 Study 2

While Chapter 2 investigated general relationships between RAI and RSPM, as well as RAI and SES, the second study will focus on further relationships between subscales within the Elaborated RAI. Additionally, the current study aims to examine response fluency, rather than simply speed or accuracy considered separately. This is an important development as, within the behavioural field, response fluency is widely considered to capture more aspects of the behaviour than response accuracy or speed considered separately (Binder, 1993, 1996). According to RFT, a more developed relational responding repertoire results in faster response times with higher accuracy. That is, where accuracy is maximised, there is always room to improve speed. On the other hand, speed alone is of no interest if the accuracy is not at least satisfactory.

Two recent studies that have already interrogated the elaborated RAI to some extent have already been published. Specifically, Cummins et al. (2023) analysed the intercorrelations between the measure's subscales, and Cummins (2023) investigated individual-level precision and item characteristics using Item Response Theory. Given these important advances, however, no study to date has conducted a factor analysis on the RAI, either in its current form nor for any of its predecessors. Such an analysis is crucial to fully understanding how clusters of relational response repertoires may be related and interrelated. As this analysis is explanatory, it may be the case that the entire set of relational responding constitutes a single factor or that factors split along surprising lines which would raise important questions for RFT-based interpretations of these skills. Indeed, McLoughlin (2022) argued that a factor analysis would offer valuable insight into the existing factor structure and as a result, allow a more comprehensive understanding of how RFT's proposition of relational responding proficiency as an underlying skill of intelligence relates to *g* theory's proposition of a single underlying factor. Cummins

(2023) also pointed out that it is of primary importance to RFT that its conceptualisation of relational frames is based on function, rather than topography. Therefore, a confirmatory factor analysis might provide insight, for example, into how the “difficulty" of a particular relational task might better define a factor than the topography of the task (e.g., how the tasks are grouped together under the umbrella of a particular frame labelled by researchers as involving tasks of the same type conceptually). If this is the case, then items may cluster into factors irrespective of which relational frame they “belong” to.

Given all of the above, the current study aims to investigate which subscales of the RAI relate closest with the RSPM. It will further investigate which subscales are best predicted by SES and how SES impacts fluency on both the full RAI and its individual subscales. Finally, a factor analysis will provide insight into the underlying factor structure.

3.1.2 Research Questions

Overall, this study aimed to provide a more holistic and detailed insight into the role of specific RAI subscales in order to assess, which subscales might be more closely correlated with RSPM performance or more substantially predicted by SES. As all of these hypotheses were exploratory in nature, no predictions around the outcome were made before analyses were run.

3.1.2.1 RQ1. How do the individual RAI subscales correlate with RSPM?

In order to gain a more holistic understanding into how the RAI and RSPM correlate, this research question will analyse individual RAI subscales and RSPM correlations using linear regression models in which RAI subscales scores were modelled as predictor variables and RSPM was modelled as the criterion variable. This allows insight into differences of the individual subscale’s correlations with RSPM.

3.1.2.2 RQ2. How does SES predict performance on the individual RAI subscales?

Again, to provide a more holistic understanding of the correlations between SES and relational responding, the current research question investigates to what degree SES predicts performance on individual RAI subscales using a multiple regression model in which the four SES variables were modelled as the predictor variables and RAI subscales performances were modelled as the criterion variable.

3.1.2.3 RQ3. How does SES predict fluency on the full RAI, RSPM, and the individual RAI subscales?

Another approach to gain a more holistic insight is to analyse the correlational impact that SES has on fluency. Therefore, the current research question was analysed using multiple regression models in which the four SES variables were modelled as predictor variables and the RAI subscale performances were modelled as criterion variables.

3.1.2.4 RQ4. Does the RAI's current composition confirm to its underlying factor structure?

To analyse whether the current composition adheres to the underlying factor structure, a Confirmatory Factor Analysis based on the composition's eight-factor model will be conducted. While no specific hypotheses are made in advance, the factor structure could either confirm the composition based on topographical features (i.e., factor clusters adhere to the relational frame to which the items belong), or possibly to the difficulty level.

3.2 Methodology

3.2.1 Participants

The current analysis was based on a deeper investigation of patterns within the data collected and reported in Chapter 2, combined with datasets from published data. Given that the factor analysis reported here required as much statistical power as possible, data collected for the Study 1 (N = 156), was combined with the data sets of Cummins et al. (2023; N = 153), and Cummins (2023; N = 264) resulting in a total sample size of 573. Full permission was provided by the first author of both studies who was also a co-supervisor of the current work.

3.2.2 Settings and Materials

As mentioned above, the current study used secondary data from Study 1 for all analyses. Moreover, the factor analysis used data from Study 1, as well as Cummins et al. (2023) and Cummins´s (2023). Therefore, the demographic, RAI, and RSPM data were collected as described in section 2.2.

Fluency scores were calculated by the following formula, as introduced by Colbert (2015), providing a correct-incorrect differential score per minute, thus accounting for both speed and accuracy, while paying attention not to neglect incorrect responses as behaviour:

$$60{,}000\left(\frac{\begin{array}{c}Total\ correct\ responses\\ -\ Number\ of\ incorrect\ responses\end{array}}{Time\ to\ complete\ RAI}\right)$$

Fluency scores for the RSPM were computed with the same formula using RSPM data.

3.2.3 Ethical Considerations

Since the current study used secondary data, the collection of which has been approved by the Maynooth University Ethics Research Committee (BSRESC-2023-37306), and the Ghent University Ethics Committee for the Cummins et al. (2023) and Cummins (2023; approval number 2020/74) studies, no new ethical considerations had to be taken into account, as relevant procedures were complied with (see section 2.2.4).

3.3 Results

3.3.1 Descriptive Statistics

As can be seen in Table 6, the mean fluency for the full RAI was 2.60, indicating an average of 2.6 correct responses per minute (while accounting for incorrect responses). The maximum number of correct responses per minute was 10.84 for the full RAI. The highest mean fluency score was achieved for the difference subscales, with 6.40, while the lowest was .70 for the mathematical subscales.

Table 6

Descriptive Statistics for Full RAI, RAI-M, RSPM, and the RAI Subscales Fluency

Variable	Mean	Variance	SD	Minimum	Maximum	Range	Skewness	Kurtosis
Total RAI	2.60	3.65	1.91	-4.49	10.84	15.34	.70	3.04
RAI-M	2.83	3.90	1.98	-3.42	12.12	15.54	1.06	1.69
RSPM	2.15	1.78	1.33	-4.18	5.33	9.51	-.50	3.03
Opposition	1.69	4.43	2.10	-7.53	9.93	17.46	.15	4.14
Difference	6.40	75.99	8.72	-46.77	75.00	121.77	2.11	33.35
Quantity	3.61	87.56	9.36	-6.21	102.88	109.09	8.80	87.45
Temporal	4.12	193.71	13.92	-15.43	143.28	158.71	7.97	71.66
Containment	1.13	200.29	14.15	-143.55	48.60	192.14	-73.28	73.28
Analogy	2.49	34.15	5.84	-35.59	13.30	48.89	-4.29	2.61
Deictic	2.01	25.74	5.07	-22.48	28.74	51.22	.92	13.21
Mathematical	.70	97.66	9.88	-82.28	54.07	136.35	-3.30	39.04

Note. $N = 156$

3.3.2 Inferential Statistics

3.3.2.1 Correlational Matrix

A Spearman's rho correlations matrix was run to investigate the correlations between age, own and parental income and education, as well as fluency on the full RAI, its individual subscales, and RSPM. Specific correlation coefficients are shown in Table 7. Significant correlations were found between parental income with fluency on the quantity and temporal subscales. Moreover, own income significantly predicted fluency on the mathematical subscales, while parental education significantly predicted total RAI and RAI-M fluency, as well as fluency on the opposition, difference, quantity, temporal, containment, and mathematical subscales. Total RAI and RAI-M fluency scores significantly predicted RSPM fluency and all subscale fluency scores apart from the quantity and analogy, analogy and temporal, and analogy and deictic subscale correlations, reaching coefficients between .16 (analogy and opposition) and .61 (opposition and difference).

Table 7

Spearman's rho Correlation Matrix for RAI and RSPM Fluency Scores and Demographic Variables

Variables	Age (CI [95%])	1 (CI [95%])	2 (CI [95%])	3 (CI [95%])	4 (CI [95%])	5 (CI [95%])	6 (CI [95%])	7 (CI [95%])	8 (CI [95%])	9 (CI [95%])	10 (CI [95%])	11 (CI [95%])	12 (CI [95%])	13 (CI [95%])	14 (CI [95%])
1. Parental Income	-.02 (-.18, .15)	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2. Own Income	.04 (-.12, .20)	.63** (.52, .72)	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Parental Education	.01 (-.54, .17)	.33** (.18, .47)	.22** (.06, .37)	-	-	-	-	-	-	-	-	-	-	-	-
4. Own Education	.39** (.24, .51)	-.09 (-.24, .08)	.03 (-.13, .19)	.20* (.04, .35)	-	-	-	-	-	-	-	-	-	-	-
5. RAI	-.06 (-.22, .11)	.15 (-.01, .31)	.05 (-.12, .21)	.32** (.17, .46)	.05 (-.11, .21)	-	-	-	-	-	-	-	-	-	-
6. RAI-M	-.05 (-.21, .11)	.13 (-.03, .29)	.02 (-.14, .18)	.32** (.16, .46)	.06 (-.11, .22)	.99** (.99, .99)	-	-	-	-	-	-	-	-	-
7. RSPM	-.12 (-.28, .04)	.08 (-.09, .24)	.09 (-.07, .25)	.06 (-.10, .22)	-.11 (-.27, .05)	.48** (.35, .69)	.47** (.33, .59)	-	-	-	-	-	-	-	-
8. Opposition	-.07 (-.23, .09)	.06 (-.10, .22)	-.06 (-.22, .11)	.20* (.04, .35)	.00 (-.16, .16)	.65** (.55, .74)	.67** (.57, .75)	.37** (.22, .50)	-	-	-	-	-	-	-
9. Difference	-.05 (-.21, .12)	.09 (-.08, .25)	.03 (-.13, .20)	.21** (.05, .36)	.01 (-.16, .17)	.73** (.64, .80)	.74** (.66, .80)	.39** (.24, .52)	.61** (.50, .70)	-	-	-	-	-	-
10. Quantity	.02 (-.15, .18)	.16* (.00, .32)	.07 (-.09, .23)	.35** (.20, .48)	.04 (-.13, .20)	.74** (.65, .80)	.73** (.65, .80)	.38** (.23, .51)	.43** (.29, .55)	.43** (.29, .56)	-	-	-	-	-
11. Temporal	-.05 (-.21, .12)	.16* (.00, .31)	.03 (-.13, .19)	.20* (.04, .35)	-.02 (-.18, .14)	.71** (.62, .78)	.71** (.62, .78)	.46** (.32, .58)	.39** (.25, .52)	.47** (.33, .58)	.48** (.35, .60)	-	-	-	-
12. Containment	-.03 (-.19, .13)	.10 (-.07, .25)	.01 (-.16, .17)	.25* (.09, .40)	.07 (-.09, .23)	.61** (.49, .70)	.60** (.49, .70)	.29** (.13, .43)	.31** (.16, .45)	.29** (.13, .43)	.49** (.36, .61)	.40** (.26, .53)	-	-	-
13. Analogy	.08 (-.09, .24)	-.03 (-.19, .13)	.04 (-.13, .20)	-.01 (-.17, .15)	.03 (-.14, .19)	.31** (.16, .46)	.33** (.17, .46)	.14 (-.02, .30)	.16* (.00, .31)	.21** (.05, .36)	.16 (-.01, .31)	.14 (-.02, .29)	.21** (.05, .36)	-	-
14. Deictic	-.05 (-.21, .11)	.01 (-.16, .17)	.00 (-.16, .16)	.10 (-.07, .26)	.13 (-.03, .29)	.44** (.31, .57)	.45** (.31, .57)	.21* (.05, .36)	.18* (.02, .33)	.20** (.04, .35)	.20* (.04, .35)	.34** (.19, .48)	.22** (.06, .37)	.05 (-.11, .21)	-
15. Mathematical	-.03 (-.14, .19)	.15 (-.01, .31)	.20* (.04, .35)	.16* (.00, .32)	.03 (-.14, .19)	.58** (.31, .57)	.50** (.37, .61)	.50** (.32, .58)	.21** (.05, .36)	.39** (.25, .52)	.45** (.32, .57)	.41** (.26, .53)	.37** (.23, .51)	.18* (.02, .33)	.18* (.12, .33)

**. Correlation is significant at the .01 level (2-tailed); *. Correlation is significant at the .05 level (2-tailed).

3.3.2.2 RQ1. Correlation Between Individual RAI Subscales and RSPM Performance

To reduce the increased risk of a Type I error, which is due to the number of statistical tests performed, resulting in a Family-Wise Error Rate (FWER) of $(1 - (1 - .05)^{8}) = .34$, the following analyses will also be inspected with a Bonferroni-corrected α level of 0.5/8 = .006, which brings the FWER back to .05. While the Bonferroni correction decreases the risk of a Type I error, it simultaneously increases the risk of a Type II error (false negative; Vanderweele & Mathur, 2019). Therefore, a Benjamini-Hochberg correction was performed. As all the below analyses showed a p-value of <.001, the p-values were ranked in the order of analysis performed. As the Bonferroni-adjusted α level and the lowest Benjamini-Hochberg critical value were both .006, and therefore all p-values still fall below the critical value, all results remained significant.

A linear regression analysis found that RAI opposition subscale scores significantly predicted RSPM scores ($F(1, 153) = 13.84$, $p < .001$, $R^2 = .08$) (RAI, $\beta = 17.03$, $p < .001$, CI (95%) 7.99, 26.07).

A linear regression analysis found that RAI difference subscale scores significantly predicted RSPM scores ($F(1, 153) = 41.53$, $p < .001$, $R^2 = .21$) (RAI, $\beta = 20.10$, $p < .001$, CI (95%) 13.94, 26.26).

A linear regression analysis found that RAI quantity subscale scores significantly predicted RSPM scores ($F(1, 153) = 71.47$, $p < .001$, $R^2 = .31$) (RAI, $\beta = 26.01$, $p < .001$, CI (95%) 19.93, 32.09).

A linear regression analysis found that RAI temporal subscale scores significantly predicted RSPM scores ($F(1, 153) = 104.80$, $p < .001$, $R^2 = .40$) (RAI, $\beta = 26.97$, $p < .001$, CI (95%) 21.76, 32.17).

A linear regression analysis found that RAI containment subscale scores significantly predicted RSPM scores ($F(1, 153) = 44.86$, $p < .001$, $R^2 = .22$) (RAI, $\beta = 18.92$, $p < .001$, CI (95%) 13.34, 24.49).

A linear regression analysis found that RAI analogy subscale scores significantly predicted RSPM scores ($F(1, 153) = 13.69$, $p < .001$, $R^2 = .08$) (RAI, $\beta = 19.49$, $p < .001$, CI (95%) 9.08, 29.89).

A linear regression analysis that RAI deictic subscale scores significantly predicted RSPM scores ($F(1, 153) = 17.01$, $p < .001$, $R^2 = .10$) (RAI, $\beta = 12.73$, $p < .001$, CI (95%) 6.65, 18.81).

A linear regression analysis found that RAI mathematical subscale scores significantly predicted RSPM scores ($F(1, 153) = 103.28$, $p < .001$, $R^2 = .40$) (RAI, $\beta = 26.26$, $p < .001$, CI (95%) 21.15, 31.36).

3.3.2.3 RQ2. SES as a Predictor of Individual RAI Subscale Performance

Coefficients for insignificant multiple regression models can be found in Appendix F. To reduce the increased risk of a Type I error, which is due to the number of statistical tests performed, resulting in a FWER of $(1 – (1 - .05)^{8}) = .34$, the following analyses will also be inspected with a Bonferroni-corrected α level of 0.5/8 = .006, which brings the FWER back to .05. Again, to protect against an inflated Type II error risk (Vanderweele & Mathur, 2019), a Benjamini-Hochberg correction was performed and the results will be reported and interpreted accordingly.

Multiple regression indicated that the overall model did not significantly explain the variance in accuracy on the RAI Opposition subscale ($R^2 = -.00$, $F(4, 151) = .92$, $p = .457$). The findings remain non-significant even with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .038.

Multiple regression indicated that the overall model did not significantly explain the variance in accuracy on the RAI Difference subscale (R^2 = .01, F(4, 151) = 1.33, p = .262). The findings remain non-significant even with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .031.

Multiple regression indicated that the overall model explained 9.2% of the variance (R^2 = .09, F(4, 151) = 4.93, p < .001). It was found that parental education significantly predicted accuracy on the RAI Quantity subscale(β= .03, p < .001, *95% CI* = .01, .05) while the other variables did not. The findings remain significant even with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .006.

Multiple regression indicated that the overall model explained 3.7% of the variance (R^2 = .04, F(4, 151) = 2.47, p = .047). It was found that parental education significantly predicted accuracy on the RAI Temporal subscale (β= .02, p = .036, *95% CI* = .00, .04) while the other variables did not. The findings do not remain significant with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .025.

Multiple regression indicated that the overall model explained 5.5% of the variance (R^2 = .06, F(4, 151) = 3.27, p = .013). It was found that parental education significantly predicted accuracy on the RAI Containment subscale(β= .03, p = .004, *95% CI* = .01, .05) while the other variables did not. The findings do not remain significant with the Bonferroni-adjusted α level of .006. However, they do remain significant with the Benjamini-Hochberg critical value of .013.

Multiple regression indicated that the overall model did not significantly explain the variance in accuracy on the RAI Analogy subscale (R^2 = -.01, F(4, 151) = .76, p = .556). The findings remain non-significant even with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .05.

Multiple regression indicated that the overall model did not significantly explain the variance in accuracy on the RAI Deictic subscale (R^2 = .00, F(4, 151) = .84, p = .502). The findings remain non-significant even with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .044.

Multiple regression indicated that the overall model explained 4.2% of the variance (R^2 = .04, F(4, 151) = 2.71, p = .033). It was found that parental education significantly predicted accuracy on the RAI Mathematical subscale (β= .02, p = .040, *95% CI* = .00, .04) while the other variables did not. The findings do not remain significant with the Bonferroni-adjusted α level of .006 and the Benjamini-Hochberg critical value of .019.

3.3.2.4 RQ3. SES as a Predictor of RAI, RSPM, and Individual RAI Subscale Fluency

Coefficients for insignificant multiple regression models can be found in Appendix F. To reduce the increased risk of a Type I error, which is due to the number of statistical tests performed, resulting in a FWER of $1 – (1 - .05)^{11} = .43$, the following analyses will also be inspected with a Bonferroni-corrected α level of 0.5/11 = .005, which brings the FWER back to .05. Again, to protect against an inflated Type II error risk (Vanderweele & Mathur, 2019), a Benjamini-Hochberg correction was performed and the results will be reported and interpreted accordingly.

Multiple regression indicated that the overall model explained 8.1% of the variance in RAI fluency ($R^2 = .08$, $F(4, 151) = 4.42$, $p = .002$). It was found that parental education significantly predicted RAI Fluency ($\beta = .31$, $p < .001$, *95% CI* = .15, .52) while the other variables did not. The findings remain significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .005.

Multiple regression indicated that the overall model did not significantly explain the variance in RSPM Fluency ($R^2 = .03$, $F(4, 150) = 1.09$, $p = .362$). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .036.

Multiple regression indicated that the overall model explained 8.4% of the variance in RAI-M fluency ($R^2 = .08$, $F(4, 151) = 4.54$, $p = .002$). It was found that parental education significantly predicted RAI-M Fluency ($\beta = .30$, $p < .001$, *95% CI* = .16, .54) while the other variables did not. The findings remain significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .009.

Multiple regression indicated that the overall model explained 4.7% of the variance in fluency on the opposition subscale (R^2 = .05, F(4, 151) = 2.92, p = .023). It was found that parental education significantly predicted fluency on the RAI Opposition subscale (β= .24, p = .007, *95% CI* = .08, .50) while the other variables did not. The findings fail to remain significant with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .018.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Difference subscale (R^2 = -.02, F(4, 151) = .43, p = .785). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .045.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Quantity subscale (R^2 = .01, F(4, 151) = 1.52, p = .199). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .027.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Temporal subscale (R^2 = .03, F(4, 151) = 2.17, p = .075). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .023.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Containment subscale (R^2 = .00, F(4, 151) = .1.14, p = .342). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .032.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Analogy subscale (R^2 = -.02, F(4, 151) = .27, p = .897). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .05.

Multiple regression indicated that the overall model did not significantly explain the variance in fluency on the RAI Deictic subscale (R^2 = -.01, F(4, 151) = .74 p = .569). The findings remain non-significant even with the Bonferroni-adjusted α level of .005 and the Benjamini-Hochberg critical value of .041.

Multiple regression indicated that the overall model explained 6.1% of the variance in fluency on the mathematical subscale (R^2 = .06, F(4, 151) = 3.53, p = .009). It was found that own income (β= 1.09, p = .002, *95% CI* = .42, 1.76) and parental income (β= -.94, p = .019, *95% CI* = -1.72, -.16) predicted fluency on the RAI Mathematical subscale, while the other variables did not significantly predict fluency on the RAI Mathematical subscale. The findings fail to remain significant with the Bonferroni-adjusted α level of .005. However, they do remain significant with the Benjamini-Hochberg critical value of .014.

3.3.2.5 RQ5. RAI Factor Structure

A Confirmatory Factor Analysis was run in Mplus. Cross-loadings were not allowed and variances were not fixed. Instead, Mplus fixes the factor loading of the first indicator to one by default, while the other loadings and factor variance are freely estimated. Fit statistics are displayed in Table 8, and desired values, as proposed by Kline (2005) and Hooper et al. (2008), are shown for readers unfamiliar with factor analyses. The results show the current eight-factor model to not be a good fit for the data. However, relative to the one-factor model, the eight-factor model was a better fit based on almost all parameters. Tables 9 and 10 show factor loadings and correlations, respectively.

Table 8

Confirmatory Factor Analysis Model Fit Statistics of the RAI

	χ^2	*df*	*p*	CFI	TLI	RMSEA (90% CI)	SRMR	AIC	BIC	ssaBIC
Desired Values			>.05	≥ .90	≥ .95	<.08	<.08	Smaller IC desired (relative to alternative model)		
Eight-factor Model	12551.70	7972	< .001	.591	.583	.032 (.031, .033)	.056	79851	81644	80336
One-factor Model	13429.76	8000	< .001	.515	.507	.034 (.033, .035)	.055	80804	82475	81256

Note: N = 573; Estimator = MLR (Robust Maximum Likelihood), χ^2 = Chi-square Goodness of Fit statistic; *df* = degrees of freedom; *p* = Statistical significance; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA (90% CI) = Root-Mean-Square Error of Approximation with 90% confidence intervals; SRMR = Standardised Root-Mean Square Residual; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; ssaBIC = sample size adjusted BIC

The Confirmatory Factor Analysis found the eight-factor model to be a poor fit for the current data, suggesting that the data do not fit based on the topographical features, after which the RAI is modelled. That is, the performance across all of the trials of the test is not easily parsed in terms of relational frame labels as they would have expected to be given their classification within RFT. Almost all factor loadings (see Table 9) were significant, though some in the negative direction, suggesting a poor model fit. In terms of factor correlations displayed in Table 10, all factors significantly and positively correlated with each other, ranging from .37 (difference and deictic) to .93 (quantity and temporal). However, despite the poor fit of the eight-factor model, it was found that the eight-factor model fit the data slightly more accurately than the one-factor model.

Table 9

Standardised Factor Loadings (Standard Errors) for the Eight-Factor Model of the RAI

	Opposition	Difference	Quantity	Temporal	Containment	Analogy	Deictic	Mathematical
Item 1	.61** (.04)	.44** (.05)	.32** (.05)	.35** (.04)	.29** (.06)	.33** (.05)	.26** (.07)	.24** (.05)
Item 2	.29** (.05)	.43** (.05)	.34** (.04)	.38** (.04)	.26** (.05)	.44** (.05)	.25** (.07)	.28** (.05)
Item 3	.17** (.05)	-.02 (.05)	.38** (.04)	.38** (.04)	.38** (.06)	.37** (.05)	.46** (.05)	.29** (.05)
Item 4	.44** (.04)	.47** (.05)	.37** (.04)	.36** (.04)	.40** (.06)	.46** (.05)	.19* (.06)	.32** (.05)
Item 5	.43** (.04)	.42** (.05)	.33** (.04)	.48** (.04)	.48** (.06)	.27** (.05)	.25** (.06)	.03 (.05)
Item 6	.26** (.05)	.52** (.04)	.40** (.04)	.35** (.04)	.28** (.05)	.50** (.04)	.43** (.04)	.13* (.05)
Item 7	.20** (.05)	.43** (.05)	.41** (.04)	.32** (.04)	.37** (.05)	.62** (.04)	.11* (.06)	.17* (.05)
Item 8	.44** (.04)	.47** (.05)	.32** (.04)	.33** (.04)	.38** (.05)	.64** (.04)	.42** (.06)	.24** (.05)
Item 9	.48** (.04)	.50** (.05)	.34** (.05)	.42** (.04)	.36** (.05)	-.46** (.04)	.46** (.06)	.45** (.04)
Item 10	.27** (.05)	.51** (.05)	.40** (.04)	.36** (.04)	.37** (.05)	.44** (.05)	.38** (.06)	.41** (.04)
Item 11	-.24** (.05)	-.13** (.05)	.34** (.04)	.40** (.04)	.36** (.06)	.09 (.07)	.44** (.05)	.16* (.06)
Item 12	-.41** (.04)	.52** (.04)	.40** (.04)	.40** (.04)	.45** (.05)	.47** (.05)	.23** (.05)	.20** (.06)
Item 13	.50** (.04)	.50** (.05)	.29** (.04)	.46** (.04)	.48** (.04)	-.06 (.06)	.17* (.06)	.41** (.04)
Item 14	.29** (.05)	.50** (.05)	.33** (.04)	.37** (.04)	.37** (.06)	.07 (.06)	.50** (.06)	.40** (.04)
Item 15	-.53** (.04)	.59** (.04)	.39** (.04)	.36** (.04)	.30** (.06)	-.41** (.05)	.21* (.06)	.32** (.05)
Item 16	-.27** (.05)	.56** (.04)	.36** (.04)	.41** (.04)	.38** (.05)	-.17* (.06)	.43** (.05)	.33** (.05)

Note. $N = 573$.

**. Correlation is significant at the .01 level.

*. Correlation is significant at the .05 level.

Table 10

Factor Correlations (Standard Errors) for the Eight-Factor Model of the RAI

	Opposition	Difference	Quantity	Temporal	Containment	Analogy	Deictic
Difference	.71 (.04)	–	–	–	–	–	–
Quantity	.77 (.04)	.76 (.04)	–	–	–	–	–
Temporal	.78 (.04)	.67 (.04)	.93 (.03)	–	–	–	–
Containment	.73 (.05)	.59 (.06)	.85 (.04)	.80 (.04)	–	–	–
Analogy	.61 (.05)	.66 (.05)	.76 (.04)	.68 (.05)	.68 (.05)	–	–
Deictic	.43 (.07)	.37 (.06)	.51 (.07)	.51 (.07)	.42 (.07)	.46 (.07)	–
Mathematical	.56 (.07)	.58 (.05)	.84 (.05)	.75 (.05)	.76 (.05)	.70 (.05)	.54 (.07)

Note. N = 573. All effects are statistically significant at the .001 level.

3.4 Discussion

3.4.1 Summary

Research question 1 aimed to investigate the correlation between individual RAI subscales and RSPM performance. It was found that all subscales except for the opposition subscale significantly correlated with RSPM performance. The specific implications will be discussed in section 3.4.2. Research question 2 aimed to investigate which RAI subscales would be significantly predicted by SES. It was found that the quantity, temporal, containment, and mathematical subscales were all significantly predicted by the four SES variables. After Bonferroni and Benjamini-Hochberg corrections the four SES variables no longer significantly predicted accuracy on the temporal subscale and fluency on the containment subscale was still significantly predicted by the four SES subscales according to the Benjamini-Hochberg but not the Bonferroni correction. Parental education was the sole significant predictor in all analyses, except for the mathematical subscale, where own and parental income were the significant predictors. No significant correlations were found for SES and the opposition, difference, analogy, and deictic subscales. This will be discussed in detail in section 3.4.3.

The third research question aimed to investigated SES as a predictor of fluency on the full RAI and its individual subscales, as well as RSPM. While the four SES variables significantly predicted performance on the overall RAI, RAI-M, as well as the opposition and mathematical subscales. Performance on the RSPM, difference, quantity, temporal, containment, analogy, and deictic subscales was not significantly predicted by the four SES variables. After Bonferroni and Benjamini-Hochberg corrections the four SES variables no longer significantly predicted fluency on the opposition subscale and fluency on the mathematical subscale was still significantly predicted by the four SES variables according to the Benjamini-Hochberg but not the Bonferroni correction. This will also be discussed in detail in section 3.4.3.

Finally, the fourth research question analysed the RAI's factor structure and found the current composition of the RAI to not be a good fit for the data tested in the current study. However, the eight-factor model was found to be a slightly better fit than the one-factor model. The implications of this will be discussed in section 3.4.4.

3.4.2 Convergence Between the RAI Subscales and RSPM

As mentioned in section 2.4.2, an alternative explanation for why RAI scores do not fully predict RSPM scores might be that there are essential relational frames, relevant to performance on RSPM, which are not yet integrated into the RAI. The possibility that there might be relational frames relevant to RSPM performance which are not yet currently assessed by the RAI is supported by the current study's finding that individual subscales contributed differently to RSPM performance. One should, of course, be careful with the interpretation of the verb "contributed", as the current thesis only assessed correlations, but RFT would argue for a causal link. In other words, the finding highlights that a combination of various frames is the best predictor of RSPM performance. More specifically, the current thesis found all individual subscale scores to significantly correlate with RSPM performance. However, there were substantial differences in the strength of those relationships. To illustrate, the temporal and mathematical subscales (40% each) accounted for five times the level of covariance as the opposition and analogy subscales (8% each). Quantity accounted for 31% of the covariance and thus was followed by containment (22%), which in turn was closely followed by the difference subscale (21%). Finally, the deictic subscale only accounted for 10% of the covariance in RSPM performance. Indeed, this finding alone has multiple implications. One such implication is that different intellectual activities and intellectual ability tests require proficiency in different relational skills and will be explored in more detail below.

The difference in the various correlations between RSPM and RAI subscale scores supports the assertion that different relational skills are drawn upon in different intellectual activities and across different intellectual ability tests. This stands to reason because many gold standard tests of intelligence have formed their content based on the factor structure of the test as it has evolved over the years rather than based on basic behavioural processes. Given that, for example, psychometricians do not entertain the idea of a relational frame of dialectic relational responding, let alone acknowledge its relevance to a whole host of intellectual skills, there would be no reason to make sure that test items adequately draw upon this skill set as part of the overall intelligence assessment. However, in the development of a range of questions to assess the traditional factors such as numerical reasoning, verbal reasoning, and so on, the whole gamut of relational frame skills might be drawn upon to a greater or lesser extent across various tests purely fortuitously. As a result, various correlations between the RAI and various measures of intelligence would be expected. Indeed, as discussed earlier (section 1.5.3), certain IQ test items can be conceived of as direct measures of relational responding based only upon one or two relational frames. WAIS Vocabulary and Information items, for instance, require the solving of relational tasks based upon opposition relations, whereas the Similarities subtest assesses proficiency of opposition and distinction relational responding (Colbert, 2015). Indeed, some subscales appear to be better proxy measures of intellectual performance than intentionally devised short-form measures of IQ. For instance, the Oklahoma Premorbid Intelligence Estimate (OPIE, Schoenberg et al., 2002), a short version of the WAIS-III, was found to correlate with RSPM with a correlation coefficient of only .35 (Spinks et al., 2009).

The different correlation coefficients obtained between the individual RAI subscales might suggest that performance on RSPM depends on the test-takers repertoire in the various relational responding domains. Indeed, this interpretation tallies with the observed statistical importance of performance on both temporal and mathematical relational responding for predicting RSPM scores and the relatively low importance of opposition, analogy, and deictic subscale scores in this regard. Interestingly, the finding here that temporal relational skills are predictive of performance on various types of intelligence assessments is in line with the findings of Colbert (2015) and O'Hora (2008), who both found this skill to predict Full, Verbal, and Performance IQ. However, the *relative importance* of temporal relations in predicting intellectual skills compared to other relational framing skills does not dovetail well with the findings of Colbert (2015). That is, while Colbert (2015) found that responding in accordance with a relational frame of opposition was a very good predictor of performance on intelligence tests, the current study found the opposition subscale to be one of the weakest predictors of RSPM performance. One possible explanation for this is that the intelligence measures used in Colbert (2015) are more highly reliant on verbal skills than the RSPM. The fact that the RSPM is conceived of as a nonverbal test of intelligence bears this out and further illustrates how convergences between various measures of intelligence are likely themselves the product of third variables rather than bona fide approximations of the same construct. Incidentally, that third variable could be relational responding fluency. Future research might examine this issue more extensively by administering a wide battery of intelligence tests across a large sample of participants and finding in a regression analysis if relational skills are indeed the best predictor of performance on all other tests individually and collectively. This would be the near-perfect study to address the RFT assertion that all intellectual skills are built upon the foundation of fluent derived relational responding.

Regardless of the explanation as to why differences in subscale correlations were found across studies using different measures, and why different relational frame assessment subscales differentially predicted RSPM performance in the current study, the observation of this very variance is a key insight to the fact that however measured, real-world intelligence measures will likely draw upon different relational skills differentially if they continue to assess real-world knowledge rather than underlying abstracted relational skills. That is, in the sense that intelligence tests will continue to evolve, and for new ones to be developed, based on the same principles that the factors identified across numerous intelligence tests over the past century are likely the only ones worth assessing, there will inevitably be imbalances within and across tests and how they draw upon various relational skill repertoires. Inevitably, this will also result in constant variation in how relational skill assessments correlate with traditional intelligence tests. All of this serves to illustrate that the divergences in correlations between relational skill assessments and gold standard tests is not necessarily due to failings within the relational assessment approach but may even point to failings within the approach taken thus far in the development of gold standard tests, which appeared to assess crucial relational skills indirectly, haphazardly and disproportionately. At the same time, however, if the pragmatic concerns of the behaviour analyst are to enhance repertoires that are of use in real-world education and workplaces as they currently stand, it may be strategic to develop interventions such as the SMART intervention, to consciously disproportionately establish fluencies across the various subscales for stated and clear pragmatic ends. It is precisely this agility, made possible by a functional understanding of the behaviours of interest that undermines any effort to arrive at a final definition of what is meant by intelligence, even if that is in purely functional terms. Intelligence is a concept that is more of a burden than an asset within the field. As the current argument shows, it leads only to speculation about the bases for inter-correlations rather than getting to the root of the precise skills that need to be trained and assessed in the most efficient way possible.

One interesting observation within the current data set relates to the varying baseline levels across the various relational skills subsets within and across participants. More specifically, one might assume that high scores on one particular subscale are indicative of near-perfect performance and, therefore, the lack of a need for further intervention to enhance the relevant relational skill. However, it is important to understand that these subscales, or the overall test for that matter, are not standardised in the psychometric sense. Nor is there any need for them ever to be standardised, although such a strategy could be useful if taken on board with precautions protecting against the reification of the construct under assessment.

The high degree of correlation between RAI and RSPM performance provide some insight into the functional understanding of the RSPM itself, thereby highlighting the importance of conducting research which analyses intelligence tests in relational terms as part of the wider effort to develop a fully functional understanding of the phenomenon known as human intelligence. Specifically, both the RAI and RPSM are conceptualised as measures of deductive reasoning ability. On the other hand, one is supposed to be highly culturally independent (RSPM), while the other is supposed to be subject to the test-taker's educational history, which will inevitably vary across the population (RAI). While differences in environmental impact on both measures will be discussed in detail in section 3.4.3, the implications of the observed high degree of correlation across these test types, with differing content and conceptual orientations, raises the interesting question as to how the RSPM could be face valid. To be more specific, the RSPM assessment is widely recognised as a measure of analogical reasoning (e.g., Prade & Richard, 2009, 2010, 2013). However, the current thesis found the analogy subscale of the RAI to be the weakest predictor of RSPM, scores, accounting for only 8% of the variance between the measures. That is assume, for a moment, that the analogy assessment administered here as part of the RAI was respectively valid as a measure of analogical reasoning. Indeed, the tasks administered are of a type common across measures of analogical reasoning and research within the behavioural field (e.g., Stewart et al., 2001, 2013). If this is the case, it strongly suggests that the RSPM is deficient as a measure of analogical reasoning to something approaching the degree assessed by the variance and the correlation between its outcome scores and scores on the RAI analogy subscale. Notwithstanding that, of course, several factors will also contribute to this variance; the important point remains that of all the subscales on the RAI, this is the one that should perhaps have correlated best with overall scores on the RSPM. Of course, it is also important to consider that perhaps the poor correlation found in the current thesis just highlights that the RAI's analogical subscale needs further improvement. Such a conclusion is supported by the fact that the RSPM has been widely validated against other measures of analogical reasoning (e.g., Mekik et al., 2018). Indeed, Cummins (2023) pointed out that the analogical subscale of the RAI is in need of improvement to bring it in line with the standards expected of a test of analogical reasoning. In this particular instance, therefore, we find that perhaps the correlation analysis conducted here provides important information regarding areas in which the RAI can improve methodologically (rather than conceptually).

The finding that performance on the opposition subscale was only weakly related to RSPM performance might attest to the RSPM's claimed cultural independence and its use as a nonverbal measure of intelligence. That is, the frame of opposition is of particular importance to verbal abilities due to its crucial role in word-word and word-object relations. Given that performance on the RSPM was only weakly correlated with proficiency on the opposition subscale, this suggests that perhaps, RSPM are indeed, relatively free of interference from verbal abilities. On the other hand, the RAI is still conceived of as a verbal measure. That is, RFT sees all intellectual skills as involving the verbal process of symbolic or derived relational responding. Despite various topographical differences in how stimuli are related to each other according to different patterns that define the various relational frames, all of the processes, including mathematical skills, are perceived ultimately as verbal skills in functional terms. Thus, it is somewhat of a paradox, as there is at the same time support for the case that the RSPM assessment is nonverbal even according to the findings of the current study, and yet strong correlations overall between performances on the two tests.

O'Hora et al. (2008) suggested a possible explanation for such a correlation between ostensibly verbal and nonverbal tests of intelligence, drawing upon a form of problem-solving, which, within the RFT literature, is conceived of as *pragmatic verbal analysis.* Pragmatic verbal analysis describes a specific form of relational responding whereby arbitrarily applicable relational responding occurs under the control of non-arbitrary physical relations in the real world. To simplify, this means that even when engaging in tasks traditionally interpreted as non-verbal (e.g., placing a round peg in a round hole), the individual still either requires verbal abilities to solve these tasks initially or they respond to the nonverbal stimuli verbally due to an extensive history of derived relational responding to almost everything in their environment (e.g., naming things privately, discriminating arbitrary relations between events related via formal continua such as size or distance). In the current context, this means that individuals solving RSPM tasks may engage in verbal behaviour in the form of identifying the governing relations between the relata in the task and responding accordingly. Indeed, Colbert (2015) further explained that this argument also applies to the Block Design task, another task traditionally conceived as non-verbal. However, it is important to remember that, from an RFT perspective, relational skills are pervasive for verbal beings and events in the environment are likely responded to in terms of derived relations. In other words, the generalised response of arbitrarily applicable relational responding is a feature of verbal behaviour. That is, while it is, in principle, possible to solve some RPSM tasks based purely on spatial relations that are visibly observable along formal continua, it is likely extremely rare for a human individual to respond to any task completely nonverbally. Thus, in an indirect way, the RSPM cannot function as a nonverbal test so long as it is administered to a verbal organism who will always bring a long history of arbitrarily applicable relational responding to such tasks.

3.4.3 SES as a Predictor of RAI Subscale and RSPM Performance and Fluency

Adding to section 2.4.3, SES was found to significantly predict performance (both accuracy and fluency) on the full RAI as well as various subscale scores, but not RSPM. While the specific implications of this were discussed, it is nonetheless important to note that, despite the statistical significance, even in the best cases, SES only accounted for 9.2% of the covariance in performance. It would be inaccurate to automatically equate this statistical significance with practical importance, however (Gelman & Stern, 2006). That is, just because a correlation is statistically significant, that does not mean that the identified relationship is meaningful in the real world. In other words, most of the variants in scores on relational skill fluency are not accountable for by the SES of the participant and their family. At the same time, however, there is a strong probability that the contribution of SES was, in fact, underestimated in the current research due to the limited variance found in SES and RAI scores within the current data set. That being noted, it is important to look at those variables that best predicted overall performance on the RAI. It is interesting therefore to focus attention on the fact that performance on RAI subscales, and RSPM performance, were substantially better at predicting performance on the full RAI, than any aspect of SES. Of course, the issue here was not merely to predict likely performance proficiency on a relational abilities test from the point of view of developing a proxy measure. Rather, the issue was to predict performance on these types of tests based on factors that should be expected to at least somewhat impact performance to a small extent as part of an effort to understand all of the variables that might influence performance and, therefore, understand the relational ability skill set more fully. What might in traditional psychometrics be referred to as a confound or a test bias can, from an RFT perspective, be simply considered explanatory variables.

In summary, it is important not to misinterpret the small impact of socioeconomic status variables on RAI scores, as evidence that most of the performance on a relational test is predicted by factors "within" the individual themselves. This is very tempting from a statistical point of view when other variables in the environment of the individual have been controlled for (i.e., SES). However, variations in performance across individuals, and strong inter-correlations in performances of different types within individuals, can reasonably be accounted for by slight variations in intellectual skills training throughout the lifetimes of the individuals to date, and consistencies in the exposure to various forms of relational skill training within the life of each individual. That is, it is unlikely for an individual to be exposed to an environment in which only one subset of relational skills is trained to a high standard while others are entirely neglected. Thus, strong correlations between various forms of relational responding but, at the same time, degrees of disparity based on the extent to which those particular skills are trained within the culture and the education system, should be expected. Similarly, one should expect to see differences across individuals in their performance on batteries of such relational skills, given differences in SES, and surely also the influence of genetic and physiological variables. The latter, however important in determining final outcomes of educational efforts, are not to be understood as causes within the behavioural paradigm, but as contextual factors that scientists should take an interest in manipulating to as great an extent as possible in the unending effort to gain maximal control over human behaviour.

When looking at the impact that the four SES variables had on RAI performance, the reader is reminded that performance on the mathematics subscale was in fact predicted only by own and parental income levels. This finding is somewhat surprising and the interpretation of this effect is admittedly speculative. One possibility is that mathematics as a school subject stands out as a key subject that students struggle with and require additional tuition (Gonzalez-DeHass et al., 2024). Given that children from economically privileged backgrounds are more likely to be able to afford and receive private tuition for mathematics, it may be the subject that most differentiates privileged from non-privileged children. Any such increased access to additional supports is more likely to be reflected in mathematics assessments before it is reflected in other relational skill tests. While an alternative explanation could be that this finding is based on a Type 1 error due to the multiple correlations run in the current analysis and indeed appeared likely to be the most conservative position one should take, adjusting alpha levels based on the Benjamini-Hochberg correction showed that the finding remained significant. The Bonferroni correction, however, rendered the finding no longer significant. To allow for accurate conclusions, a more detailed investigation into this possible correlation could be an avenue for future studies. Moreover, a less-than-ideal distribution of test scores compromises the reliability and representativeness of the current data set. The same observation was made by Cummins (2023) who linked this data distribution issue to the psychometric properties of the RAI. Given this and a range of issues raised in the previous section regarding data distributions and the meaning of correlations between various measures, it is fitting that attention will now be turned to the psychometric properties of the RAI and how it might be improved.

3.4.3 The RAI's Factor Structure

The final research question investigated in the current thesis related to the RAI's factor structure. No study to date has conducted a factor analysis of the RAI, either in its current nor any previous iteration. The current composition of the RAI consists of eight factors. That is, the subscales are divided based on the relational frames being assessed in the subscales. As outlined in section 3.1.2.4, this was an exploratory analysis, and therefore, there were no *apriori* hypotheses. On the one hand, it might be expected that the factors identified would have corresponded more or less to the eight subscales under analysis. On the other hand, a lack of model fit might not be surprising. This is first and foremost because, unlike traditional intelligence measures, the RAI's development was not guided by factor analyses in the first instance. This is, in fact, a strength of the test and speaks to the fact that the importance of these relational skills under analysis has been established through the empirical and functional analysis of these various skill sets. More particularly, however, the eight various frames were conceived of hierarchically such that during assessments the complexity of the relational responding tasks were arranged from least to most complex using a Gutmann structure. Indeed, this is itself a clue as to why the current eight-factor model might not be a good fit for the most recently developed RAI. That is, the items might load onto factors based on shared complexity of various items across the relational frame subscales, rather than on to the relational frames themselves. For example, all of the least complex task types across all of the frames may constitute a factor that could be identifiable statistically, whereas all of the most complex tasks across all of the frames combined may constitute another distinct factor. And indeed, there may be other currently unknown reasons for why the present composition of the RAI might not have a good model fit. In an effort to make sense of these findings in more detail, the following section will consider more closely the findings of the factor analysis.

Overall, the eight-factor model based on the current composition of the RAI failed to provide a statistically significantly better model fit than a null model. That is, the topographical factor-model does not predict the actual underlying factor structure of the RAI. More specifically, the factor loadings on most items were only within the .30 to .50 range and therefore, not particularly impressive. Indeed, only four subscales, namely quantity, temporal, containment, and deictic had significant and positive factor loadings. The mathematical subscale had one nonsignificant factor loading, while the difference subscale had one nonsignificant as well as one negative factor loading. What particularly stood out, however, were the opposition and analogy subscales loadings. That is, the opposition subscale had four negative factor loadings and the analogy had four negative as well as three insignificant factor loadings, indicating 25% and 33% of problematic items respectively. Interestingly, the latter finding is not surprising as it is in line with those of Cummins' (2023) that the analogy and opposition subscales are insufficient to provide information on the relationship between overall subscale score and the participant's performance at any given level of ability (i.e., poor discriminability). Therefore, the case may be getting clearer that the opposition and analogy subscales in particular, fall short of currently acceptable psychometric standards. This might not necessarily be attributable to poor design, as outlined by Cummins (2023), but more seriously may reflect theoretical flaws in the assumptions of RFT regarding the separability and discriminability of these relational operants. This topic deserves focus in future research because it lies at the very heart of the theoretical position underlying the RAI.

The finding that the overall model fit was not acceptable is also unsurprising. As pointed out above, the composition of the RAI is based on theoretically important and functionally understood differences in response patterns across the different sets of relational frame tasks, rather than based upon statistically guided and psychometrically informed test construction techniques. Again, this is to be applauded insofar as the various relational repertoires are considered important in their own right. It would fly in the face of the behaviour analytic emphasis on bottom-up and inductive research to allow factor analysis to determine how various sets of relational skills are parsed. While such an approach is possible and would surely lead to a more psychometrically coherent test, it would result in a set of constructs that themselves were defined merely psychometrically rather than understood functionally from the ground up in terms of a coherent set of behavioural principles linked to known etymologies. But nevertheless, in an effort to at least understand the degree to which the test in its current form has psychometric integrity, it is important to be aware of the current state of affairs regarding the factor structure.

One caveat, however, is that where a factor analysis model does not model cross-loadings which are actually there, the sensitivity of fit can be impacted considerably (Cao & Lian, 2023). Given the significant correlations between the factors found in both the current thesis and previous studies (e.g., Cummins et al., 2023), such cross-loadings can be assumed, although they must be empirically analysed in order to allow for definitive conclusions. Therefore, one should not prematurely assume that there is no validity whatsoever in the overall factor structure and thus the absence of a fit must be interpreted with this in mind. The need to exercise caution in interpretation of the model fit is further highlighted by the finding that the eight-factor model was found to be a slightly better fit compared to the one-factor model. More specifically, the one-factor model has been a widely replicated finding within the intelligence testing literature, yet was still found to be a poorer fit than the eight-factor model considered here. This highlights, not only that the poor fit for the single factor model might be due to the exclusion of cross-loadings in the factor analysis, but perhaps more importantly, supports RFT's conceptualisation of intelligent behaviour as involving fluency in a series of related but conceptually separable relational operants. That is, the superiority of the RAI's eight-factor model might be a defining feature of a functional account of intelligence based on a range of core intellectual skills that cannot be easily considered as a single operant and yet involves a series of skills that likely share components (i.e., some relational skill repertoires likely build upon others and yet are separable both functionally and conceptually in some ways). Given the crucial theoretical implications of such a conclusion, this issue requires consideration in future conceptual and empirical work.

The current factor analysis is the first ever conducted on the RAI and thus offers valuable insight into the governing factor structure of the measure. Improvements to the RAI can be guided by this factor analysis, which will, in turn, enhance its psychometric integrity. More specifically, improving the RAI based on the suggestions made by Cummins (2023), by such means as increasing the number of response options and making any other adjustments that may increase the spread of performances (e.g., the use of strict response windows), would improve inter-participant discriminability and likely the fit of the test two and eight-factor model or even as a single factor.

The factor analysis conducted here also provided information on factor correlations. Factor loadings and factor correlations can provide valuable information about how strongly the factors correlate with each other statistically speaking and whether a one-factor model might be more appropriate than the RAI's proposed eight-factor model. The current study found all factors (i.e., RAI subscales) to intercorrelate significantly with positive correlation coefficients. What is noticeable is that the opposition, quantity, temporal, containment and mathematical subscales were, in most cases, highly correlated with other RAI subscales. The finding that the opposition subscale was consistently relatively highly correlated with the other subscales supports RFT's assertion that this frame is fundamental to the acquisition of additional relational frames.

Chapter 4

General Discussion

4. Discussion

4.1 Research Summary and Main Findings

The current thesis aimed to investigate a total of seven research questions, which can be further subdivided into three main overarching research questions with individual components. These are: (1) assessing the convergence between the RAI and RSPM (investigated in terms of full RAI as well as its individual subscales and analysed for both accuracy and fluency), (2) assessing SES as a predictor of RAI (full and individual subscales) and RSPM performance in terms of both accuracy and fluency, and finally, (3) assessing whether the RAI's current composition aligns with its underlying factor structure. Results showed strong correlations between RAI and RSPM performance, although there were important differences between the correlations based on which RAI subscales were analysed. This was discussed in more detail in sections 2.4.2 and 3.4.2. Furthermore, the current thesis found SES to generally predict performance and fluency on the full RAI, but not RSPM. This was discussed in more detail in section 2.4.3 and 3.4.3. Additionally, a confirmatory factor analysis found the current eight-factor model of the RAI to not be a good fit for the data analysed in Study 2, but still a better fit than a one-factor model. The implications of this were discussed in detail in section 3.4.4. The following will discuss global considerations, limitations, and implications for future studies in sections 4.2, 4.3, and 4.4 respectively.

4.2 Global Considerations

Overall, performance on the full RAI was found to be significantly correlated with SES, with parental education standing out as the sole significant predictor. While individual subscales were not all significantly correlated with SES, in cases where they were, the importance of parental education was consistently reflected across analyses. This, as discussed above, likely illustrates the importance that the quality of educational interactions children have with their parents at home has for the development of intellectual ability.

The current thesis further found significant correlations between RAI and RSPM performance. As explained above, this suggests that relational responding proficiency indeed underlies intelligent behaviour (at least as measured by standardised tests). However, the correlational nature of the conducted analyses inherently prohibits an investigation into any functional properties of this correlation. Therefore, it is important to mention that, of course, it is possible that rather than relational abilities underlying performance on the RSPM, something that is measures by RSPM is what underlies performance on the RAI. Alternatively, it could also be possible that neither of those options are true and instead, performance is underpinned by some other, perhaps currently unknown or at least currently unidentified, factor. While this is a very real possibility, RFT would argue that it has the more fundamental of these processes identified and therefore allows for the causal relationship in which relational responding proficiency determines performance on measures such as RSPM. This assumption is further supported by the success of the SMART studies presented in Chapter 1, in which training relational responding resulted in substantial IQ increases.

The full RAI was found to have high internal reliability. However, individual subscales showed variance in their internal reliability, with some still showing high correlation coefficients while others performed rather poorly. This demonstrates that the RAI appears to measure one underlying skill but highlights the importance of improving the individual subscales. Consistent with the aforementioned findings, for instance, the analogy subscale performed particularly poorly. Indeed, this finding was reflected across all analyses conducted in the current thesis. That is, performance on the analogy subscale was neither significantly correlated with RSPM performance nor with SES, and it showed the poorest overall factor loadings out of all of the individual subscales. Given the psychometric issues associated with the analogy subscale that were outlined here and in Cummins (2023), it is unsurprising that the analogy subscale fails to significantly correlate with any relevant factors assessed in the current study. The finding that performance on the RAI Analogy subscale correlates poorly with measures of intelligence is not special to the current thesis. More specifically, Colbert et al. (2018) found that the Analogy subscale failed to significantly correlate with Full Scale and Verbal IQ and that the exclusion of this subscale even increased the RAI+'s overall correlation with both of these indices. The authors explained this finding to be particularly surprising as analogy is believed to be a form of higher-order relational responding due to its focus on relations between relations. Perhaps, therefore, it was the case that the reason Analogy was not significantly related to IQ is because traditional IQ tests fall short on assessing analogy. Interestingly, however, Colbert et al. (2018) found the Analogy subscale to significantly correlate with Performance IQ, and indeed, the current study found significant correlations with RSPM. However, importantly, this correlation was one of the weakest out of all the subscales and indeed, performance on the analogy subscale was not related to SES. While this finding needs to be explored in more detail and, indeed, perhaps might be solely a result of the subscale's poor psychometric properties, it might also be the case that the assumptions about what is meant by highly developed relational responding repertoires might have to be revisited given the now consistent findings that this particular relational scale stands apart as unrelated to most other apparently relevant relational skills and measures of relational fluency.

Despite the various nuances in the results outlined to date, an important takeaway message is that the results generally highlight the strong correlation between relational responding and intellectual performance. This not only further supports well-established literature (Barnes, 1994; Barnes et al. 1995; Catania, 1996, 1998; Hayes & Barnes, 1997; Hayes, 1991; Healy et al., 1998; Horne & Lowe, 1996; Lowekron, 1998; Roche & Barnes, 1997; Smeets et al., 1997), but further supports the assertions made by RFT as far back as the first seminal text on the subject (Hayes et al., 2001). Given that certain subscales were found to be more strongly correlated with traditional measures of intelligence than other subscales, both here, and in previous studies (e.g., Colbert et al., 2018), a closer investigation into the connection of specific intellectual abilities and specific relational responding skills is warranted. Colbert (2015), for instance, found the frame of opposition to be the most important subscale to predict intelligence, while the current study found the opposition subscale to be one of the weakest predictors of intelligence. As Colbert (2015) used a verbal IQ measure while the current thesis used a performance IQ measure, this disparity might highlight that certain subscales are more important to verbal IQ than to performance IQ. A study investigating each subscale as it relates to a variety of intelligence test measures would offer valuable insight into the specific relevance of each relational frame to different conceptualisations of intelligence. The findings of such a study could help guide the development of a revised SMART protocol depending on the aims of the scientist. That is, identifying which specific relational frames are of particular importance to intelligence as defined within various models will allow to refine the protocol to relevant frames only. Combined with Cummins et al.'s (2023) approach, in which the inter-correlations between subscales were used to guide training, with the aim to train fewer, but simultaneously more generalisable frames, could lead to the development of a truly efficient relational skills training program that would target performance on particular intelligence tests. Of course, it is also an option for the behaviour analyst to continue to develop the RAI merely in terms of internal psychometric integrity and without respect to its convergent validity with traditional measures precisely so as not to reinvent the wheel. It is important to remember, that it is an important starting point for the current research that the measure of an acceptable assessment of relational ability will not be determined by its convergence with other measures that are topographically dissimilar and conceptually incompatible.

4.6 Limitations of the Current Study

Apart from the methodological limitations already discussed in the respective sections, there are a number of factors that may limit the reliability and generalisability of the current findings. One very obvious limitation relates to the sample of participants employed. While the sample was sufficient in size, it consisted of exclusively twenty- to twenty-four-year-olds, all of which currently reside and have grown up in either Ireland or the UK. This, of course, impacts the conclusions that can be made about theoretical and practical implications in general and caution in doing so is advised. A somewhat smaller, but still relevant, issue has to do with the choice of independent variables that were used to assess SES. More specifically, given that parental income was assessed only as a proxy index of access to education, it would have been more accurate to treat it as a moderator variable, rather than a predictor variable. That is, when assessing the impact of own education, the analyses should have controlled for parental income, instead of assessing the direct impact that parental income has on performance alone. Also important in relation to the SES investigated here is the interpretability of the obtained regression coefficients. An increased likelihood of causal dependencies between the included variables in a multiple regression model makes the individual coefficients difficult to interpret. Given that parental education is very likely to impact parental income and the participants' education, just like own education likely influences own income, this can lead to collider bias. While the current analyses still met the requirement of not exceeding the allowed limit for inter-variable correlations, it is nonetheless an important question to consider in future work.

Another somewhat easily fixable methodological issue facing the current study was that it was difficult to exert good laboratory control over the behaviour of participants during the assessments. The vast majority of participants conducted the study without any experimental supervision. Therefore, performance was subject to a number of factors, such as environmental distractors but also individual motivation and adherence to the provided instructions.

Another limitation relates to the choice of intelligence assessment used here. Specifically, while the strong, significant correlation found between RAI and RSPM performance supports the strong link between relational abilities and intellectual skills, it is important to remember that the measure used as an index for intelligence, RSPM, consists of only one item type. Traditional gold standard IQ tests, however, usually not only consist of a wide variety of item types, but are further split into types of intelligence being assessed; namely verbal and performance IQ. Therefore, an investigation is warranted in order to provide more detailed and arguably more accurate insight into the correlations between relational abilities and intellectual performance. More specifically, such an analysis (e.g., Colbert et al., 2018), would allow for conclusions about which particular IQ test items are best predicted by which relational skills. Investigating these correlations with a particular focus on the individual subscales would additionally enhance the level of detail by providing further insight into the specific frames relevant for particular intellectual skills. While RSPM might be limited to just one item-type, however, it has been argued to be the best estimate of *g* available. Though this argument has been contested (Gignac, 2015), RSPM nonetheless serve as a sufficient estimate of IQ and therefore the current analysis is not necessarily affected by whether or not RSPM are indeed the best correlate of *g*.

4.7 Implications for Future Research

While the current thesis was the first attempt to identify the role of SES in predicting levels of relational abilities, there is room for future studies to investigate this impact more systematically. As every psychologist knows, the insight that correlational analyses can provide into any given mechanism is limited. While correlations are invaluable in first identifying whether or not there is any potentially functional relationship between variables worthy of further analysis, they are by no means sufficient to allow for any final conclusions regarding such functional relationships. Therefore, future research should adopt a more experimental design to study the impact of SES on the acquisition of relational responding proficiency. That is, further intervention work is required that systematically examines the impact of training on individual relational frames or clusters of such frames on the performance on various intelligence tests and their sub tests. Not only should the aforementioned limitations be considered in the assessment of SES, but the specific impact of individual SES factors should be studied in closer detail. Undoubtedly, such an endeavour will require both ample time and effort, but given the crucial importance of its implications, it is certainly a resolution worth the investment. Should a study like that of Hart and Risley (1995) be replicated in the future, the crucial impact of relational responding repertoires to intellectual development that has been established in the past two decades, should undoubtedly be incorporated into the study. That is, when investigating parent-child interactions, attention should be paid into how those interactions guide the establishment of relational responding repertoires. Such a study would offer direct insight into how SES, through its impact on parent-child interactions, predicts the development of a relational responding skills repertoire, which in turn affects intellectual performance.

Given the RAI's verbal nature, an interesting avenue for future research might be to examine how linguistic processes impact performance on the RAI. On the one hand, this could take the form of investigating how the RAI translates into other languages and what impacts that has on performance, as well as possible correlations with other variables. As the RAI has not been translated into other languages, this has not yet been done. Alternatively, though perhaps of less value to the behaviour analyst, an interesting question is whether linguistic deficits, such as non-verbal (or perhaps more precisely, non-vocal) individuals perform on this measure. Similarly, future studies could investigate the impact of anaduralia (i.e., the absence of an internal verbal monologue) or perhaps even aphantasia (the inability to picture things in the mind). Some theorists believe that language itself plays a fundamental role in human cognition (Bowerman & Levinson, 2001; Lupyan, 2015) and that its unique properties, missing from other representational modalities, allow language to become a valuable interface in the manipulation of mental states (Carruthers, 2002; Clark, 1997; Lupyan & Bergen, 2015). More specifically, its categorical and compositional nature allows for categorisation and abstraction, which is impossible through perception, which is restricted to specific objects thus precluding categorisation and abstraction. Therefore, some researchers have vaguely hinted at the idea that certain mental representations augmented by language, might facilitate certain types of reasoning (Baldo et al., 2015; Boutonnet & Lupyan, 2015; Holmes & Wolff, 2013; Lupyan, 2017; Lupyan & Thompson-Schill, 2012). This proposition has, however, not been tested in published studies to date.

4.8 Conclusion

The current thesis suggests an important role for parental education in the development of relational responding. Combined with the strong link found between relational responding proficiency and intellectual ability, this finding highlights the crucial impact that early experiences likely have on intellectual development. However, the thesis additionally points to some of the RAI's psychometric shortcomings, which might compromise our interpretation of several of the more nuanced findings. Nevertheless, the current findings provide some excellent starting points for improvement of the RAI and provide a springboard for the development of more effective and efficient interventions that will not only further extend our knowledge of relational skills and how best to assess them but help the behaviour analyst to make further inroads to the functional analysis of the previously amorphous concept "intelligence".

References

Abrami, P. C., Cholmsky, P., & Gordon, R. (2001). *Statistical analysis for the social sciences: An interactive approach*. Allyn & Bacon.

Aciego, R., García, L., & Betancort, M. (2012). The benefits of chess for the intellectual and social-emotional enrichment in schoolchildren. *The Spanish journal of psychology*, *15*(2), 551–559. https://doi.org/10.5209/rev_sjop.2012.v15.n2.38866

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs?. *Psychological Bulletin, 131,* 30– 60.

Aichele, S., Rabbitt, P., & Ghisletta, P. (2015). Life span decrements in fluid intelligence and processing speed predict mortality risk. *Psychology and Aging, 30*(3), 598–612. https://doi.org/10.1037/pag0000035

Alloway, T. (2012). Learning and fluid intelligence across the life span. In N. M. (Eds.), *Encyclopedia of the Sciences of Learning* (pp. 1788-1790). Springer. https://doi.org/10.1007/978-1-4419-1428-6_1780

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature neuroscience*, *2*(11), 1032–1037. https://doi.org/10.1038/14833

Andrews, E., & Croal, P. (2023, January 25). *Average household income, UK: Financial year ending 2022*. Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/i ncomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending 2022

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development, 13,* 479-513.

Angelsen, N., Vik, T., Jacobsen, G., & Bakketeig, L. (2001). Breast feeding and cognitive development at age 1 and 5 years. *Archives of Disease in Childhood, 85*(3), 183-188. https://doi.org/10.1136%2Fadc.85.3.183

Aron, A., & Aron, E. N. (2003). *Statistics for psychology* (3rd ed.). Prentice- Hall.

Axelrod, B. N. (2002). Validity of the Wechsler Abbreviated Scale of Intelligence and other very short forms of estimating intellectual functioning. Assessment, 9, 17–23.

Baldo, J. V., Paulraj, S. R., Curran, B. C., & Dronkers, N. F. (2015). Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay. *Frontiers in Psychology, 6*, Article 1523. https://doi.org/10.3389/fpsyg.2015.01523

Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs, 23,* 191–234.

Ballesteros, S., Mayas, J., Prieto, A., Barrera, P. T., González, C. P., De León, L. P., Reales, J. M., & Waterworth, J. A. (2015). A randomized controlled trial of brain training with non-action video games in older adults: Results of the 3-month follow-up. *Frontiers in Aging Neuroscience*, *7*(MAR). https://doi.org/10.3389/fnagi.2015.00045

Ballesteros, S., Prieto, A., Mayas, J., Toril, P., Pita, C., de León, L. P., Reales, J. M., & Waterworth, J. (2014). Brain training with non-action video games enhances aspects of cognition in older adults: A randomized controlled trial. *Frontiers in Aging Neuroscience*, *6*(OCT). https://doi.org/10.3389/fnagi.2014.00277

Barnes-Holmes, D., Regan, D., Barnes-Holmes, Y., Commins, S., Walsh, D., & Stewart, I. et al. (2005a). Relating Derived Relations as a Model of Analogical Reasoning: Reaction Times and Event-Related Potentials. *Journal of Experimental Analysis of Behaviour*, *84*(3), 435-451. http://dx.doi.org/10.1901/jeab.2005.79-04

Barnes-Holmes, D., Staunton, C., Whelan, R., Barnes-Holmes, Y., Commins, S., & Walsh, D. et al. (2005b). Derived Stimulus Relations, Semantic Priming, and Event-Related Potentials: Testing a Behavioral Theory of Semantic Networks. *Journal of Experimental Analysis of Behaviour*, *84*(3), 417-433. http://dx.doi.org/10.1901/jeab.2005.78-04

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. (2001a). The development of self and perspective-taking: A relational frame analysis. *Behavioural Development Bulletin*, *10*(1), 42-45. http://dx.doi.org/10.1037/h0100482

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001b). Exemplar training and a derived transformation of function in accordance with symmetry. *Psychological Record*, *51*(2), 287–308. https://doi.org/10.1007/BF03395400

Barnes-Holmes, Y., Barnes-Holmes, D., Smeets, P. M., Strand, P., & Friman, P. (2004). Establishing relational responding in accordance with more-than and less-than as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy*, *4*, 531-558.

Barnes-Holmes, D., Regan, D., Barnes-Holmes, Y., Commins, S., Walsh, D., Stewart, I., Smeets, P., Whelan, R., & Dymond, S. (2005). Relating derived relations as a model of analogical reasoning: Reaction times and event-related potentials. *Journal of the Experimental Analysis of Behavior*, *84*(3), 435-451.

Barnes, C. A., McNaughton, B. L., Mizumori, S. J., Leonard, B. W., & Lin, L. H. (1990). Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in brain research*, *83*, 287–300. https://doi.org/10.1016/s0079-6123(08)61257-1

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record, 44,* 91-124.

Barnes, D., Browne, M., Smeets, P., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three-to six-year-old children. *The Psychological Record*, *45*, 405-430.

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002). Genetic and environmental influences on the development of intelligence. *Behavior Genetics, 32*(4), 237–249. https://doi.org/10.1023/A:1019772628912

Basso, A., De Renzi, E., Faglioni, P., Scotti, G., & Spinnler, H. (1973). Neuropsychological evidence for the existence of cerebral areas critical to the performance of intelligence tasks. Brain, 96(4), 715–728. https://doi.org/10.1093/brain/96.4.715

Bates, B. T., Zhang, S., Dufek, J. S., & Chen, F. C. (1996). The effects of sample size and variability on the correlation coefficient. *Medicine and Science in Sports and Exercise*, *28*(3), 386–391. https://doi.org/10.1097/00005768-199603000-00015

Beck, C., Garcia, Y., Brothers, L., Mahoney, A., Rancourt, R. C., & Andrews, M. (2023). A systematic review of the impact of derived relational responding technology in raising intelligence scores. *The Psychological Record, 73*(3), 339–361. https://doi.org/10.1007/s40732-023-00546-0

Beckwith, J. (1999). Simplicity and complexity: Is IQ ready for genetics? *Current Psychology of Cognition, 18*, 161-169.

Bell, P., Staines, P., & Michell, J. (2001). *Logical psych: Reasoning, explanation & writing in psychology.* UNSW Press.

Bellinger, D. C., Stiles, K. M., & Needleman, H. L. (1992). Low-level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics*, *90*(6), 855-861.

Bennett, G. K., Seashore, H. G., & Westman, A. G. (1990). *The Differential Aptitude Test (5th ed.).* Psychological Corporation.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music makes the world go round: The impact of musical training on non-musical cognitive functions - a review. *Frontiers in Psychology*, *January 7*, 2023. https://doi.org/10.3389/fpsyg.2015.02023

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily Applicable Comparative Relations: Experimental Evidence for a Relational Operant. *Journal of Applied Behavior Analysis*, *40*(1), 45–71. https://doi.org/10.1901/jaba.2007.7-06

Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00926

Bialystok, E., & Senman, L. (2004). Executive Processes in Appearance-Reality Tasks: The Role of Inhibition of Attention and Symbolic Representation. *Child Development, 75*(2), 562–579. https://doi.org/10.1111/j.1467-8624.2004.00693.x

Binder, C. (1993). Behavioral Fluency: A New Paradigm. *Educational Technology*, *33*(10), 8–14. http://www.jstor.org/stable/44428106

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*(2), 163–197.

Binet, A. (1904a). Nos commissions de travail [Our working committee]. *Bulletin de la Société Libre pour l'Etude Psychologique de l'Enfant*, *4*(14), 337–346.

Binet, A. (1904b). Sommaire des travaux en cours à la Société de psychologie de l'enfant [Summary of the work in progress at the society of child psychology]. *L'Année Psychologique*, *10*, 116–130.

Binet, A., & Henri, V. (1895). La mémoire des phrases [Memory phrases]. *L'Année Psychologique*, *1*, 24–59.

Binet, A., & Simon, T. (1905). Sur le necessite d'etablit un diagnostic scientifique des etats inferieurs de l'intelligence. (On the necessity of establishing a scientific diagnosis of inferior states of intelligence.). *L'Année Psychologique*, *11*, 163–190.

Binet, A., & Simon, T. (1916). New investigation upon the measure of the intellectual level among school children. (*L'Année Psych.*, 1911, pp. 145-201). In A. Binet, T. Simon & E. S. Kite (Trans.), *The development of intelligence in children (The Binet-Simon Scale)* (pp. 274–329). Williams & Wilkins Co. https://doi.org/10.1037/11069-005

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

Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence, 33*(1), 93–106. https://doi.org/10.1016/j.intell.2004.07.008

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65,* 60–66. https://doi.org/10.1016/j.intell.2017.07.004

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*(3), 383-405. https://doi.org/10.1076/jcen.24.3.383.981

Boring, E. G. (1961). Intelligence as the Tests Test It. In J. J. Jenkins & D. G. Paterson (Eds.), *Studies in individual differences: The search for intelligence* (pp. 210–214). Appleton-Century-Crofts. https://doi.org/10.1037/11491-017

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience, 35*(25), 9329-9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Bowerman, M., & Levinson, S. C. (2001). *Language acquisition and conceptual development*. Cambridge University Press.

Boyle, S., Roche, B., Dymond, S., & Hermans, D. (2016). Generalisation of fear and avoidance along a semantic continuum. *Cognition and Emotion*, *30*(2), 340–352. https://doi.org/10.1080/02699931.2014.1000831

Brand, C. (1987). Bryter still and bryter? *Nature, 328*(6126), 110. https://doi.org/10.1038/328110a0

Brans, R. G., Kahn, R. S., Schnack, H. G., van Baal, G. C., Posthuma, D., van Haren, N. E., Lepage, C., Lerch, J. P., Collins, D. L., Evans, A. C., Boomsma, D. I., & Hulshoff Pol, H. E. (2010). Brain plasticity and intellectual ability are influenced by shared genes. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *30*(16), 5519–5524. https://doi.org/10.1523/JNEUROSCI.5841-09.2010

Brant, A. M., Haberstick, B. C., Corley, R. P., Wadsworth, S. J., DeFries, J. C., & Hewitt, J. K. (2009). The developmental etiology of high IQ. *Behavior genetics*, *39*(4), 393–405. https://doi.org/10.1007/s10519-009-9268-x

Brant, A. M., Munakata, Y., Boomsma, D. I., DeFries, J. C., Haworth, C. M., Keller, M. C., ... & Hewitt, J. K. (2013). The nature and nurture of high IQ: an extended sensitive period for intellectual development. *Psychological science*, *24*(8), 1487-1495.

Brant, A. M., Munakata, Y., Boomsma, D. I., Defries, J. C., Haworth, C. M., Keller, M. C., Martin, N. G., McGue, M., Petrill, S. A., Plomin, R., Wadsworth, S. J., Wright, M. J., & Hewitt, J. K. (2013). The nature and nurture of high IQ: an extended sensitive period for intellectual development. *Psychological science*, *24*(8), 1487–1495. https://doi.org/10.1177/0956797612473119

Brinch, C. N., & Galloway, T. A. (2012). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences*, *109*(2), 425–430. https://doi.org/10.1073/pnas.1106077109

Broadbent, D. E., Cooper, P. F., Fitzgerald, P., & Parkes, K. R. (1982). The cognitive failures questionnaire and its correlates. British Journal of Clinical Psychology, 21, 1–16.

Brody, N. (1992). *Intelligence* (2nd ed.). Academic Press.

Brown, K. J., Sinatra, G. M., & Wagstaff, J. M. (1996). Exploring the Potential of Analogy Instruction to Support Students' Spelling Development. *The Elementary School Journal*, *97*(1), 81–99. https://doi.org/10.1086/461850

Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *The Journal of Genetic Psychology, 93*(2), 199–228.

Burt, C. (1963). Is intelligence distributed normally?. *British Journal of Statistical Psychology*, *16*(2), 175-190.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376. https://doi.org/10.1038/nrn3475

Caemmerer, J. M., Keith, T. Z., & Reynolds, M. R. (2020). Beyond individual intelligence tests: Application of Cattell-Horn-Carroll theory. *Intelligence, 79,* Article 101433. https://doi.org/10.1016/j.intell.2020.101433

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's Reading Comprehension Ability: Concurrent Prediction by Working Memory, Verbal Ability, and Component Skills. *Journal of Educational Psychology, 96*(1), 31-42. https://doi.org/10.1037/0022-0663.96.1.31

Camara, W. J., Nathan, J. S., & Puente, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31*(2), 141–154. https://doi.org/10.1037/0735-7028.31.2.141

Cao, C., & Liang, X. (2023). The Impact of Ignoring Cross-loadings on the Sensitivity of Fit Measures in Measurement Invariance Testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *31*(1), 64-80.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246–251. https://doi.org/10.1016/j.lindif.2008.10.002

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Carruthers, P. (2002). The cognitive functions of language. *Behavioural and Brain Sciences, 25*(6), 657-674. https://doi.org/10.1017/s0140525x02000122

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record, 61*(2), 173–198. https://doi.org/10.1007/BF03395755

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational Frame Theory and Human Intelligence. *European Journal of Behaviour Analysis, 11*(1), 37-51. https://doi.org/10.1080/15021149.2010.11434333

Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. M. (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences, 47,* 222–235. https://doi.org/10.1016/j.lindif.2016.03.001

Catania, A. C. (1996). On the origins of behaviour structure. In T. R. Zentall & P. M. Smeets (Eds.) *Stimulus class formation in humans and animals* (pp. 3- 12). North Holland: Elsevier Science.

Catania, A. C. (1998). *Learning* (4th ed.). Prentice-Hall.

Catron, D. W., & Thompson, C. C. (1979). Test–retest gains in WAIS scores after four retest intervals. *Journal of Clinical Psychology, 35*, 352–357.

Cattell, R. (1971). *Abilities*. Houghton Mifflin.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*(3), 153–193. https://doi.org/10.1037/h0059973

Ceci, S. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, *27*(5), 703-722. http://dx.doi.org/10.1037/0012-1649.27.5.703

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*(5), 703–722. https://doi.org/10.1037/0012-1649.27.5.703

Ceci, S. J. (1996). On intelligence: A bio-ecological treatise on intellectual development. Cambridge, MA: Harvard University Press.

Central Statistics Office. (2017, June 21). *Household Budget Survey 2015-2016*. Central Statistics Office. https://www.cso.ie/en/releasesandpublications/ep/p-hbs/hbs20152016/hinc/

Central Statistics Office. (2024, March 7). *Survey on Income and Living Conditions (SILC) 2023)*. Central Statistics Office. https://www.cso.ie/en/releasesandpublications/ep/p-silc/surveyonincomeandlivingconditionssilc2023/householdincome/

Cheramie, G. M., Stafford, M. E., Boysen, C., Moore, J., & Prade, C. (2012). Relationship between the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV) and Woodcock-Johnson-III Normative Update (NU): Tests of Cognitive Abilities (WJ-III: COG). Journal of Education and Human Development, 5(1), 1–9.

Chomsky, N. (1959). Review of Skinner's Verbal Behaviour. *Language, 35,* 26-58.

Clark, A. (1997). Being there: Putting brain, body, and world together again.: MIT Press.

Colbert, D. (2015). Assessing the efficacy of a relational skills training intervention in improving intellectual function in a sample of high IQ adults. [Master's thesis, National University of Ireland, Maynooth]. Maynooth University Research Archive Library. https://mural.maynoothuniversity.ie/7539/

Colbert, D., Dobutowitsch, M., Roche, B., & Brophy, C. (2017). The proxy-measurement of intelligence quotients using a relational skills abilities index. *Learning and Individual Differences, 57,* 114–122. https://doi.org/10.1016/j.lindif.2017.03.010

Colbert, D., Malone, A., Barrett, S., & Roche, B. (2020). The Relational Abilities Index+: Initial Validation of a Functionally Understood Proxy Measure for Intelligence. *Perspectives on Behaviour Science*, *43*, 189-213. https://doi.org/10.1007/s40614-019-00197-z

Colbert, D., Tyndall, I., Roche, B., & Cassidy, S. (2018). Can SMART training really increase intelligence? A replication study. *Journal of Behavioral Education, 27*(4), 509–531. https://doi.org/10.1007/s10864-018-9302-2

Conway, A. R. A., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 233–270). Elsevier Academic Press.

Corben, L. A., Georgiou-Karistianis, N., Fahey, M. C., Storey, E., Churchyard, A., Horne, M., et al. (2006). Towards an understanding of cognitive function in Friedreich ataxia. *Brain Research Bulletin, 70*(3), 197–202.

Costa, A., & Gallés, N. S. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews Neuroscience, 15*(5), 336–345. https://doi.org/10.1038/nrn3709

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*(2), 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

Council of Economic Advisers. (1999). *Families and the labor market, 1969-1999: Analysing the "time crunch".* https://clintonwhitehouse4.archives.gov/media/pdf/famfinal.pdf

Cowan, N., & Alloway, T. (2008). *Development of working memory in childhood* (2nd Ed.) Psychology Press.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In Crawford, J. R., Parker, D. M., & McKinlay, W. (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). Erlbaum.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.

Cummins, J. (2023). On the measurement of relational responding. *Journal of Contextual Behavioral Science*, *30*, 155-168.

Cummins, J., Nevejans, M., Colbert, D., & De Houwer, J. (2023). On the structure of relational responding. *Journal of Contextual Behavioural Science, 27*, 16-25. https://doi.org/10.1016/j.jcbs.2022.11.006

Daneman, M., & Carpenter, P. A. (2004). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*(4), 450–466. https://doi.org/10.1016/s0022-5371(80)90312-6

Day, E. A., Arthur, W., Bell, S. T., Edwards, B. D., Bennett, W., Mendoza, J. L., et al. (2005). Ability-based pairing strategies in the team-based training of a complex skill: Does the intelligence of your training partner matter? *Intelligence, 33*(1), 39–65.

De Rose, J. C., De Souza, D. G., & Hanna, E. S. (1996). Teaching reading and spelling: Exclusion and stimulus equivalence. *Journal of Applied Behaviour Analysis*, *29*(4), 451-469.

Deary, I. J., & Johnson, W. (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *International journal of epidemiology*, *39*(5), 1362–1369. https://doi.org/10.1093/ije/dyq072

Deary, I. J., & Smith, P. (2004). Intelligence research and assessment in the United Kingdom. In R. J. Sternberg (Ed.), *International handbook of intelligence* (pp. 1–48). Cambridge University Press.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Department of Further and Higher Education, Research, Innovation and Science. (3 October 2022). *Minister Harris welcomes publication of OECD report on third level education.* [Press release]. https://www.gov.ie/en/press-release/70777-minister-harris-welcomes-publication-of-oecd-report-on-third-level-education/#:~:text=Ireland%20ranks%20third%20in%20the,the%20OECD%20average%20of%2016.1%25

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior, 46*(3), 243–257. https://doi.org/10.1901/jeab.1986.46-243

Dickens, W., & Flynn, J. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*(2), 346-369. http://dx.doi.org/10.1037/0033-295x.108.2.346

Dixon, M. R. (2014a). *The PEAK relational training system: Direct training module*. Shawnee Scientific Press.

Dixon, M. R. (2014b). *The PEAK relational training system: Generalisation module*. Shawnee Scientific Press.

Dixon, M. R. (2015). *The PEAK relational training system: Equivalence module*. Shawnee Scientific Press.

Dixon, M. R. (2016). *PEAK relational training system: Transformation module*. Shawnee Scientific Press.

Dixon, M. R., Belisle, J., McKeel, A., Whiting, S., Speelman, R., Daar, J. H., & Rowsey, K. (2017). An Internal and Critical Review of the PEAK Relational Training System for Children with Autism and Related Intellectual Disabilities: 2014-2017. *The Behavior analyst*, *40*(2), 493–521. https://doi.org/10.1007/s40614-017-0119-4

Dixon, M. R., Paliliunas, D., Barron, B. F., Schmick, A., & Stan- ley, C. (2021). Randomized controlled trial evaluation of ABA content on IQ gains in children with autism. *Journal of Behavioral Education, 30*, 455–477. https://doi.org/10.1007/ s10864-019-09344-7

Dixon, M. R., Yi, Z., & Chastain, A. N. (2022). PEAK relational train- ing system. In J. B. Leaf, J. H. Cihon, J. L. Ferguson, & M. J. Weiss (Eds.), *Handbook of applied behavior* *analysis interventions for autism: Integrating research into practice* (pp. 341–360). Springer.

Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the discriminative and eliciting functions of generalized relational stimuli. *Journal of the experimental analysis of behavior*, *88*(2), 179-197.

Douglas, J. W. B. *The Home and the School*. 1964. Harper Collins Distribution.

Doumas, L. A., Morrison, R. G., & Richland, L. E. (2018). Individual differences in relational learning and analogical reasoning: A computational model of longitudinal change. *Frontiers in Psychology*, *9*, 1235.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*(5), 613-635.

Dugdale, N., & Lowe, C. F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior*, *73*(1), 5–22. https://doi.org/10.1901/jeab.2000.73-5

Dunn, P. K. (2021). *Scientific Research and Methodology: An introduction to quantitative research in science and health.* Chapman & Hall.

Dutton, E., & Lynn, R. (2014). Regional differences in intelligence and their social and economic correlates in Finland. *Mankind Quarterly*, *54*(3/4), 447.

Dutton, E., & Lynn, R. (2015). A negative Flynn Effect in France, 1999–2008/2009. Intelligence, 51, 67–70.

Dutton, E., van der Linden, D., & Lynn, R. (2016). The negative Flynn Effect: A systematic literature review. *Intelligence*, *59*, 163-169.

Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). The development of attention skills in action video game players. *Neuropsychologia*, *47*(8–9), 1780–1789. https://doi.org/10.1016/j.neuropsychologia.2009.02.002

Dymond, S., & Barnes, D. (1995). A transfer of self-discrimination response functions through equivalence relations. *Journal of the Experimental Analysis of Behaviour, 62*(2), 251-267. http://dx.doi.org/10.1901/jeab.1994.62-251

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behaviour, 64*(2), 163-184. https://doi.org/10.1901/jeab.1995.64-163

Dymond, S., & Roche, B. (Eds.). (2013). *Advances in relational frame theory: Research and application* .New Harbinger Publications, Inc..

Dymond, S., Roche, B., & Barnes-Holmes, D. (2003). The continuity strategy, human behavior and, and behavior analysis. *The Psychological Record, 53*(3), 333–347.

Edwards, C. A., Jagielo, J. A., Zentall, T. R., & Hogan, D. E. (1982). Acquired equivalence and distinctiveness in matching to sample by pigeons: Mediation by reinforcer-specific expectancies. *Journal of Experimental Psychology: Animal Behavior Processes, 8*(3), 244–259. https://doi.org/10.1037/0097-7403.8.3.244

Edwards, L., Figueras, B., Mellanby, J., & Langdon, D. (2011). Verbal and spatial analogical reasoning in deaf and hearing children: The role of grammar and vocabulary. *Journal of Deaf Studies and Deaf Education*, *16*(2), 189–197. https://doi.org/10.1093/deafed/enq051

Esher, F., Raven, I., & Earl, C. (1942). Discussion on testing intellectual capacity in adults. *Proceedings of the Royal Society of Medicine, 35*, 779–785.

Eysenck, H. J. (1971). Relation between intelligence and personality. *Perceptual and motor skills*, *32*(2), 637-638.

Eysenck, H. J. (1973) *The measurement of intelligence.* Williams and Wilkins Company.

Eysenck, H. J. (1998). *Dimensions of personality*. Transaction Publishers.

Farrington-Flint, L., Canobi, K. H., Wood, C., & Faulkner, D. (2007). The role of relational reasoning in children's addition concepts. *British Journal of Developmental Psychology, 25*(2), 227-245. https://doi.org/10.1348/026151006X108406

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2008). G*Power Version 3.1.2 [computer software]. Universität Kiel, Germany.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

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.

Flanagan, D. P., and Dixon, S. G. (2014). "The Cattell-Horn-Carroll theory of cognitive abilities," in *Encyclopedia of Special Education* (Hoboken, NJ: John Wiley & Sons, Inc.). doi: 10.1002/9781118660584.ese0431

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation.* Allyn & Bacon.

Fletcher, R.B. & Hattie, J. (2011). *Intelligence and intelligence testing*. Routledge: Taylor & Francis Group.

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan executive function system and the Woodcock–Johnson III tests of cognitive abilities. *Psychology in the Schools, 47*, 721–738. https://doi.org/ 10.1002/pits.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Flynn, J. R. (1998). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science*, *30*(4), 541–553. https://doi.org/10.1017/S0021932098005410

Flynn, J. R. (2007). *What is intelligence?: Beyond the flynn effect*. *What Is Intelligence? Beyond the Flynn Effect*. Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*(3), 353-365.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* Basic Books.

Gardner, H. (1997). *Frames of mind: The theory of multiple intelligences (10th anniversary ed.).* Basic Books.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and Arithmetical Cognition: Patterns of Functions and Deficits in Children at Risk for a Mathematical Disability. *Journal of Experimental Child Psychology*, *74*(3), 213–239. https://doi.org/10.1006/jecp.1999.2515

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M (magnitude) errors. *Perspectives on Psychological* Science, 9(6), 641-651. https://doi.org/10.1177/1745691614551642

Gelman, A., & Stern, H. (2006). The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. *In Amsel, E. et Al Language, Literacy , and Cogntive Development* (pp. 87-120). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Gignac, G. E. (2006). Evaluating the subtest ‘g’ saturation level via the single trait-correlat- ed uniqueness (STCU) SEM approach: Evidence in favor of crystallized subtests as the best indicators of ‘g’. Intelligence, 34, 29–46.

Gignac, G. E. (2015). Raven’s is not a pure measure of general intelligence: Implications for *g* factor theory and the brief measurement of *g*. *Intelligence, 52*, 71-79. https://doi.org/10.1016/j.intell.2015.07.006

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. (2001). The concepts of ‘sameness’ and ‘difference’ in an insect. *Nature*, *410*(6831), 930-933. http://dx.doi.org/10.1038/35073582

Glenberg, A. M. (1996). *Learning from data: An introduction to statistical reasoning*. Erlbaum.

Gobet, F., & Campitelli, G. (2006). Educational benefits of chess instruction. A critical review. In T. Redman (Ed.), *Chess and education. Selected essays from the Kolanowski Conference* (pp. 124–143). University of Texas at Dallas.

Goleman, D. (1995). *Emotional intelligence.* Bantam Books, Inc.

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society, 13*(2), 197-208. https://doi.org/10.1017/S1355617707070038

Gonzalez-DeHass, A. R., Furner, J. M., Vásquez-Colina, M. D., & Morris, J. D. (2024). Undergraduate students' math anxiety: The role of mindset, achievement goals, and parents. *International Journal of Science and Mathematics Education*, *22*(5), 1037-1056.

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *Journal of Experimental Education, 74*, 251–266. https://doi.org/10.3200/JEXE.74.3.249-266

Gore, N. J., Barnes-Holmes, Y., & Murphy, G. (2010). The Relationship between Intellectual Functioning and Relational Perspective-Taking. *International Journal of Psychology & Psychological Therapy, 10,* 1-17.

Gottfredson, L. S. (1998). The general intelligence factor. *Scientific American Inc.*, *9*, 24-29.

Gottfried., A. W. (1984). *Home environment and early cognitive development: Longitudinal research*. Academic Press.

Green, C. S., & Bavelier, D. (2012). Learning, attentional control, and action video games. *Current Biology, 20*(6), 197-206. https://doi.org/10.1016/j.cub.2012.02.012

Gregory, R. (1994). Seeing intelligence. In J. Khalfa (Ed). *What is intelligence?* (pp. 13- 26). Cambridge, England: Cambridge University Press.

Groth-Marnat, G. (2003). *Handbook of Psychological Assessment*. John Wiley & Sons.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence*, *29*(6), 523-535.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. https://doi.org/10.1016/0160-2896(84)90008-4

Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in Systems Neuroscience*, *8*(34). https://doi.org/10.3389/fnsys.2014.00034

Hakes, D. T., James, C. T., & Lloyd, R. H. (1965). Grammatical frames as stimuli in a four-stage stimulus-equivalence paradigm. *Psychonomic Science, 2*(12), 383–384. https://doi.org/10.3758/BF03343509

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences, 14*(11), 497 - 505. https://doi.org/10.1016/j.tics.2010.08.005

Hammer, P. (2022, April). Reasoning-learning systems based on non-axiomatic reasoning system theory. In *International Workshop on Self-Supervised Learning* (pp. 89-107). PMLR.

Hansen, K. T., Heckman, J. J., & Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of econometrics*, *121*(1-2), 39-98.

Harmon, K., Strong, R., & Pasnak, R. (1982). Relational responses in tests of transposition with Rhesus monkeys. *Learning and Motivation*, *13*(4), 495–504. https://doi.org/10.1016/0023-9690(82)90006-6

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology, 28*(6), 1096–1105. https://doi.org/10.1037/0012-1649.28.6.1096

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Hart, B., Risley, T. R., & Kirby, J. R. (1997). Meaningful differences in the everyday experience of young American children. *Canadian Journal of Education, 22*(3), 323.

Haworth, C. M., Dale, P., & Plomin, R. (2008). A Twin Study into the Genetic and Environmental Influences on Academic Performance in Science in nine-year-old Boys and Girls. *International journal of science education*, *30*(8), 1003. https://doi.org/10.1080/09500690701324190

Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., Bartels, M., Posthuma, D., Boomsma, D. I., Davis, O. S., Kovas, Y., Corley, R. P., Defries, J. C., Hewitt, J. K., Olson, R. K., Rhea, S. A., Wadsworth, S. J., Iacono, W. G., McGue, M., Thompson, L. A., Hart, S. A., Petril, S., A., Lubinski, D., & Plomin, R. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular psychiatry*, *15*(11), 1112–1120. https://doi.org/10.1038/mp.2009.55

Hayes, J., & Stewart, I. (2016). Comparing the effects of derived relational training and computer coding on intellectual potential in school-age children. *The British Journal of Educational Psychology*, *86*(3), 397–411. https://doi.org/10.1111/bjep.12114

Hayes, S. C. (1991). A relational control theory of stimulus equivalence. In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 19-40). Context Press.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* Kluwer Academic/Plenum Publishers.

Hayes, S., & Barnes, D. (1997). Analyzing Derived Stimulus Relations Requires More Than The Concept Of Stimulus Class. *J Exp Anal Behav*, *68*(2), 235-244.

Hayes, S.C., Fox, E., Gifford, E.V., Wilson, K.G., Barnes-Holmes, D. & Healy, O. (2001). Derived relational responding as learned behavior. In. S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of language and cognition*. Plenum Press.

Healy, O., Barnes, D., & Smeets, P. M. (1998). Derived relational responding as an operant: The effects of between-session feedback. *The Psychological Record*, 48, 511–536.

Healy, W., & Fernald, G. M. (1911). Tests for practical mental classification. *The Psychological Monographs, 13*(2), i–54. https://doi.org/10.1037/h0093055

Herman, L. M., & Thompson, R. K. R. (1982). Symbolic, identity, and probe delayed matching of sounds by the bottlenosed dolphin. *Animal Learning & Behavior*, *10*(1), 22–34. https://doi.org/10.3758/BF03212042

Herrnstein, R. J. (1971). IQ. *Atlantic Monthly, 228*, 43-64.

Herrnstein, R. J., & Murray, C. A. (1994). *The bell curve: Intelligence and class structure in American life.* Free Press.

Holmes, K. J., & Wolff, P. (2013). Spatial language and the psychological reality of schematisation. *Cognitive Processing, 14*(2), 205-208. https://doi.org/10.1007/s10339-013-0545-5

Holyoak, K. J. (2012). Analogy and relational reasoning. *The Oxford handbook of thinking and reasoning*, 234-259.

Hooper, D., Coughlan, J., & Mullen, M. (2008, September). Evaluating model fit: a synthesis of the structural equation modelling literature. In *7th European Conference on research methodology for business and management studies* (Vol. 2008, pp. 195-200).

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Allyn & Bacon.

Horn, J., & Cattell, R. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, *26*, 107-129. http://dx.doi.org/10.1016/0001- 6918(67)90011-x

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185-241.

Hotz, V. J., Wiemers, E. E., Rasmussen, J., & Koegel, K. M. (2023). The role of parental wealth and income in financing children's college attendance and its consequences. *Journal of Human Resources*, *58*(6), 1850-1880.

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: The basic account. *The Wiley handbook of contextual behavioral science*, 129-178.

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology, 19*(5), 640-648. https://doi.org/10.1097/ede.0b013e31818131e7

Isaacs, E. B., Fischl, B. R., Quinn, B. T., Chong, W. K., Gdian, D. G., & Lucas, A. (2010). Impact of breast milk on intelligence quotient, brain size, and white matter development. *Pediatric Research, 67*(4), 357-362.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning - implications for training and transfer. *Intelligence*, *38*(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001

Jaeggi, S., Buschkuehl, M., Jonides, J., & Perrig, W. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy Of Sciences, 105*(19), 6829-6833. http://dx.doi.org/10.1073/pnas.0801268105

Jensen, A. (1969). How Much Can We Boost IQ and Scholastic Achievement. *Harvard Educational Review*, *39*(1), 1-123. http://dx.doi.org/10.17763/haer.39.1.l3u15956627424k7

Jensen, A. (1980). Chronometric analysis of intelligence. Journal of Social & Biological Structures, 3, 103–122.

Jensen, A. R. (1973). Level I and level II abilities in three ethnic groups. *American Educational Research Journal, 10*(4), 263–276.

Jensen, A. R. (1987). Further evidence for Spearman's hypothesis concerning black–white differences on psychometric tests. *Behavioral and Brain Sciences, 10*(03), 512–519.

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

Jensen, A. R., & Cattell, R. B. (2006). Abilities: Their Structure, Growth, and Action. *The American Journal of Psychology*, *87*(1/2), 290. https://doi.org/10.2307/1422024

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2016). The Cattell–Horn–Carroll model of cognition for clinical assessment. *Journal of Psychoeducational Assessment, 1*, 1–21. https://doi.org/10.1177/0734282916651360.

Johnson, W., Penke, L., & Spinath, F. M. (2011). Heritability in the era of molecular genetics: Some thoughts for understanding genetic influences on behavioural traits. *European Journal of Personality*, *25*(4), 254-266. https://doi.org/10.1002/per.836

Joseph, J. (1998). The equal environment assumption of the classical twin method: A critical analysis. *The Journal of Mind and Behavior*, 325-358.

Kamphaus, R. (2005). Clinical assessment of child and adolescent intelligence (2nd Ed.). Springer-Verlag

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Allyn & Bacon.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence, 40*(2), 123–138. https://doi.org/10.1016/j.intell.2012.01.009

Kazemi, F., Yektayar, M., & Abad, A. M. B. (2012). Investigation the impact of chess play on developing meta-cognitive ability and math problem-solving power of students at different levels of education. In *Procedia - Social and Behavioral Sciences* (Vol. 32, pp. 372–379). https://doi.org/10.1016/j.sbspro.2012.01.056

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650. https://doi.org/10.1002/pits.

Kline, T. J. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage publications.

Knox, H. A. (1914). A scale, based on the work at Ellis Island, for estimating mental defect. *Journal of the American Medical Association*, *62*(10), 741-747.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of cognitive neuroscience*, *16*(8), 1412-1425.

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry, 27*(3), 151–177. https://doi.org/10.1080/1047840X.2016.1153946

Kovacs, K., & Conway, A. R. A. (2019). What is IQ? Life beyond "general intelligence". *Current Directions in Psychological Science, 28*(2), 189–194. https://doi.org/10.1177/0963721419827275

Krijnen, W. P. (2004). Positive loadings and factor correlations from positive covariance matrices. *Psychometrika, 69*(4), 655–660. https://doi.org/10.1007/BF02289861

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433. https://doi.org/10.1016/S0160-2896(05)80012-1

Kyllonen, P. C., Lohman, D. F., & Woltz, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology*, *76*(6), 1325–1345. https://doi.org/10.1037/0022-0663.76.6.1325

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence*, *14*(3), 309-325.

Lashley, K. S. (1938). Conditional reactions in the rat. *The Journal of Psychology*, *6*(2), 311-324.

Layzer, D. (1972). Science or superstition? (A physical scientist looks at the IQ controversy). *Cognition*, *1*(2-3), 265-299.

Lee, E. S. (1951). Negro intelligence and selective migration: a Philadelphia test of the Klineberg hypothesis. *American Sociological Review, 16,* 227–233. https://doi.org/10.2307/2087699

Leech, A., & Barnes-Holmes, D. (2020). Training and testing for a transformation of fear and avoidance functions via combinatorial entailment using the Implicit Relational Assessment Procedure (IRAP): Further exploratory analyses. *Behavioural Processes*, *172*, 104027.

Lenroot, R. K., Schmitt, J. E., Ordaz, S. J., Wallace, G. L., Neale, M. C., Lerch, J. P., Kenneth, S. K., Evans, A. C., & Giedd, J. N. (2009). Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human brain mapping*, *30*(1), 163-174. https://doi.org/10.1002/hbm.20494

Lezak, M. D. (1995). Neuropsychological assessment (3rd ed.). New York: Oxford University Press.

Li, S. C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the coulings 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

Linehan, C., Kirman, B. and Roche, B. (2015) 'Gamification as behavioral psychology' in Walz, S.P. and Deterding, S. (eds.) The Gameful World: Approaches, Issues, Applications. MIT Press.

Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of experimental child psychology*, *56*(2), 201–239. https://doi.org/10.1006/jecp.1993.1032

Llabre, M. M. (1984). Standard progressive matrices. In D. J. Keyser, & R. C. Sweetland (Eds.), Test critiques, Vol. 1. (pp. 595–602). Test Corporation of America.

Lovecky, D. V. (2004). *Different Minds: Gifted Children With AD/HD, Asperger Syndrome, and Other Learning Deficits*. Jessica Kingsley.

Lowenkron, B. (1998). Some logical functions of joint control. *Journal of the Experimental Analysis of Behavior, 69*, 327-354.

Lupyan, G. (2015). The centrality of language in human cognition. *Language Learning, 66*(3), 516-553. https://doi.org/10.1111/lang.12155

Lupyan, G. (2017). The paradox of the universal triangle: Concepts, language, and prototypes. *Quarterly Journal of Experimental Psychology, 70*(3), 389-412. https://doi.org/10.1080/17470218.2015.1130730

Lupyan, G., & Bergen, B. (2015). How language programs the mind. *Topics in Cognitive Science, 8*(2), 408-424. https://doi.org/10.1111/tops.12155

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General, 141*(1), 170–186. https://doi.org/10.1037/a0024904

Lynn, R. (1990). The role of nutrition in secular increases in intelligence. *Personality and Individual Differences, 11*(3), 273–285. https://doi.org/10.1016/0191-8869(90)90241-I

Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence*, *37*(1), 16–24. https://doi.org/10.1016/j.intell.2008.07.008

Mackay, H. A. (1985). Stimulus equivalence in rudimentary reading and spelling. *Analysis and Intervention In Developmental Disabilities*, *5*(4), 373–387. https://doi.org/10.1016/0270-4684(85)90006-0

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood: Fast-Track Report. *Developmental Science*, *11*(3). https://doi.org/10.1111/j.1467-7687.2008.00671.x

Marr, M. J. (2015). Mathematics as verbal behavior. *Behavioural Processes*, *113*, 75-80.

Martin, T. A., Donders, J., & Thompson, E. (2000). Potential of and problems with new measures of psychometric intelligence after traumatic brain injury. Rehabilitation Psychology, 45, 402–408.

Martinez, M. E. (2013). *Future bright: A transforming vision of human intelligence*. Oxford University Press.

Martschenko, D. (2018, January 29). The IQ test wars: Why screening for intelligence is still so controversial. *Independent. https://www.independent.co.uk/news/education/the-iq-test-wars-why-screening-for-intelligence-is-still-so-controversial-a8179176.html*

May, R. J., Stewart, I., Baez, L., Freegard, G., & Dymond, S. (2017). Arbitrarily applicable spatial relational responding. *Journal of the Experimental Analysis of Behaviour, 107*(2), 234-257. https://doi.org/10.1002/jeab.250

May, R. J., Tyndall, I., McTiernan, A., Roderique-Davies, G., & McLoughlin, S. (2022). The impact of the SMART program on cognitive and academic skills: A systematic review and meta-analysis. *British Journal of Educational Technology, 53*(5), 1244–1261. https://doi.org/10.1111/bjet.13192

McCutchen, D. (1996). A capacity theory of writing: working memory in composition. *Educational Psychology Review*, *8*(3), 299–325. https://doi.org/10.1007/BF01464076

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–119). The Guilford Press.

McGrew, K. S., & Evans, J. J. (2005). Internal and external factorial extensions to the Cattell-Horn-Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise. *Institute for Applied Psychometrics*.

McGurn, B., Starr, J. M., Topfer, J. A., Pattie, A., Whiteman, M. C., Lemmon, H. A., Whalley, L. J., & Deary, I. J. (2004). Pronunciation of irregular words is preserved in dementia, validating premorbid IQ estimation. *Neurology*, *62*(7), 1184–1186. https://doi.org/10.1212/01.wnl.0000103169.80910.8b

McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *Psychological Record*, 54(1), 115-144. https://doi.org/10.1007/BF03395465

McLoughlin, S., & Roche, B. T. (2023). ACT: A process-based therapy in search of a process. *Behaviour Therapy, 54*(6), 939-955. https://doi.org/10.1016/j.beth.2022.07.010

McLoughlin, S., & Stewart, I. (2017). Empirical advances in studying relational networks. *Journal of Contextual Behavioural Science, 6*(3), 329-342. https://doi.org/10.1016/j.jcbs.2016.11.009

McLoughlin, S., Tyndall, I., & Pereira, A. (2018). Piloting a brief relational operant training program: Analyses of response latencies and intelligence test performance. *European Journal of Behaviour Analysis, 19*(2), 228-246. https://doi.org/10.1080/15021149.2018.1507087

McLoughlin, S., Tyndall, I., & Pereira, A. (2020). Convergence of multiple fields on a relational reasoning approach to cognition. *Intelligence*, *83*, 101491.

McLoughlin, S., Tyndall, I., & Pereira, A. (2022). Relational operant skills training increases standardized matrices scores in adolescents: A stratified active-controlled trial. *Journal of Behavioral Education, 31*(2), 298–325. https://doi.org/10.1007/s10864-020-09399-x

McLoughlin, S., Tyndall, I., & Pereira, A., & Mulhern, T. (2021). Non-verbal IQ gains from relational operant training explain variance in educational attainment: An active-controlled feasibility study. *Journal of Cognitive Enhancement, 5*, 35-50. https://doi.org/10.1007/s41465-020-00187-z

Mekik, C. S., Sun, R., & Dai, D. Y. (2018). Similarity-Based Reasoning, Raven's Matrices, and General Intelligence. In *IJCAI* (Vol. 18, pp. 1576-82).

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Metz, Brittany Leigh, "A Comparison of the WJ-III T est of Cognitive Abilities and the WAIS-III" (2005). Theses, Dissertations and Capstones. Paper 762.

Metz, Brittany Leigh, "A Comparison of the WJ-III Test of Cognitive Abilities and the WAIS-III" (2005). Theses, Dissertations and Capstones. Paper 762.

Miendlarzewska, E. A., & Trost, W. J. (2014). How musical training affects cognitive development: rhythm, reward and other modulating variables. *Frontiers in neuroscience*, *7*, 279. https://doi.org/10.3389/fnins.2013.00279

Mingroni, M. A. (2004). The secular rise in IQ: Giving heterosis a closer look. *Intelligence*, *32*(1), 65-83. https://doi.org/10.1016/S0160-2896(03)00058-8

Mingroni, M. A. (2007). Resolving the IQ Paradox: Heterosis as a Cause of the Flynn Effect and Other Trends. *Psychological Review*, *114*(3), 806–829. https://doi.org/10.1037/0033-295X.114.3.806

Molenaar, D., Kő, N., Rózsa, S., & Mészáros, A. (2017). Differentiation of cognitive abilities in the WAIS-IV at the item level. *Intelligence, 65,* 48–59. https://doi.org/10.1016/j.intell.2017.10.004

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*(4), 327–328. https://doi.org/10.1016/j.intell.2009.04.005

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). Brief report: The training and assessment of relational precursors and abilities (TARPA): A preliminary analysis. *Journal of Autism and Developmental Disorders*, *40*(9), 1149–1153. https://doi.org/10.1007/s10803-010-0968-0288

Mortensen, E. L., Michaelsen, K. F., Sanders, S. A., & Reinisch, J. M. (2002). The association between duration of breastfeeding and adult intelligence. *JAMA: Journal of the American Medical Association, 287*(18), 2365–2371. https://doi.org/10.1001/jama.287.18.2365

Myerson, J., Hale, S., Zheng, Y., Jenkins, L., & Widaman, K. F. (2003). The difference engine: A model of diversity in speeded cognition. *Psychonomic Bulletin and Review*, *10*(2), 262–288. https://doi.org/10.3758/BF03196491

Naglieri, J. A., & Bardos, A. N. (1997). General Ability Scale for Adults (GAMA). National Computer Systems.

Naglieri, J. A., & Bardos, A. N. (1997). General Ability Scale for Adults (GAMA). National Computer Systems.

Neisser, U. (1998). Introduction: Rising test scores and what they mean. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 3–22). American Psychological Association.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual.* NFER-Nelson.

Nelson, H. E., & Willison, J. (1991). *National adult reading test (NART)*. Nfer-Nelson.

Nicolas, S., Andrieu, B., Croizet, J.-C., Sanitioso, R. B., & Burman, J. T. (2013). Sick? Or slow? On the origins of intelligence as a psychological object. *Intelligence, 41*(5), 699–711. https://doi.org/10.1016/j.intell.2013.08.006

Nippold, M. A., & Sullivan, M. P. (1987). Verbal and perceptual analogical reasoning and proportional metaphor comprehension in young children. *Journal of Speech & Hearing Research, 30*(3), 367–376. https://doi.org/10.1044/jshr.3003.367

Nissen, H. W., Riesen, A. H., & Nowlis, V. (1938). Delayed response and discrimination learning by chimpanzees. *Journal of Comparative Psychology*, *26*(2), 361–386. https://doi.org/10.1037/h0056387

Njiokiktjien, C. H., & Verschoor, C. A. (1998). Attention deficits in children with low performance IQ: arguments for right hemisphere dysfunction. *Fiziologiia cheloveka*, *24*(2), 16–22.

O'Carroll, R. (1995). The assessment of premorbid ability: A critical review. *Neurocase, 1*(1), 8389–89.

O'Connor, J., Rafferty, A., Barnes-Holmes, D., & Barnes-Holmes, Y. (2009). The role of verbal behavior, stimulus nameability, and familiarity on the equivalence performances of autistic and normally developing children. *Psychological Record*, *59*(1), 53–74. https://doi.org/10.1007/BF03395649

O'Connor, M., Farrell, L., Munnelly, A., & McHugh, L. (2017). Citation analysis of relational frame theory: 2009-2016. *Journal of Contextual Behavioural Science, 6*(2), 152-158. https://doi.org/10.1016/j.jcbs.2017.04.009

O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the wais-iii. *Psychological Record*, *55*(1), 155–175. https://doi.org/10.1007/BF03395504

O'Hora, D., Peláez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudhary, T. (2008). Temporal relations and intelligence : Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, *58*(4), 559-584. https://doi.org/10.1007/BF03395638

O'Neill, J. (2017, January 10). *Household disposable income and inequality in the UK: financial year ending 2016*. Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2016

O'Toole, C., & Barnes-Holmes, D. (2009). Three chronometric indices of relational responding as predictors of performance on a Brief Intelligence Test: The importance of relational flexibility. *Psychological Record, 59*(1), 119-132. https://doi.org/10.1007/BF03395652

O'Toole, C., Barnes-Holmes, D., Murphy, C., O'Connor, J., & Barnes-Holmes, Y. (2009). Relational flexibility and human intelligence: Extending the remit of skinner's verbal behavior. *International Journal of Psychology and Psychological Therapy*, *9*(1), 1–17.

Oommen, A. (2014). Factors influencing intelligence quotient. *Journal of Neurology and Stroke, 1*(14), 129-133. https://doi.org/10.15406/jnsk.2014.01.00023

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., … Ballard, C. G. (2010). Putting brain training to the test. *Nature*, *465*(7299), 775–778. https://doi.org/10.1038/nature09042

Peal, E., & Lambert, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied, 76*(27, Whole No. 546), Peterson, J. (1925). *Early conceptions and tests of intelligence.* World Book Company. https://doi.org/10.1037/11569-000

Perkins, F. T. (1931). A further study of configurational learning in goldfish. *Journal of Experimental Psychology, 14,* 508-538.

Peterson, J. (1925). *Early* Company. *conceptions and tests of intelligence.* World Book https://doi.org/10.1037/11569-000

Petrill, S., Saudino, K., Cherny, S., Emde, R., Fulker, D., Hewitt, J., & Plomin, R. (1998). Exploring the Genetic and Environmental Etiology of High General Cognitive Ability in Fourteen- to Thirty-Six-Month-Old Twins. *Child Development*, *69*(1), 68-74. http://dx.doi.org/10.1111/j.1467- 8624.1998.tb06133.x

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn Effect (1909–2013). Perspectives on Psychological Science, 10(3), 282–306.

Pike, A., Reiss, D., Hetherington, E., & Plomin, R. (1996). Using MZ Differences in the Search for Nonshared Environmental Effects. *J Child Psychol & Psychiat*, *37*(6), 695-704. http://dx.doi.org/10.1111/j.1469-7610.1996.tb01461.x

Plomin, R. (1990). *Nature and nurture: An introduction to human behavioral genetics*. Brooks/Cole.

Plomin, R. (1997). *Behavioral genetics*. W.H. Freeman.

Plomin, R., & DeFries, J. (1998). The Genetics of Cognitive Abilities and Disabilities. *Sci Am*, *278*(5), 62-69. http://dx.doi.org/10.1038/scientificamerican0598-62

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*(2), 309–322. https://doi.org/10.1037/0033-2909.84.2.309

Pollitt, E., Gorman, K. S., Engle, P. L., Martorell, R., et al. (1993). Early supplementary feeding and cognition: Effects over two decades. *Monographs of the Society for Research in Child Development, 58*(7), 1–99. https://doi.org/10.2307/1166162

Prade, H., & Richard, G. (2009). Analogy, paralogy and reverse analogy: Postulates and inferences. *Annual Conference on Artificial Intelligence* (pp. 306–314). Springer. https://doi.org/10.1007/978-3-642-04617-9_39

Prade, H., & Richard, G. (2010). Reasoning with logical proportions. *Twelfth International Conference on the Principles of Knowledge Representation and Reasoning*.

Prade, H., & Richard, G. (2013). From analogical proportion to logical proportions. *Logica Universalis*, *7*, 441–505. https://doi.org/10.1007/s11787-013-0089-6

Qian, M., Wang, D., Watkins, W. E., Gebski, V., Yan, Y. Q., Li, M., & Chen, Z. P. (2005). The effects of iodine on intelligence in children: a meta-analysis of studies conducted in China. Asia Pacific Journal of Clinical Nutrition, 14(1), 32-42.

Rammsayer, T. H., & Troche, S. J. (2016). Validity of the worst performance rule as a function of task complexity and psychometric *g*: On the crucial role of *g* saturation. *Journal of Intelligence, 4*(1), Article 5. https://doi.org/10.3390/jintelligence4010005

Raven, J. C. (1940). Matrix tests. *Mental Health, 1*, 10–18.

Raven, J. C. (1941). Standardisation of progressive matrices. *The British Journal of Medical Psychology, 19*, 137–150.

Raven, J. C. (1966). *Advanced progressive matrices*. Psychological Corporation.

Raven, J. C., Court, J. H., & Raven, J. (1994). *Advanced progressive matrices: Sets I and II. Manual for Raven's progressive matrices and vocabulary scales*. Oxford Psychologists Press.

Raven, J. C., Raven, J. C., & Court, J. H. (1962). *Coloured progressive matrices*. HK Lewis.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4. Advanced progressive matrices*. Oxford Psychologist Press.

Rawat, P. S., Singh, S., Mahdi, A. A., & Mehrotra, S. (2022). Environmental lead exposure and its correlation with intelligence quotient level in children. *Journal of Trace Elements in Medicine and Biology*, *72*, 126981.

Rey, A. (1958). L'examen clinique en psychologie. Paris: Presses Universitaire de France.

Richardson, K., & Norgate, S. H. (2015). Does IQ really predict job performance? *Applied Developmental Science, 19*(3), 153–169. https://doi.org/10.1080/10888691.2014.983635

Rindermann, H., Becker, D., & Coyle, T. R. (2016). Survey of expert opinion on intelligence: Causes of international differences in cognitive abilities. *Frontiers in Psychology, 7*(399). https://doi.org/10.3389/fpsyg.2016.00399

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological science*, *29*(8), 1358-1369.

Roche, B., & Barnes, D. (1996). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record*, *46*, 451-475.

Roche, B., & Barnes, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of behavior*, *67*(3), 275-301.

Roche, B., & Barnes, D. (1997). The behavior of organisms?. *The Psychological Record*, *47*, 597-618.

Roche, B., Barnes-Holmes, D., Smeets, P. M., Barnes-Holmes, Y., & McGeady, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record, 50,* 267-291.

Roche, B., Cummins, C., Cassidy, S., Dillon, A., Moore, L., & Grey, I. (2023). The effect of SMART relational skills training on intelligence quotients: Controlling for individual differences in attentional skills and baseline IQ. *Journal of Contextual Behavioural Science, 28*, 185-197. https://doi.org/10.1016/j.jcbs.2023.03.012

Roth, B., Romeyke, S., Spinath, F. M., Domnick, F., Schäfer, S., & Becker, N. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, *53*, 118–137. https://doi.org/10.1016/j.intell.2015.09.002

Rudebeck, S. R., Bor, D., Ormond, A., O'Reilly, J. X., & Lee, A. C. H. (2012). A Potential Spatial Working Memory Training Task to Improve Both Episodic Memory and Fluid Intelligence. *PLoS ONE*, *7*(11). https://doi.org/10.1371/journal.pone.0050431

Sackheim, H.A., Freeman, J., McElhiney, M., Coleman, E., Prudic, J., & Devanand, D.P. (1992). Effects of major depression on estimates of intelligence. *Journal of Clinical and Experimental Neuropsychology*, 14, 268–288.

Sala, G., & Gobet, F. (2016). Do the benefits of chess instruction transfer to academic and cognitive skills? A meta-analysis. *Educational Research Review, 18,* 46-47. https://doi.org/10.1016/j.edurev.2016.02.002

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*(4), 671–685. https://doi.org/10.1037/dev0000265

Saloojee, H., & Pettifor, J. M. (2001). Iron deficiency and impaired child development. British Medical Journal, 323(7326), 1377-1378. https://doi.org/10.1136/bmj.323.7326.1377

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19*, 532–545. https://doi.org/10.1037/0894-4105.19.4.532.

Scarr-Salapatek, S. (1971). Race, Social Class, and IQ: Population differences in heritability of IQ scores were found for racial and social class groups. *Science*, *174*(4016), 1285-1295.

Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype→ environment effects. *Child development*, 424-435.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*(8), 511–514. https://doi.org/10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G., & Moreno, S. (2010). Music lessons, pitch processing, and g. *Psychology of Music, 38*(2), 209–221.

Schlinger, H. D. (2003). Myth of intelligence. *The Psychological Record*, *53*, 15–32. https://doi.org/Article

Schoenthaler, S. J., & Bier, I. D. (2000). The effect of vitamin-mineral supplementation on juvenile delinquency among American schoolchildren: a randomised, double-blind placebo-controlled trial. *Journal of alternative and complementary medicine, 6*(1), 7–17. https://doi.org/10.1089/acm.2000.6.7

Schubert, T., Strobach, T., & Karbach, J. (2014). New directions in cognitive training: on methods, transfer, and application. *Psychological Research*, *78*(6), 749–755. https://doi.org/10.1007/s00426-014-0619-8

Sharpe, K., & O'Carroll, R. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: A Canadian study. *British Journal of Clinical Psychology, 30*(4), 381–384.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*(7084), 676-679.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., Clasen, L., Evans, A., Rapoport, J. L., Giedd, J. N., & Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *28*(14), 3586–3594. https://doi.org/10.1523/JNEUROSCI.5309-07.2008

Shayer, M., & Ginsburg, D. (2009). Thirty years on: A large anti-Flynn effect? (II): 13- and 14-year-olds. Piagetian tests of formal operations norms 1976–2006/7. British Journal of Educational Psychology, 79(Pt. 3), 409–418.

Shayer, M., Ginsburg, D., & Coe, R. (2007). Thirty years on: A large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. British Journal of Educational Psychology, 77, 25–41.

Sheshagiri, S., Khaskaran. J. K., Patel, K. S., & Rajagopala, S. (2016). Factors influencing intelligence quotient of school going children in and around Jamnagar – A survey study. *Journal of Research in Traditional Medicine, 2*(5), 122-129. *10.21276/jrtm.2016/83*

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research. 14*, 5–13.

Sidman, M. (1994). *Equivalence relations and behavior*. Authors Cooperative.

Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior, 74*(1), 127–146. https://doi.org/10.1901/jeab.2000.74-127

Siegler, R. (1992). The other Alfred Binet. *Developmental Psychology, 28*, 179-190.

Silva, M. A. (2008). Development of the WAIS-III: A brief overview, history and description. Graduate Journal of Counseling Psychology, 1(1), 1–19.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest, Supplement*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Skinner, B. F. (1957). *Verbal behavior.* Appleton-Century-Crofts.

Slattery, B., & Stewart, I. (2014). Hierarchical classification as relational framing. *Journal of the experimental analysis of behavior*, *101*(1), 61-75.

Smeets, P., Barnes, D., & Roche, B. (1997). Functional Equivalence in Children: Derived Stimulus–Response and Stimulus–Stimulus Relations. *Journal Of Experimental Child Psychology*, *66*(1), 1-17. http://dx.doi.org/10.1006/jecp.1997.2378

Smith, B., Smith, T., Taylor, L., & Hobby, M. (2005). Relationship between intelligence and vocabulary 1,2. *Perceptual And Motor Skills*, *100*(1), 101-108. http://dx.doi.org/10.2466/pms.100.1.101-108

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal Of Psychology*, *15*(2), 201. http://dx.doi.org/10.2307/1412107

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

Spearman, C. (1946). Theory of general factor. *British Journal of Psychology. General Section, 36*(3), 117–131.

Spinks, R., McKirgan, L. W., Arndt, S., Caspers, K., Yucuis, R., & Pfalzgraf, C. J. (2009). IQ estimate smackdown: Comparing IQ proxy measures to the WAIS-III. Journal of the International Neuropsychological Society, 15(4), 590–596.

Staats, A., & Burns, G. (1981). *Intelligence and child development*. The Journal Press.

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: The best non-cognitive predictor of academic achievement? *Educational Psychology, 34*(1), 9–28. https://doi.org/10.1080/01443410.2013.814194

Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading research quarterly*, 267-283.

Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behaviour, 56*(3), 519-555. https://doi.org/10.1901/jeab.1991.56-519

Stein, A. D., Behrman, J. R., DiGirolamo, A., Grajeda, R., Martorell, R., Quisumbing, A., & Ramakrishnan, U. (2005). Schooling, educational achievement, and cognitive functioning among young Guatemalan adults. *Food and nutrition bulletin*, *26*(2 Suppl 1), S46–S54. https://doi.org/10.1177/15648265050262S105

Sternberg, R. (2008). Increasing fluid intelligence is possible after all. *Proceedings Of The National Academy Of Sciences*, *105*(19), 6791-6792. http://dx.doi.org/10.1073/pnas.0803396105

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

Stewart, I., & McElwee, J. (2009). Relational responding and conditional discrimination procedures: An apparent inconsistency and clarification. The Behaviour Analyst, 32, 309–317. https://doi.org/10.1007/bf03392194

Stewart, I., & Roche, B. (2013). Relational Frame Theory: An overview. In S. Dymond, & B. Roche (Eds.) *Advances in Relational Frame Theory & Contextual Behavioral Science: Research & Application* (pp. 124-147). New Harbinger Press.

Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *Psychological Record, 54*(4), 531-552. https://doi.org/10.1007/BF03395491

Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). Generating derived relational networks via the abstraction of common physical properties: A possible model of analogical reasoning. *The Psychological Record*, *51*, 381-408.

Stewart, I., Tarbox, J., Roche, B., & O'Hora, D. (2013). Education, intellectual development, and relational frame theory. In S. Dymond, & B. Roche (Eds.) *Advances in Relational Frame Theory: Research & Application* (pp. 178-198). New Harbinger.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401-426.

Stricker, C., Mao, J., Cassidy, S., Colbert, D., & Roche, B. (2024). A Relational Frame Theory-Based Intervention for Improving Reading and Mathematical Competencies Among School Children. *Journal of Behavioral Education*, 1-23.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. Intelligence, 32, 349–362.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical Problem Solving and Working Memory in Children with Learning Disabilities: Both Executive and Phonological Processes Are Important. *Journal of Experimental Child Psychology*, *79*(3), 294–321. https://doi.org/10.1006/jecp.2000.2587

Taylor, E. M. (1959). The appraisal of children with cerebral deficits. Cambridge, MA: Harvard University Press.

Teasdale, T. W., & Owen, D. R. (2008). Secular declines in cognitive test scores: A reversal of the Flynn effect. Intelligence, 36(2), 121–126.

The Psychological Society of Ireland. (2019). Code of professional ethics. (1978, amended effective November 9, 1991, May 16, 1998, November 13, 1999, and October 19, 2019). https://www.psychologicalsociety.ie/source/Code%20of%20Professional%20Ethics%20(Oct%202019).pdf

Thomas, M. D., McGrath, A., Sugden, N., Weekes, C., & Skilbeck, C. E. (2021). Investigating the National Adult Reading Test (NART-2) and Wechsler Test of Adult Reading Test (WTAR) in predicting Wechsler Abbreviated Scale of Intelligence—Second edition (WASI-II) scores in an Australian sample. *Australian Psychologist, 56*(5), 372–381. https://doi.org/10.1080/00050067.2021.1937923

Thorndike, R. L. (1986). Historical and theoretical perspectives. In R. L. Thorndike, E. P. Hagen, & J. M. Sattler (Eds.), The Standford–Binet intelligence scale (Fourth edition). Technical manual. (pp. 1–7). Riverside.

Toril, P., Reales, J. M., & Ballesteros, S. (2014). Video game training enhances cognition of older adults: A meta-analytic study. *Psychology and Aging*, *29*(3), 706–716. https://doi.org/10.1037/a0037507

Toril, P., Reales, J. M., Mayas, J., & Ballesteros, S. (2016). Video Game Training Enhances Visuospatial Working Memory and Episodic Memory in Older Adults. *Frontiers in Human Neuroscience*, *10*, 206. https://doi.org/10.3389/fnhum.2016.00206

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

Turner, M.A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning- disabled individuals with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 40,* 189-201.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence, 33*(1), 67–81. https://doi.org/10.1016/j.intell.2004.08.003

van der Linde, I., & Bright, P. (2024). The prediction limits of the National Adult Reading Test and its abbreviated and international variants. *Journal of the International Neuropsychological Society*, *30*(9), 812–818. doi:10.1017/S1355617724000420

VanderWeele, T. J., & Mathur, M. B. (2019). Some desirable properties of the Bonferroni correction: Is the Bonferroni correction really so bad?. *American Journal of Epidemiology, 188*(3), 617-618. https://doi.org/10.1093/aje/kwy250

Varriale, V., van der Molen, M. W., & De Pascalis, V. (2018). Mental rotation and fluid intelligence: A brain potential analysis. *Intelligence, 69,* 146–157. https://doi.org/10.1016/j.intell.2018.05.007

Vaskinn, A., & Sundet, K. (2001). Estimering av premorbid IQ: En norsk versjon av National Adult Reading Test/Estimating premorbid IQ: A Norwegian version of National Adult Reading Test. *Tidsskrift for Norsk Psykologforening, 38*(12), 1133–1140.

Vernon, M. D. (1947). Different types of perceptual ability. *British journal of psychology*, *38*(2), 79.

Vernon, P. E. (1947). The variations of intelligence with occupation, age, and locality. *British Journal of Statistical Psychology, 1*(1), 52–63.

Vetterli, C., & Furedy, J. (1997). Correlates of intelligence in computer measured aspects of prose vocabulary: word length, diversity, and rarity. *Personality And Individual Differences, 22*(6), 933-935. http://dx.doi.org/10.1016/s0191- 8869(96)00265-6

Vial, M., & Hugon, M. A. (1998). *La commission Bourgeois (1904–1905).* CTNERHI.

Villatte, M., Monestès, J.-L., McHugh, L., i Baquè, E. F., & Loas, G. (2010).Adopting the perspective of another in belief attribution: Contribution of Relational Frame Theory to the understanding of impairments in schizophrenia. *Journal of Behaviour Therapy and Experimental Psychiatry, 41*(2), 125-153. https://doi.org/10.1016/j.jbtep.2009.11.004

Wainer, H. (2002). On the automatic generation of test items: Some whens, whys, and hows. In S. H. Irvine, & P. C. Kyllonen (Eds.), *Item generation for test development* (pp. 287–314). Lawrence Erlbaum Associates.

Walker, S. Y., Pierre, R. B., Christie, C. D. C., & Chang, S. M. (2013). Neurocognitive function in HIV-positive children in a developing country. *International Journal of Infectious Diseases, 17*(10), e862–e867.

Wang, P., Liu, H. H., Zhu, X. T., Meng, T., Li, H. J., & Zuo, X. N. (2016). Action video game training for healthy adults: A meta-analytic study. *Frontiers in Psychology*, *7,* 907. https://doi.org/10.3389/fpsyg.2016.00907

Waterhouse, L. (2006). Inadequate Evidence for Multiple Intelligences, Mozart Effect, and Emotional Intelligence Theories. *Educational Psychologist, 41*(4), 247–255. https://doi.org/10.1207/s15326985ep4104_5

Wechsler, D. (1992). *Wechsler intelligence scale for children–Third edition–UK Manual.* Psychological Corp, Europe.

Wechsler, D. (1997). WAIS-III WMS-III technical manual. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI).* The Psychological Corp.

Wechsler, D. (2004a). *Wechsler Intelligence Scale for Children (WISC) (4th ed.).* Pearson Assessment.

Wechsler, D. (2004b). *WISC-IV-UK administration and scoring manual*. Harcourt Assessment.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed). Pearson Assessment.

Weil, T. M., Hayes, S. C., & Capurro, P. (2011). Establishing a deictic relational repertoire in young children. *The Psychological Record*, *61*, 371-390.

White, D. A. & Breen, M. (1998). Edutainment: Gifted education and the perils of misusing multiple intelligence. *Gifted Child Today, 4*(2).

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461–481. https://doi.org/10.1037/0033-2909.91.3.461

White, M. B., & Hall, A. E. (1980). An overview of intelligence testing. *Educational Horizons, 58*(4), 210-216. https://www.jstor.org/stable/42924403

Williams, B., Myerson, J., & Hale, S. (2008). Individual differences, intelligence, and behavior analysis. *Journal of the experimental analysis of behavior*, *90*(2), 219-231.

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. Personality & Individual Differences, 54(8), 871–876.

Wright, A. A., & Delius, J. D. (1994). Scratch and Match: Pigeons Learn Matching and Oddity With Gravel Stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(1), 108–112. https://doi.org/10.1037/0097-7403.20.1.108

Wulfert, E., Greenway, D., Farkas, P., Hayes, S., & Dougher, M. (1994). Correlation between self-reported rigidity and rule-governed insensitivity to operant contingencies. *J Appl Behav Anal*, *27*(4), 659-671. http://dx.doi.org/10.1901/jaba.1994.27-659

Yuan, P., Voelkle, M. C., & Raz, N. (2018). Fluid intelligence and gross structural properties of the cerebral cortex in middle-aged and older adults: A multi-occasion longitudinal study. *NeuroImage, 172*, 21-30. https://doi.org/10.1016/j.neuroimage.2018.01.032

Zimmerman, L.L. & Woo-Sam, J.M. (1973). *Clinical interpretations of the Wechsler Adult Intelligence Scale*. Grune and Stratton.

Appendices

Appendix A. Demographic Questionnaire

Please provide the following information to help us with this research.

D1. What is your age in years? (please type number)

[ ]

D2. What is your gender? (please check button)

- o *Female*
- o *Male*
- o *Non-binary*
- o *Prefer not to say*
- o *Other*

D3. Are you a permanent resident of the UK or Ireland?

- o *Yes*
- o *No*

D4. What was your family's average annual disposable income when you were growing up? Disposable income refers to the total amount of money that is left for living costs after all taxes and other wage deductions have been made by the employer **(please check the button that's closest to your value)**

≤€9,800 (Ireland)	**OR**	*≤£11,400 (UK)*
€9,801 – €15,700 *(Ireland)*	**OR**	*£11,401 – £15,800 (UK)*
€15,701 – €22,400 *(Ireland)*	**OR**	*£15,801 – £20,400 (UK)*
€22,401 – €28,600 *(Ireland)*	**OR**	*£20,401 – £25,900 (UK)*
€28,601 – €34,800 *(Ireland)*	**OR**	*£25,901 – £32,100 (UK)*
€34,801 – €41,700 *(Ireland)*	**OR**	*£32,101 – £39,300 (UK)*
€41,701 – €50,600 *(Ireland)*	**OR**	*£39,301 – £48,100 (UK)*
€50,601 – €61,600 *(Ireland)*	**OR**	*£48,101 – £60,100 (UK)*
€61,601 – €76,600 *(Ireland)*	**OR**	*£60,101 – £80,700 (UK)*
*≥€76,60*0 *(Ireland)*	**OR**	*≥£80,700 (UK)*

D5. What is your household annual disposable income now?
Disposable income refers to the total amount of money that is left for living costs after all taxes and other wage deductions have been made by the employer **(please check the button that's closest to your value)**

≤€12,700 *(Ireland)*	**OR**	≤£12,300 *(UK)*

€12,701 – €19,900 *(Ireland)*	**OR**	£12,301 – £21,400 *(UK)*
€19,901 – €28,300 *(Ireland)*	**OR**	£21,401 – £27,100 *(UK)*
€28,301 – €35,600 *(Ireland)*	**OR**	£27,101 – £32,400 *(UK)*
€35,601 – €42,900 *(Ireland)*	**OR**	£32,401 – £37,000 *(UK)*
€42,901 – €51,500 *(Ireland)*	**OR**	£37,001 – £43,400 *(UK)*
€51,501 – €61,300 *(Ireland)*	**OR**	£43,401 – £49,600 *(UK)*
€61,301 – €72,900 *(Ireland)*	**OR**	£49,601 – £58,600 *(UK)*
€72,901 – €89,100 *(Ireland)*	**OR**	£58,601 – £70,500 *(UK)*
≥€89,100 *(Ireland)*	**OR**	>£70,500 *(UK)*

D6. What is the highest degree or level of education your primary caregiver has completed? *(By primary caregiver, we mean the parent/caregiver you have spent the most time with growing up. If you have spent equal (or roughly equal) amounts of time with multiple adults, please indicate the highest of those levels of education.* ***Please check button)***

- *No formal education*
- *Some secondary school*
- *Junior Certificate / GCSE Examinations*
- *Leaving Certificate / A Levels*
- *Certificate of Higher Education/Diploma*
- *College/ University Degree (Undergraduate)*
- *Master's Degree (Postgraduate)*
- *Ph.D*

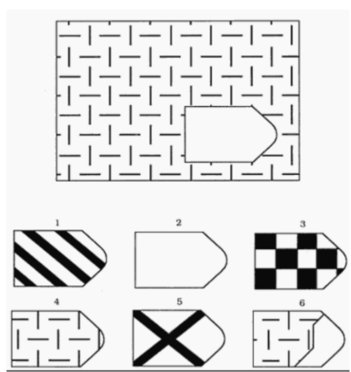
D8. What is the highest degree or level of education you have completed or are currently completing? (please check button)

- *No formal education*
- *Some secondary school*
- *Junior Certificate / GCSE Examinations*
- *Leaving Certificate / A Levels*
- *Certificate of Higher Education/Diploma*
- *College/ University Degree (Undergraduate)*
- *Master's Degree (Postgraduate)*
- *Ph.D*

Appendix B. Information Sheet

This research is being conducted by Carina Kaufmann (carina.kaufmann.2024@mumail.ie) a postgraduate student at the Department of Psychology, Maynooth University, under the supervision of Prof. Bryan Roche (contact: Bryan.T.Roche@mu.ie / 01 708 6026). It is the responsibility of this student to adhere to professional ethical guidelines in their dealings with participants and the collection and handling of data. If you have any concerns about participation, you may refuse to participate or withdraw at any stage.

The study aims to investigate a new test measuring what psychologists call “relational abilities”, which are the skills you need to reason logically. This refers essentially to the ability to understand relationships between objects, such as whether or not they are the same, opposite, smaller than each other, and so on, and to reason logically using these concepts. We would like to understand the relationship between peoples performance on a test for these skills and their performance on tests that measure related skills, in particular a test called Ravens Matrices. For this reason, you will also be asked to complete a Ravens Matrices test, which involves solving simple visual puzzles that require you to spot patterns amongst a range of 2D objects, such as those below. The task is to choose which of the small shapes numbered 1-6 fits best into the space in the larger shape at the top.



Full instructions will be provided at the relevant point.

As part of the study, you will be asked to provide some very basic information about yourself (concerning your age and gender, as well as your socioeconomic background, more specifically, level of education and income). Like the rest of your data, this is completely confidential and anonymous.

You should decline to participate in the study if you are not fluent in English or are not aged between 20 and 24 years. If at any point during the study, you decide to withdraw from the study you are free to simply exit the software by closing your browser. If you choose to do so your data cannot be included in the study. However, if you do complete all stages, it will not be possible to withdraw your data because of the anonymous nature of your participation.

At the end of the study you will be provided a randomly generated four-digit code on-screen. You should note this for proof of participation if you need it for any reason (such as remuneration via Prolific or for course credit).

The data gathered will be compiled and analysed at a group level only and submitted in a postgraduate thesis. This data may also be used as part of analyses for a scientific publication. All data collected will be retained on a university computer in the Department of Psychology for a duration of 10 years as per university regulations. No personally identifying information will be gathered or stored in any form.

At the conclusion of your participation, you will be provided with more information about the purpose of the study, and you will be invited to email the researchers with any further queries you may have.

Please note that this research can only be conducted on a desktop or a laptop computer and should take place in a quiet environment. This task requires concentration, and the data may be of no use to the researchers if it is conducted where there are any distractions or on a small device. **Data will be screened to ensure that participants have engaged with all stages of the task to ensure data quality. Your participation is only useful to us if you complete all stages with concentration.**

While we will hold no personal data of any kind on participants, it must be recognized that, in some circumstances, the confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of investigation by lawful authority. In such circumstances, the University will take all reasonable steps within the law to ensure that confidentiality is maintained to the greatest possible extent.

If during your participation in this study you feel that the information and guidelines provided have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the National University of Ireland Maynooth Ethics Committee at research.ethics@mu.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

Appendix C. Consent Form

By proceeding you are confirming that you have read and understood the information provided to you. You are also confirming that you are between 20 and 24 years, proficient in the English language and a permanent resident of the UK or Ireland. Finally, you are agreeing that you understand that it is not possible to withdraw your data following participation because all data is completely anonymous.

By proceeding you are agreeing that you are between 20 and 24 years of age and that you will conduct this research participation in a quiet relaxed environment at a desktop or laptop computer without distraction. You are also agreeing that you understand that data will have to be screened for elimination of participants whose performance shows evidence of non-engagement.

When you are ready, please press the proceed button below.

Appendix D. Debrief Sheet

Thank you for taking the time to participate in this study. The purpose of the study was to investigate a new measure of intelligence called the Elaborated Relational Abilities Index (RAI). The measure is based on a theory called Relational Frame Theory, which is a theory about language and cognition. Cognition refers to processes associated with learning, knowledge representation, reasoning, perception, attention, memory, and problem solving. This theory has provided evidence over the past number of years that intelligence can in fact be enhanced if only researchers target the right skills. Those skills have been identified as *relational skills,* and it turns out that they are quite easy to teach. However, Psychologists do not yet have a sufficiently well worked out measure of these skills. One of the purposes of this study was to see how scores on a new version of one such available test relates to the performance of individuals on a more well established and traditional test of intelligence.

Psychologists have long known, however, that educational opportunities can influence one's reasoning and cognitive abilities. For this research we also wanted to know how people's income might be an indicator of the opportunities they have received and therefore their performance on these types of tests. This is why we were interested in your parental and personal income and level of education. A second aim of the study was to examine the relationship between the education level of your parents and your own relational skill levels which we expect to correlate positively.

Should you have any questions or concerns about the study you can contact me at carina.kaufmann.2024@mumail.ie or my supervisor for this research, Prof. Bryan Roche at Bryan.T.Roche@mu.ie / 01 708 6026.

Your 4-digit code for proof of participation is C1H0S780.

Appendix E. Original Multiple Regression Analyses Performed for Chapter 2 and Chapter 3

Chapter 2

Sociodemographic Predictors of RAI Performance

Multiple regression was used to test whether own and parental income and education were predictors of RAI performance. The results of the regression indicated that four predictors explained 6% of the variance (R^2 = .06, F(4, 145) = 3.23, p = .014). It was found that parental education significantly predicted RAI performance (β= .22, p = .015, *95% CI* = .00, .03) while the other variables did not significantly predict RAI performance.

Sociodemographic Predictors of Raven's Performance

Multiple regression was used to test whether own and parental income and education were predictors of Raven's performance. The results of the regression indicated that the four predictors did not significantly explain the variance in Raven's performance (R^2 = .03, F(4, 143) = 2.28, p = .063).

Sociodemographic Predictors of RAI-M

Multiple regression was used to test whether own and parental income and education were predictors of RAI-M performance. The results of the regression indicated that four predictors explained 4.8% of the variance (R^2 = .05, F(4, 145) = 2.89, p = .024). It was found that parental education significantly predicted RAI-M performance (β= .21, p = .019, *95% CI* = .00, .03) while the other variables did not significantly predict RAI-M performance.

Chapter 3

Sociodemographic Predictors of Opposition

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Opposition subscale. The results of the regression indicated that the four predictors did not significantly explain the variance in performance on the RAI Opposition subscale (R^2 = -.01, F(4, 147) = .61, p = .656).

Sociodemographic Predictors of Difference

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Difference subscale. The results of the regression indicated that four predictors explained 5.7% of the variance (R^2 = .06, F(4, 146) = 3.28, p = .013). It was found that parental education significantly predicted performance on the RAI Difference subscale (β= .19, p = .041, *95% CI* = .00, .04) while the other variables did not significantly predict performance on the RAI Difference subscale.

Sociodemographic Predictors of Quantity

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Quantity subscale. The results of the regression indicated that four predictors explained 10% of the variance (R^2 = .10, F(4, 145) = 5.14, p < .001). It was found that parental education significantly predicted performance on the RAI Quantity subscale (β= .31, p < .001, *95% CI* = .02, .05) while the other variables did not significantly predict performance on the RAI Quantity subscale

Sociodemographic Predictors of Temporal

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Temporal subscale. The results of the regression indicated that four predictors explained 3.9% of the variance (R^2 = .04, F(4, 146) = 2.51, p = .044). It was found that parental education significantly predicted performance on the RAI Temporal subscale (β= .19, p = .034, *95% CI* = .02, .04) while the other variables did not significantly predict performance on the RAI Temporal subscale.

Sociodemographic Predictors of Containment

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Containment subscale. The results of the regression indicated that four predictors explained 5.9% of the variance (R^2 = .06, F(4, 147) = 3.38, p = .011). It was found that parental education significantly predicted performance on the RAI Containment subscale(β= .26, p = .004, *95% CI* = .01, .05) while the other variables did not significantly predict performance on the RAI Containment subscale.

Sociodemographic Predictors of Analogy

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Analogy subscale. The results of the regression indicated that the four predictors did not significantly explain the variance in performance on the RAI Analogy subscale (R^2 = -.01, F(4, 145) = .64, p = .637).

Sociodemographic Predictors of Deictic

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Deictic subscale. The results of the regression indicated that the four predictors did not significantly explain the variance in performance on the RAI Deictic subscale (R^2 = .00, F(4, 152) = .90, p = .466).

Sociodemographic Predictors of Mathematical

Multiple regression was used to test whether own and parental income and education were predictors of the performance on the RAI Mathematical subscale. The results of the regression indicated that four predictors explained 5.7% of the variance (R^2 = .06, F(4, 146) = 3.26, p = .013). It was found that parental education significantly predicted performance on the RAI Mathematical subscale (β= .18, p = .044, *95% CI* = .00, .04) while the other variables did not significantly predict performance on the RAI Mathematical subscale.

Sociodemographic Predictors of RAI Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI Fluency. The results of the regression indicated that four predictors explained 5.5% of the variance (R^2 = .06, F(4, 145) = 3.19, p = .015). It was found that parental education significantly predicted RAI performance (β= .26, p = .004, *95% CI* = .08, .40) while the other variables did not significantly predict RAI Fluency.

Sociodemographic Predictors of Raven's Fluency

Multiple regression was used to test whether own and parental income and education were predictors of Raven's Fluency. The results of the regression indicated that the four predictors did not significantly explain the variance in Raven's Fluency (R^2 = .00, F(4, 143) = .98, p = .420).

Sociodemographic Predictors of RAI-M Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI-M Fluency. The results of the regression indicated that four predictors explained 5.7% of the variance (R^2 = .06, F(4, 145) = 3.27, p = .013). It was found that parental education significantly predicted RAI-M Fluency (β= .26, p = .004, *95% CI* = .08, .42) while the other variables did not significantly predict RAI-M Fluency.

Sociodemographic Predictors of Opposition Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI Opposition Fluency. The results of the regression indicated that the four predictors did not significantly explain the variance in RAI Opposition Fluency (R^2 = .03, F(4, 147) = 2.19, p = .073).

Sociodemographic Predictors of Difference Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI Difference Fluency. The results of the regression indicated that the four predictors did not significantly explain the variance in RAI Difference Fluency (R^2 = .00, F(4, 146) = 1.14, p = .338).

Sociodemographic Predictors of Quantity Fluency

Multiple regression was used to test whether own and parental income and education were predictors of Fluency on the RAI Quantity subscale. The results of the regression indicated that four predictors explained 8.4% of the variance (R^2 = .08, F(4, 145) = 4.42, p = .002). It was found that parental education significantly predicted Fluency on the RAI Quantity subscale (β= .31, p < .001, *95% CI* = .19, .64) while the other variables did not significantly predict Fluency on the RAI Quantity subscale.

Sociodemographic Predictors of Temporal Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI Temporal Fluency. The results of the regression indicated that the four predictors did not significantly explain the variance in RAI Temporal Fluency (R^2 = .00, F(4, 146) = 1.00, p = .412).

Sociodemographic Predictors of Containment Fluency

Multiple regression was used to test whether own and parental income and education were predictors of RAI Containment Fluency. The results of the regression indicated that the four predictors did not significantly explain the variance in RAI Containment Fluency (R^2 = -.01, F(4, 147) = .79, p = .54).

Sociodemographic Predictors of Analogy Fluency

Multiple regression was used to test whether own and parental income and education were predictors of Fluency on the RAI Analogy subscale. The results of the regression indicated that the four predictors did not significantly explain the variance in Fluency on the RAI Analogy subscale (R^2 = .02, F(4, 145) = 1.68, p = .159).

Sociodemographic Predictors of Deictic Fluency

Multiple regression was used to test whether own and parental income and education were predictors of Fluency on the RAI Deictic subscale. The results of the regression indicated that the four predictors did not significantly explain the variance in Fluency on the RAI Deictic subscale (R^2 = .01, F(4, 152) = 1.35, p = .254).

Sociodemographic Predictors of Mathematical Fluency

Multiple regression was used to test whether own and parental income and education were predictors of the Fluency on the RAI Mathematical subscale. The results of the regression indicated that four predictors explained 4.2% of the variance (R^2 = .04, F(4, 146) = 2.63, p = .037). It was found that own education significantly predicted Fluency on the RAI Mathematical subscale (β= -.25, p = .004, *95% CI* = -2.89, -.54) while the other variables did not significantly predict Fluency on the RAI Mathematical subscale.

Appendix F. Multiple Regression Models – Full Coefficients

Accuracy

Variable	*β*	*p*	*95% CI*
RAI (R^2 = .06, $F(4, 151) = 3.56$, p = .008)			
Parental Income	.00	.740	-.01, .01
Own Income	.00	.317	.00, .01
Parental Education	.02	.008	.00, .03
Own Education	.00	.910	-.02, .02
RSPM (R^2 = .03, $F(4, 150) = 1.98$, p = .100)			
Parental Income	.10	.755	-.55, .75
Own Income	.28	.316	-.27, .83
Parental Education	.54	.187	-.27, 1.35
Own Education	-.95	.102	-2.09, .19
RAI-M (R^2 = .06, $F(4, 151) = 3.27$, p = .013)			
Parental Income	.00	.796	-.01, .01
Own Income	.00	.343	.00, .01
Parental Education	.02	.010	.00, .03
Own Education	.00	.910	-.02, .02
RAI Opposition (R^2 = .00, $F(4, 151) = .92$, p = .457)			
Parental Income	.00	.807	-.01, .01
Own Income	.00	.595	-.01, .01
Parental Education	.01	.084	.00, .03
Own Education	.00	.911	-.02, .02
RAI Difference (R^2 = .01, $F(4, 151) = 1.33$, p = .262)			
Parental Income	.00	.788	-.01, .02
Own Income	.01	.305	-.01, .02
Parental Education	.01	.248	-.01, .03
Own Education	.00	.831	-.03, .02
RAI Quantity (R^2 = .09, $F(4, 151) = 4.93$, p < .001)			
Parental Income	.00	.701	-.02, .01
Own Income	.01	.123	.00, .02
Parental Education	.03	<.001	.01, .05
Own Education	-.01	.578	-.03, .02
RAI Temporal (R^2 = .04, $F(4, 151) = 2.47$, p = .047)			
Parental Income	.01	.460	-.01, .02
Own Income	.00	.623	-.01, .02
Parental Education	.02	.036	.00, .04
Own Education	-.01	.471	-.04, .02
RAI Containment (R^2 = .06, $F(4, 151) = 3.27$, p = .013)			
Parental Income	.00	.963	-.02, .02
Own Income	.01	.433	-.01, .02
Parental Education	.03	.004	.01, .05
Own Education	.00	.781	-.02, .03
RAI Analogy (R^2 = -.01, $F(4, 151) = .76$, p = .556)			
Parental Income	.00	.789	-.01, .01
Own Income	-.01	.186	-.00, .01
Parental Education	.00	.569	-.02, .01
Own Education	.01	.467	-.01, .02

Variable	β	p	95% CI
RAI Deictic (R^2 = .00, F(4, 151) = .84, p = .502)			
Parental Income	.00	.651	-.01, .02
Own Income	.00	.984	-.01, .01
Parental Education	.01	.413	-.01, .03
Own Education	.02	.249	-.01, .05
RAI Mathematical (R^2 = .04, F(4, 151) = 2.71, p = .033)			
Parental Income	.00	.592	-.01, .02
Own Income	.01	.365	-.01, .02
Parental Education	.02	.040	.00, .04
Own Education	.00	.938	-.03, .03

Fluency

Variable	β	p	95% CI
RAI (R^2 = .08, F(4, 151) = 4.42, p = .002)			
Parental Income	.07	.369	-.08, .22
Own Income	-.05	.462	-.18, .08
Parental Education	.34	<.001	.15, .52
Own Education	-.07	.624	-.33, .20
RSPM (R^2 = .00, F(4, 150) = 1.09, p = .362)			
Parental Income	-.03	.634	-.14, .08
Own Income	.06	.222	-.04, .15
Parental Education	.04	.582	-.10, .17
Own Education	-.16	.095	-.35, .03
RAI-M (R^2 = .08, F(4, 151) = 4.54, p = .002)			
Parental Income	.09	.242	-.06, .25
Own Income	-.08	.222	-.21, .05
Parental Education	.35	<.001	.16, .54
Own Education	-.04	.747	-.32, .23
RAI Opposition (R^2 = .05, F(4, 151) = 2.92, p = .023)			
Parental Income	.09	.297	-.08, .26
Own Income	-.12	.100	-.26, .02
Parental Education	.29	.007	.08, .50
Own Education	-.12	.433	-.41, .18
RAI Difference (R^2 = -.02, F(4, 151) = .43, p = .785)			
Parental Income	.05	.893	-.67, .77
Own Income	.00	.997	-.61, .61
Parental Education	.52	.252	-.37, 1.40
Own Education	-.22	.734	-1.48, 1.04
RAI Quantity (R^2 = .01, F(4, 151) = 1.52, p = .199)			
Parental Income	-.02	.950	-.78, .74
Own Income	.34	.304	-.31, .98
Parental Education	.77	.108	.17, 1.71
Own Education	.71	.297	-2.04, .63
RAI Temporal (R^2 = .03, F(4, 151) = 2.17, p = .075)			
Parental Income	1.21	.034	.09, 2.33
Own Income	-.88	.070	-1.83, .07
Parental Education	.62	.377	-.76, 2.00
Own Education	-1.05	.295	-3.02, .92

Variable	β	*p*	*95% CI*
RAI Containment (R^2 = .00, $F(4, 151) = 1.14$, $p = .342$)			
Parental Income	-1.00	.089	-2.15, .15
Own Income	.71	.157	-.28, 1.69
Parental Education	.61	.397	-.81, 2.04
Own Education	.56	.586	-1.47, 2.59
RAI Analogy (R^2 = -.02, $F(4, 151) = .27$, $p = .897$)			
Parental Income	-.09	.720	-.57, .39
Own Income	-08	.688	-.33, .49
Parental Education	.08	.788	-.51, .68
Own Education	.32	.456	-.53, 1.17
RAI Deictic (R^2 = -.01, $F(4, 151) = .74$, $p = .569$)			
Parental Income	-.20	.351	-.61, .22
Own Income	-.15	.412	-.21, .50
Parental Education	-.10	.711	-.42, .61
Own Education	.40	.280	-.33, 1.13
RAI Mathematical (R^2 = .06, $F(4, 151) = 3.53$, $p = .009$)			
Parental Income	-.94	.019	-1.72, -.16
Own Income	1.09	.002	.42, 1.76
Parental Education	.81	.098	-.15, 1.78
Own Education	-1.25	.075	-2.62, .13